91st Congress
1st Session } JOINT COMMITTEE PRINT

THE ANALYSIS AND EVALUATION OF PUBLIC EXPENDITURES: THE PPB SYSTEM

A COMPENDIUM OF PAPERS

SUBMITTED TO THE

SUBCOMMITTEE ON ECONOMY IN GOVERNMENT

OF THE

JOINT ECONOMIC COMMITTEE

CONGRESS OF THE UNITED STATES

VOLUME 1

Part I. The Appropriate Functions of Government in an Enterprise System

Part II. Institutional Factors Affecting Efficient Public Expenditure Policy

Part III. Some Problems of Analysis in Evaluating Public Expenditure Alternatives

Printed for the use of the Joint Economic Committee

U.S. GOVERNMENT PRINTING OFFICE

27-877 WASHINGTON : 1969

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $2.75

LETTERS OF TRANSMITTAL

MAY 29, 1969.

To the Members of the Joint Economic Committee:

Transmitted herewith for the use of the members of the Joint Economic Committee and other Members of Congress is a three-volume study entitled "The Analysis and Evaluation of Public Expenditures: The PPB System," prepared for the Subcommittee on Economy in Government.

The views expressed in this document do not necessarily represent the views of members of the committee or the committee staff, but are statements of issues and alternatives intended to provide a focus for hearings and debate.

WRIGHT PATMAN,
Chairman, Joint Economic Committee.

MAY 28, 1969.

Hon. WRIGHT PATMAN,
Chairman, Joint Economic Committee,
Congress of the United States,
Washington, D.C.

DEAR MR. CHAIRMAN: Transmitted herewith is a three-volume compendium of study papers entitled "The Analysis and Evaluation of Public Expenditures: The PPB System." The contributions which it contains focus on the major policy issues and analytic problems which must be resolved if public spending programs are to become more efficient and responsive to the needs of the people.

This three-volume study is divided into six parts. Part I considers the issues of economics and equity pertinent to an optimal decision on the appropriate functions of government in a market economy. The papers in part II attempt to isolate some of the factors which influence the ability of the Federal Government to attain efficiency in public expenditure policy. In part III a few of the key problems in applying economic analysis to public expenditures are discussed and recommendations for their solution are offered. The second volume of the study (pt. IV) contains a comprehensive discussion of the Planning-Programing-Budgeting System in the Federal Government. Prepared by the Assistant Director of the Bureau of the Budget for Program Evaluation, this paper describes the plans of the Bureau for improving the system. The final two parts of the study focus directly on the PPB system. The papers in part V address the system itself and appraise and evaluate its structure and performance to date. Those in part VI direct attention to the major issues for policy analysis in each of the primary functional areas of the Federal budget.

This study contains contributions by over 50 invited experts. The subcommittee is indebted to these authors for their excellent con-

tributions, which we believe will contribute substantially to improvements in public policy and public expenditure analysis. The time and learning devoted to the preparation of these papers should do much to stimulate interest in these matters by both economists and policymakers.

Dr. Robert Haveman, who is on leave of absence from Grinnell College, is responsible for planning and preparation of the compendium, with the research and editorial assistance of Gail Steg, the administrative and secretarial assistance of Anne McAfee, and the advice and suggestions of other members of the committee's professional staff.

As the executive director's letter indicates, the compendium should not be viewed as an expression of views or conclusions of the committee staff, nor should it be viewed as an expression of views of the subcommittee or individual members.

WILLIAM PROXMIRE,
Chairman, Subcommittee on Economy in Government.

MAY 27, 1969.

Hon. WILLIAM PROXMIRE,
Chairman, Subcommittee on Economy in Government, Joint Economic Committee, U.S. Congress, Washington, D.C.

DEAR SENATOR PROXMIRE: Transmitted herewith is a three-volume study which examines the problems and possibilities in attaining efficiency in Federal Government expenditure policy. The compendium, which contains 50 papers, is entitled "The Analysis and Evaluation of Public Expenditures: The PPB System."

It is hoped that the contributions made in this study will stimulate widespread discussion among economists and policymakers on the procedures for determining the most effective among the many alternative spending proposals with which the Federal Government is confronted. These papers should assist analysts in the Federal Government in applying appropriate economic analysis to the spending programs with which they are concerned and make the Congress aware of the kinds of helpful information which explicit and quantitative policy analysis can provide them. Finally, by focusing on the PPB system, the study should provide major insight and assistance to the new administration in their efforts to improve and strengthen the evaluation of public spending programs.

The major work in planning, compiling, and editing this study was undertaken by Dr. Robert Haveman, who is on leave of absence from Grinnell College, with the advice and suggestions of other members of the staff. He was assisted in research and editorial work by Gail Steg and in administrative and secretarial work by Anne McAfee. The papers contained herein should be interpreted as representing only the opinions of their authors.

JOHN R. STARK,
Executive Director, Joint Economic Committee.

FOREWORD . . .

PPB, THE AGENCIES AND THE CONGRESS

SENATOR WILLIAM PROXMIRE*

Chairman, Subcommittee on Economy in Government

It is now 3½ years since all major Federal agencies were instructed to develop and implement planning-programing-budgeting systems. During this period there has been great activity in connection with program analysis and evaluation and a tremendous amount of discussion and debate. There has not been, however, any systematic look at how the application of the tools of economic and systems analysis has worked out in practice. There has been no comprehensive study of the lessons which have been learned, the changes which have been made, and the policies which should be followed in the future. Because of increasing interest in efforts to develop a more rational decision process, it seemed appropriate for the Joint Economic Committee to try to fill this gap.

The form chosen is that of a compendium, a collection of papers by both scholars and practitioners in the areas of public finance, systematic analysis, and program budgeting. The reason for this choice is, I think, fairly obvious—perceptions and conclusions differ widely, and there is no way to produce a "definitive" work. To attempt to do so would mean forcing a developing field of public policy analysis into a limited and artificial mold, something that we certainly wished to avoid.

Instead, our hope was that by combining the thoughtful efforts and differing perspectives of a diverse group we could produce a work that would be valuable both to Government officials interested in improving the policymaking process and to students of Government decisionmaking and innovation. As I read through the contributions made to the compendium, I am confident that this hope has been fulfilled.

THE PPB SYSTEM AND RATIONAL DECISIONMAKING

It should be emphasized that the use of PPB and systematic analysis in the Government is not a partisan issue. While originally implemented pursuant to the instruction of President Johnson, it also is supported by the new administration. As Budget Director Robert Mayo has stated, it is now quite clear that any administration needs techniques of program evaluation if it is to make effective decisions on resource allocation.

---

*This paper has been adapted from a speech given by Senator Proxmire to the Agency Program Planning Officers Group luncheon meeting on Thursday, Apr. 24, 1969. Budget Bureau Director Robert Mayo introduced Senator Proxmire.

The absence of partisan dispute over the use of PPB points to the recognition by responsible Government officials that we must be rational in our approach to public policy decisions. For, to use PPB to obtain information about the gains and losses to be anticipated from a decision is to demand no more than that the decision be rational. Properly defined, PPB is the most basic and logical planning tool which exists: it provides for the quantitative evaluation of the economic benefits and the economic costs of program alternatives, both now and in the future, in relation to analyses of similar programs.

Any decisionmaker, whether he be the head of a household or the head of a business firm, must rely on the comparison of the gains and costs of his decisions if he is to be successful at achieving his objectives. To ignore the careful consideration of gains and losses is equivalent to saying that he has no objective at all; no goal which he is attempting to achieve. While the objectives of the Federal Government are less tangible and more complex than those of a household or a business firm, they do exist, and analysis should be carried out to determine which of our alternatives will allow us to satisfy these objectives at least cost. I would add that the very effort of attempting to evaluate alternatives is of substantial assistance in determining what our objectives really are.

I have never been able to understand why we are only now getting around to the task of developing such a system of analysis and evaluation. It is even more difficult for me to understand why many official and private groups sometimes object so violently to the application of this logic to public sector choices. Obviously, they themselves demand such information before they buy a new car or trade 15 shares of one common stock for seven shares of another.

The Congressional Budgetary Process

As a U.S. Senator, I also have a strong interest in the potential of PPB for improving decisionmaking in the legislative branch as well as in the executive. This is a very important possibility because, in my view, the legislative resource allocation process is sorely in need of improvement. In a very real sense, the congressional appropriation process is a classic example of an *inexplicit*, *closed*, and *uninformed* decision process. This does not mean that the executive budgetary process is perfect, or even that it is, in fact, very good on any absolute scale of values. But it is both informed and open compared with the budgetary process which exists in the legislative branch.

In the Congress, with its appropriations committees and subcommittees, there is very little explicit consideration of program objectives or tradeoffs, of alternative means of attaining objectives, or of the benefits and costs of budget proposals this year and in the future. In short, Congress does not really give the budget a meaningful review because it fails to ask the right questions. Perhaps the primary reason for this is the traditional policy of executive branch dealings with the Congress. The executive branch comes to Congress with only one budget, with only one set of program proposals, and typically with no quantitative information on the benefits and the costs of even their own proposals. In fact, the only program area in which the Congress is presented with substantive cost-benefit evaluation information is that for water resources development. Since the Flood Control Act of 1936, project proposals in this area have been accompanied by a

benefit-cost ratio. This number enables Congressmen and Senators to get some sense of the economic value of the choices which they are making and of the implicit costs involved when they choose to accept a project with a low benefit-cost ratio despite the fact that one displaying a higher ratio is available. (Even so, the usefulness of these analyses has been impaired by the use of artificially low discount rates in computing the present values of benefits over time. This has made bad projects look far better than they should.)

A second reason why the Congress has performed so badly in the budgetary and appropriations area has to do with the interests of Congressmen and Senators. Many in the legislative branch have little interest in or patience for careful deliberations on budgetary matters. The careful consideration of alternatives requires much effort and concentrated study of the relative merits and demerits, the costs and the gains, of alternative policy proposals. This is hard and grubby work. Those not used to thinking in such terms find it easier simply to rely on the executive agencies. Unfortunately, these agencies are often more interested in selling their programs, regardless of merit, than in having Congress analyze them.

A final reason for Congress' poor performance in this area is the severe staffing constraints under which the legislative branch operates. Currently, we do not have the staff either to interpret or to evaluate the analysis done by the executive branch were it presented to us, nor does Congress have the staff to do policy analysis of its own. Indeed, in my judgment, this is one of the primary barriers to the ability of the Congress to fulfill its mandate as controller of the public purse. Dr. Jack Carlson, who is Assistant Director of the Bureau of the Budget, stated this well in his recent testimony:

> You [the Congress] have some outstanding people who can provide program evaluation, but very few. I frankly think that Congress is not very well equipped to provide that evaluation.

Nonetheless, even if the interest and the staff existed, there would still be substantial organizational problems to hinder an effective public expenditure decision process. A primary difficulty is the organization and structure of the Appropriations Committees. In considering appropriations requests from the executive, we in the Congress have organized ourselves into appropriations committees and subcommittees with each subcommittee having control over a particular portion of the budget. The subcommittees consider the executive's proposed budget, deliberate on it, perhaps amend it, and ultimately report out an appropriations bill. The structure of this arrangement is such that the powerful people on the appropriations subcommittees—the Chairmen—almost inevitably desire to see the budgets which they oversee rise. They are not interested in careful scrutiny and evaluation of their own budgets. Other budgets should be cut, of course, but everyone knows that defense (or agriculture, or space, or public works, as the case may be) is "absolutely necessary" to the further growth and prosperity of the Nation.

I happen to be on the steering committee of the Democratic Party. It is this committee which assigns the Democratic membership to the available committee vacancies. In the deliberations of this committee, there are enormous pressures to place those Senators whose States benefit from, say, public works appropriations on either the Senator In-

terior Committee or the Public Works Subcommittee of the Appropriations Committee. In fact, a Senator who is from a State which benefits substantially from these programs is, at least in the short term, rather clearly serving his own best interests and those of at least some of his constituents if he attains a seat on one of these committees. The net result of all of this, however, is that the committee structure develops a built-in bias toward higher budgets. Because the people who serve on each committee have an interest in seeing the budget for which they are responsible increase, they often fail to encourage careful evaluation and analysis of expenditures.

An example of the bias which results from this process is clearly seen by observing the State membership of the Senate Committee on Interior and Insular Affairs. The Democratic members on that committee are from Washington, New Mexico, Nevada, Idaho, Utah, North Dakota, South Dakota, Wisconsin, Montana, and Alaska. The Republican membership is from Colorado, Idaho, Arizona, Wyoming, Oregon, Alaska, and Oklahoma. With the exception of my able colleague, Gaylord Nelson, there is no Senator on this committee representing a State east of the Mississippi River. A similar kind of situation holds in the Public Works Subcommittee of the Senate Appropriations Committee. The Democratic membership of this committee represents Louisiana, Georgia, Arkansas, Washington, Florida, Mississippi, Rhode Island, Nevada, West Virginia, and Wyoming. Again, a substantial concentration of Senators from those Southern and Western States which receive major water resource appropriations. Much the same is true with the Republicans on that subcommittee, although I should add that at least two of these are from the Eastern States—Maine and New Jersey.

Largely as an outgrowth of this built-in committee bias, the relationships between the staffs of the committees and subcommittees and their counterparts in the executive agencies is hardly one of arms-length dealings. The degree of mutuality of interest between the executive staff and those on legislative branch committees is substantial. I would add that this problem is not peculiar to legislative-executive relationships. The serious colleagiality between Budget Bureau examiners who work on the military budget and their counterparts in the Pentagon has recently been the cause of much concern.

Toward an Improved Appropriation Process

Given the institutional constraints which inhibit change in this situation, is there anything which can be done to improve the congressional budgetmaking process? In my judgment, there are a number of important steps which can be taken. Many of them entail the bringing to bear of additional PPB-type information on the appropriations process. Congressmen and Senators who are concerned with national priorities and efficiency in Government must have the information and data necessary to raise and debate the right basic questions about program effectiveness and worth.

Building a Capability To Ask the Right Questions: The First Step

The most basic and elementary step which the Congress needs to take in improving the appropriation process is to develop a capability to ask the right questions. Whether this means a substantial increase in staff capability or a special office of budgetary analysis or an in-

crease in the PPB capability of the General Accounting Office is not clear. What is clear, however, is that the Congress cannot respond to the demands of the people, cannot establish proper national priorities, cannot improve the quality of its decisions, cannot properly scrutinize the executive budget unless it equips itself to ask the right questions.

The right basic questions are those having to do with the outputs of a program and its inputs and the economic values of each. They are questions concerning the total costs of program decisions, and not just the given year costs. They are questions having to do with the distribution of a program's costs and benefits among the people. We must, for example, determine the economic losses which will be sustained (or gains which will be forgone) if program X is reduced by 10 or 50 percent, or increased by 10 or 50 percent.

The following are a few examples of the kinds of questions which I have in mind:

- What, for example, are the real national security costs of removing Southeast Asia from the primary defense perimeter and what are the budgetary savings from its removal? On the basis of very little evidence and information, I am inclined to say that the costs of removing Southeast Asia may well exceed the value of the budgetary savings which we would experience. However, I cannot make a rational decision on this matter, nor can my colleagues in the Congress, unless we have the best analysis available on the costs and gains of such a policy alternative.
- What would be the national security impact of a 30-percent reduction of total U.S. ground forces, and what would be the budgetary savings from this reduction? An article in the *Congressional Quarterly** claimed that $10 billion could be cut from the defense budget with no loss of national security effectiveness. Over 50 percent of this suggested $10 billion cut was in the area of manpower. The efficiency of the Department of Defense in the handling of manpower policy is very low. Indeed, the national security costs of reducing ground forces by 30 percent may well be zero. In any case, it is evidence—data and information—on the costs and gains of that sort of decision which Congress requires if the level of rationality is to be increased.
- What are the total costs of adding a nuclear carrier force with all of its required support to our existing 15-carrier complex? What would be the gain in national security? How much elementary and secondary education could we purchase for the dollar cost of the new carrier?
- What national economic benefits would the Nation sacrifice and what national costs would it avoid, if the Trinity River project is not constructed? This project involves the creation of a channel from Dallas-Fort Worth to the Gulf of Mexico. Some observers have argued that it would be cheaper to move Dallas-Fort Worth to the Gulf than to construct this channel.
- What benefits are available from manned space flights that are not available from unmanned flights? What are the incremental costs of manned over unmanned flights? The space agency is now asking us for funds for 10 moon landings and

*Congressional Quarterly, June 28, 1968.

for the exploration of still additional planets. Those planets are going to be there 10 years from now, or even 20 years from now. On what basis can we justify the current expenditure of these funds in view of the other social objectives which we would obtain if these funds were not allocated to the space program? Moreover, some scientists believe that all of the information that we need from space flights can be obtained from unmanned flights, that manned flights are not necessary for this purpose. We need hard analysis of this decision.

- What are the real costs to the American economy of specific protectionist measures that are sought by industry, such as the oil import program? What, in hard economic terms, do similar measures by other countries cost us? Such information is essential for effective bargaining.
- How much do we spend to maintain the military capability to keep open important transportation bottlenecks, such as the Panama Canal, Gibraltar, or the Straits of Malacca? What costs would be incurred if such bottlenecks were not open?
- What is the relationship between resources put into Federal criminal investigation, prosecution and judicial activities and the outputs of those activities in terms of cases actually processed? What are the benefits obtainable through Federal payments for increasing the number of State and local law enforcement personnel versus those obtainable from increasing the support available to existing personnel? In particular, to what extent are trained police officers now used less than optimally because of a lack of subprofessionals, dictating equipment, vehicles, cameras, or other fairly elementary support items?
- Which policy of preschool education produces greater benefits: a policy which is going to reach all poverty children to at least some extent, or a program of intensive work with fewer children?
- What economic losses will be incurred in the future—in terms of loss of productivity and increased welfare costs—that could be prevented by child nutritional and health care programs? How do the benefits available from such programs compare with the benefits available from further extension of the medicare program? For each type of program, upon whom would the costs and benefits fall or accrue?
- What are the costs and benefits involved in the construction of mass transit systems in cities which do not presently have them? What should be included in our calculation of benefits, and how accurate can we be in our judgments? In the Northeastern United States, are the costs of constructing a high-speed ground transit system for intermediate intercity journeys less than those of constructing additional airport capacity?
- What is the likely yield from the Government's investment in fast breeder reactor R. & D., and how does it compare with the return that the relevant private sector would demand? Are there possibilities for international cooperation that would avoid the overlap between this work and similar work in other countries?

These are the kinds of questions that Congress needs to ask, and for which responsible executive branch agencies must develop and supply answers. In my judgment, concerned Congressmen and Senators can

reduce much gross waste from our budgets if we can first develop enough information to ask the right questions, and second, have the cooperation of the executive branch in getting answers.

In this same vein, it seems to me that the current ABM discussion which is going on in the Congress is one of the few examples of careful policy analysis by the legislative branch. It is a case in which Congress—the whole Congress—is asking the right questions about the benefits which will be achieved from this decision, about the costs which it will entail. As in good policy analysis, the question of objectives is being explicitly discussed and the interrelationships between the program proposal and the attainment of objectives is being investigated with some care. It is my belief that with more PPB-type information, the Congress can do this kind of policy analysis on increasing numbers of issues and expenditure proposals.

Gaining Access to Appropriate Data and Analysis: A Second Step

In addition to developing the capability to ask the right questions, the Congress needs to be provided with certain basic kinds of PPB-type information on an ongoing basis. The executive branch must be asked to develop this information and submit it to the Congress in appropriate form. The Bureau of the Budget must assume the leadership in this effort. Let me describe a few specific kinds of information which are essential to a more open and explicit congressional decision process.

Overview information

The first of these items of analysis and data I will call "Overview Information." We need a display of each program in the Federal budget and an estimate of its benefit-cost ratio—that is, the efficiency impact of that program. We also need information on the distributional pattern of project outputs by income level, race, and geographic location—its equity impact. This information is often as important to those of us in the legislative branch as is the efficiency information. We can frame good policy only if we have knowledge of who we are helping when we appropriate money and who is bearing the cost. Even though many of these estimates would have to be rough, they would generate a major improvement in the appropriation process by giving Congress a better perspective on the probable impacts of these public expenditures. I urge the agencies to develop this kind of information, and I urge the Bureau of the Budget to collect and supply it to the Congress for individual programs and in summary form. I should note that in recent hearings before the Subcommittee on Economy in Government, Dr. Jack Carlson of the Bureau of the Budget presented us with a sample format for this overview information and some preliminary data. The format is an excellent one. We now need the calculations to be made and the tables completed.

Budget projections

A second body of information which Congress requires is out-year budget information. For each program, what are the expenditures to which we are committed over the next 5 years because of decisions which we have already made? For each new program proposal, what are the total 5- or 10-year costs entailed by the decision? An example of what happens when we do not have this kind of information is the Higher Education Act of 1965 (Public Law 89–329). In this legisla-

tion, we provided thousands of student scholarships for the first year without really recognizing that to maintain our commitment the funding would have to double in the second year, triple in the third year, and quadruple in the fourth. By keeping the program at its present level, and refusing to honor the implied commitment, we have placed college and university administrators in an impossible position. They now either have to reduce the scholarship aid for the class which entered school last year, or they have to completely eliminate scholarship aid from this source for students currently entering school. If Congress had been oriented towards explicit consideration of the future costs of present decisions I think it would have avoided this bind.

I urge the executive branch to formulate a framework and procedure to develop this out-year budget information across the Federal budget and to present it to the Congress. Moreover, I would propose that the President use the out-year budget framework which is developed to convey his budgetary priorities to the Congress. The numbers which he would place in the appropriate slots in this framework would not commit him, and would change over time. However, they would show the level of program outlays for which commitments have already been made as well as the budgetary areas to which the President would like to see uncommitted funds devoted. They would give the Congress an ongoing description of how the President hoped to allocate the Federal budget over the next several years and how much discretionary room remains in the budget if existing laws remain unchanged. They would give the Congress a bird's-eye view of the Executive's plans and priorities. I would hope that the Bureau of the Budget could play the leadership role in developing this information.

Quantitative analysis of alternatives

The final type of information which is essential to improvements in Congress' performance of its budgetary function entails the quantitative economic analysis of alternatives. As stated earlier, when the administration comes to Congress with a new program, it typically comes with a single recommendation. If Congress is to effectively carry out its decisionmaking role, it must do more than simply accept or reject an administrative recommendation. The Congress needs to be presented with a number of alternatives which would achieve a given objective. These alternatives should be accompanied by quantitative analyses of the benefits and the costs of each. It is only slightly less than absurd that the Congress is expected to participate meaningfully in the policymaking process when it is not asked not to consider alternatives, but only to approve or disapprove or to amend slightly at the margins. This problem is especially severe in the area of defense spending and military budgets. The development of a changed policy on the part of the executive branch in this area will, I suspect, be long in coming. Current policies are rooted in the concrete of both tradition and realistic gamesmanship. Nevertheless, it is something that we should work hard to change.

The Further Development of the PPB System

All of these improvements in PPB in the legislative branch are tied to the further development of the PPB system by the executive.

As is obvious, I am a strong supporter of program analysis. I also

think the efforts that have been made recently to strengthen the process are important. In particular, the narrowing of the number of issues which receive special analytic attention was an important step, as is the insistence that these issues deal with the larger budget questions. Hopefully, agencies will be able to respond with more quantitative and more pointed analyses on the reduced list of issues. I also support the goal of increasing the role of agencies in the PPB process.

In my judgment, of high priority to the further development of PPB systems is the issuance by the Bureau of the Budget of a number of guideline documents to insure consistency in the economic analysis of public expenditures applied throughout the Federal Government. Last year, the Subcommittee on Economy in Government learned of the enormous divergence in the discounting analysis of public investment programs. The interest rates used ranged from zero percent in some programs to 20 percent in others. In testimony before the subcommittee, we learned from reputable economists that the discount rate to be used by public agencies should be at least 8 percent. This would eliminate the economic waste of diversion of resources from the private sector, where they are producing at least this return, to the public sector where, if rates of discount lower than this are applied, they will be likely to produce less. As stated earlier, I am well aware that the equity aspects are as important as the efficiency ones. However, one should not think that programs with low rates of return automatically produce equity, because they do not. Nor do I doubt our ability to find programs which meet both sets of criteria.

In the report of the Subcommittee on Economy in Government, we recommended that (1) the Bureau of the Budget should require all agencies to develop and implement consistent and appropriate discounting procedures on all Federal investments entailing future costs or benefits; and (2) the Bureau of the Budget, in conjunction with other appropriate Government agencies, should immediately undertake a study to estimate the weighted average opportunity-cost of private spending which is displaced when the Federal Government finances its expenditures. In response to these recommendations, the Bureau of the Budget has assured us that it is developing a guideline document to insure consistency in discounting practices across the Federal Government. I am anxious to see how the subcommittee recommendations are going to be implemented by the Bureau and Federal agencies.

On the basis of recent hearings before the Subcommittee on Economy in Government, I judge that Federal Government practice in benefit estimation is also extremely disparate. The issuance of a guideline document on the procedures for benefit estimation is also necessary. We need to develop a consistent concept of program benefits viewed from a national accounting stance. We need to establish a consistent procedure for handling benefits such as regional effects and secondary impacts, which are not appropriately considered from a national economic viewpoint.

In addition to increasing the role of consistent analysis through the issuance of guideline documents, the executive branch should build explicit procedures for the ongoing evaluation and appraisal of programs into new and experimental social programs. The Congress should require that provision for ongoing evaluation be included in

appropriations for these programs. We know little about the kinds of inputs and program structures which will yield the outputs we desire and if we ever hope to generate improvements in programs in the areas of education, health, labor retraining, and so on, we must have followup evaluation. This information must be available to Congress on an ongoing basis as these programs evolve.

Finally, we need a new budget analysis which breaks down and evaluates the economic impact of tax expenditures, as well as direct expenditures. In testimony before the Joint Economic Committee, Joseph Barr, former Secretary of the Treasury, pointed out that the special provisions, exceptions, and deductions in the Federal tax structure cause an enormous reallocation of the Nation's resources. The volume of these tax expenditures is huge; in some of the functional categories of the Federal budget they outweigh direct expenditures. So far we have little analysis of these expenditures; we know very little about the kinds of outputs which they are producing, and the kinds of resource diversions they entail. The Federal budget should include information on these items, as well as the information which it currently includes. I call upon the Bureau of the Budget to develop a new budget format to include a description of both direct and tax expenditures.

Conclusion

In this compendium of papers, a large number of additional proposals for improving the analysis of public expenditures are made. In my view, these papers will make a valuable contribution to the quest for an effective system of policy analysis and program evaluation in both the Congress and the executive branch. They should serve to focus attention on the importance of applying economic analysis to Government decisions, and on the extent to which the public interest suffers in the absence of such analysis. The views presented on many facets of this issue should provide valuable insights into the problems involved in rationalizing the budget process and the evaluation of policy alternatives, and should point to possible solutions to some of the difficulties currently encountered. It is my hope that this compendium will stimulate greater efforts on the part of both congressional and executive decisionmakers to enhance the effectiveness of the public policy decision process.

CONTENTS

VOLUME 1

SECTION B: SHADOW PRICES FOR UNPRICED OR IMPERFECTLY PRICED INPUTS AND OUTPUTS

SECTION C: MULTIPLE OBJECTIVES AND REGIONAL IMPACTS IN A CONSISTENT ANALYTICAL FRAMEWORK

AUTHORS LISTED ALPHABETICALLY

NOTE

The copyright material reproduced in this volume was printed with the kind consent of the copyright holders.

Section 8, title 17, United States Code, provides:

"The publication or republication by the Government, either separately or in a public document, of any material in which copyright is subsisting shall not be taken to cause any abridgement or annulment of the copyright or to authorize any use or appropriation of such copyright material without the consent of the copyright proprietor."

The subcommittee requests that no material in this volume be republished or reprinted without the permission of the author.

THE ANALYSIS AND EVALUATION OF PUBLIC EXPENDITURES: THE PPB SYSTEM

A compendium of papers submitted to the Subcommittee on Economy in Government of the Joint Economic Committee, Congress of the United States, May 1969

INTRODUCTION . . .

THE ANALYSIS AND EVALUATION OF PUBLIC EXPENDITURES: AN OVERVIEW

By Robert H. Haveman*

I

In the postwar period, both the theory and practice of public sector economics have undergone substantial change. This reorientation has been reflected in both the literature of economics dealing with public finance and governmental budgeting matters and the practice of public expenditure analysis, decisionmaking, and budgeting. From a primary concern with the principles of taxation, public finance economists have, in recent years, devoted increasing attention to the economics of public expenditure decisions. Similarly, in practice, the application of economic analysis to matters of taxation policy has been supplemented by major efforts to evaluate and appraise public expenditure alternatives. Developments in these two areas over the past decade reflect a new concern with establishing a rational and economic decisionmaking process for allocating the public budget.

Although the public budget has both a revenue and an expenditure side, pre-1950 public finance literature focused primarily on the tax side of the account. Theorizing about the economic effects of alternative tax policies on both the allocation of resources and the size of the national income was well developed by the 1950's. Moreover, substantial empirical work on the differential effects of alternative tax structures had been undertaken. Until recent years, the study of public finance was widely interpreted to be the study of the "Economics of Taxation."

Relative to the economic analysis of tax policy, prewar economics of public expenditures was undeveloped territory. For a number of reasons, public finance experts treated the expenditure side of the budget in a most cavalier fashion. While general discussions of the

*Associate Professor of Economics, Grinnell College (on leave) and senior economist, Subcommittee on Economy in Government, Joint Economic Committee. The views expressed here are not necessarily those of the Joint Economic Committee.

impact of Government spending on the allocation of resources and the size of the national income can be found prior to the 1950's, little rigorous analytical and empirical work was undertaken during that period.

In part, the lagged development of public expenditure economics is due to a lack of quantitative information on the composition and incidence of expenditure programs. It is also due to differences in the institutions surrounding public decisions on taxes and those on expenditures. While decisions on tax policy typically involve large sums of money and affect the vast majority, if not all, citizens, decisions on individual public expenditures usually involve relatively small sums of money and directly affect only limited groups of citizens. Moreover, at the Federal Government level, taxation policy is handled by a single executive agency and a single congressional committee; expenditure decisions are divided among 20 to 30 Government agencies and substantive program decisions are made by several congressional committees. While taxes are visible and painful, the benefits of public expenditures are less visible and often intangible.

The slow development of public expenditure economics was reflected in Federal Government practice. In the pre-1950 period, analyses of the effects of Federal revenue policy were undertaken as a regular matter by the U.S. Treasury Department. While the quantity and pertinence of these analyses have increased over the last two decades, they did exist in the earlier period and they did influence the development of tax policy. On the other hand, the analysis of Federal spending programs was limited in both quantity and quality. While isolated analyses of the relative worth of particular Federal spending programs or policy decisions were made, there was no systematic effort to appraise the value of alternative means of attaining Government objectives or to base public spending decisions on analysis. Clearly this absence of analysis in the governmental expenditure decision process was not unrelated to the lack of theoretical and empirical literature on the economics of public expenditures.

In the Federal Government, water resource development was the only area in which formal efforts to analyze the worth of public spending proposals were made. This was due to the Flood Control Act of 1936 which required the U.S. Army Corps of Engineers to evaluate the benefits and costs of all water resource projects, "to whomsoever they accrue." While early analytic efforts in the water resources area could be—and have been—criticized for lack of conceptual and empirical sophistication, their existence served to trigger substantial efforts to apply the logic of economics to budget allocation decisions in other areas.

II

During the late 1950's and the 1960's, the literature of economics began to reflect a concern with the effect of public expenditures on the size and composition of the gross national product, the allocation of national resources, and the distribution of income. While the factors which led to this new concern on the part of economists are many and complex, a few of them are sufficiently clear to warrant special comment.

An early influence on professional economists was the appearance in the early 1950's of the important theoretical study by W. Baumol, *Welfare Economics and the Theory of the State.*[1] Although pessimistic as to the ultimate ability of economic analysis to speak with meaning about the influence of public decisions on social welfare, this analysis made discussions of private market failure and its causes a legitimate enterprise. It also legitimized the search for a criterion by which to isolate those public decisions which increase society's economic welfare from those which do not.

In the period following publication of this volume, a number of economic theorists began to focus on issues of collective decisionmaking and efficient public choice. Their contributions challenged the classical presumption that economic efficiency would be attained with voluntary decisionmaking occurring in free markets by independent buyers and sellers and a public sector with minor economic functions. This literature examined the characteristics of free market operation and clarified those conditions in which social costs and gains diverge from their private counterparts. The circumstances in which the private market economy would fail to provide certain worthwhile outputs and would overproduce some or underproduce others became well understood. The characteristics of public goods were identified and the reasons why markets fail to produce them were identified. External effects or spillover values and the failure of private sector decisions to account for them became an integral part of standard economic theory. Market power, decreasing costs, immobilities, and lack of knowledge and their ruinous effects on the efficient operation of the private economy became part-and-parcel of realistic economic analysis. In short, "market failure" took on a significance which it did not previously have.

The ensuing developments in economic thought coincided with the relaxation by economists generally of the traditional presumption that the basic structure of the economy was that of free enterprise in which Government activities represented only a minor aberration from universal private decisionmaking. The increasing inability of neoclassical analysis to explain the behavior of markets dominated by oligopoly, advertising, and administered prices clearly contributed to this altered viewpoint. The rapid postwar growth of the public sector, both absolutely and relative to the private, also encouraged abandonment of the earlier perspective. The 20 percent of the Nation's gross national product accounted for by Government purchases of goods and services, and the 8 to 10 percent of the Nation's final output regularly devoted to the military budget, made it difficult to pass off the public sector as a minor flaw in an otherwise smoothly functioning market system.

Related to the theoretical contributions of welfare economics was a more applied and empirical line of development. If major public expenditures are necessary to attain certain worthwhile objectives—to correct for market failure—how can those proposals which contribute to society's economic welfare be separated from those that detract

[1] Cambridge: Harvard University Press, 1952. This work, it should be noted, drew on and supplemented some earlier work in welfare economics, notably that of J. R. Hicks, N. Kaldor, T. Suranyi-Unger, and T. Scitovsky.

from it? Efforts to find a functional yet theoretically correct answer to this question focused on the development of an economic criterion for evaluating public expenditure alternatives. Given the prior attempts by water resources agencies to estimate the benefits and costs of alternative proposals, these early analytic efforts looked naturally to this program area.[2] Stimulated by the growth of the disciplines of operations research and systems analysis, further studies attempted to extend quantitative and systematic economic analysis to public expenditure decisions in all areas of the Federal budget. Through these efforts, appropriate methods for measuring the future benefits and costs of expenditures in a number of areas were developed, the correct economic criterion was determined, and the issues pertinent to the appropriate definition and size of the public discount rate were delineated.

One further factor influencing recent developments in public sector economics should be mentioned. In the 1950's and 1960's, another concern, quite unrelated to public finance matters, caught the interest of economists. During this period, the field of economic development and growth, focusing on the low-income problems of the less developed nations, advanced rapidly and became a primary area of economic study. In the search for meaningful instruments of national economic planning, the appropriate criterion for choice among public investments became much discussed. Those concerned with development policy required just those concepts formulated in the public expenditure literature—social benefits and costs, the public discount rate, the valuing of public outputs, and the form of the correct criterion. It was on these issues that the two quite separate economic concerns coalesced. Because of the convergence, progress in developing appropriate concepts and methods of public expenditure analysis occurred which would not otherwise have.[3]

Simultaneous with these developments in theoretical and applied economics and, in part, because of them, there was agitation within the Federal Government and among public administrators for improved knowledge on the economic effects of public expenditures. It became increasingly recognized that effective public choice required clear knowledge of the relationship of alternative programs to the attainment of social objectives. Concern with these matters was reinforced by real world political events. The 1950's and 1960's witnessed an enormous growth in the demands of citizens for government programs to produce a wide range of outputs which until that time had either not been produced or had been produced in the private sector. With demands placed on the public sector rising faster than the resources to meet these needs, many public decisionmakers actively sought an economic criterion to aid them in choosing efficiently among the available alternatives.

---

[2] See, for example, O. Eckstein, *Water Resource Development: The Economics of Project Evaluation* (Cambridge: Harvard University Press, 1958); R. McKean *Efficiency in Government Through Systems Analysis* (New York: John Wiley & Sons, 1958); J. V. Krutilla and O. Eckstein, *Multiple-Purpose River Development* (Baltimore: Johns Hopkins Press, 1958).

[3] While its impact is more difficult to trace, the private sector-public sector debate stimulated by J. K. Galbraith's *The Affluent Society* (Boston, Houghton-Mifflin Co., 1958) could also be added to the list of factors contributing to the increased concern by economists with public expenditures and their effects. Indeed, Galbraith's plea for a larger public sector was based on a judgment that the "outputs" generated by the next public dollar spent have a greater value than those created by the next dollar spent by private consumers or businesses. Both his critics and supporters recognized that additional knowledge of the benefits and costs of public expenditures was necessary to test both his assertion and its opposite.

Because of the influence of all of these events, the flow of economic analyses of public expenditure programs grew rapidly in the 1960's. Pioneering studies in the areas of natural resources, health, education, and pollution were published. In addition, largely under the sponsorship of the RAND Corporation, the techniques of systematic analysis were applied to the evaluation of national defense decisions and policies. In these studies, analysts attempted to determine the relevant policy alternatives in a public expenditure area, to estimate the real social costs and the real social gains to be expected from each, and to point decisionmakers to the optimum alternative or set of alternatives.

Where the conceptual problems of defining the output and establishing its value could be solved, analysts calculated ratios of benefits to costs for alternative programs or for the components of programs. If a project or program demonstrated a benefit-cost ratio which exceeded 1, there was a prima facie case that the undertaking was an economic one—that the social value of the output which it produced exceeded the social value of the inputs which it drew away from other uses. In those cases in which the program output was difficult to define or measure or where the output could not be valued, analysts undertook what is called cost-effectiveness analysis. In this form of analysis, the task becomes one of searching for the most effective or lowest cost means of attaining an explicit public sector objective, rather than an evaluation and comparison of the social value of benefits and the social value of costs. Benefit-cost analysis and its concepts have been applied to public expenditures for natural resources development and, to a more limited extent, transportation and human resources investments. On the other hand, cost-effectiveness analysis has been applied to investments in the national defense and space programs. In these areas, the good produced is a public good, meaning that once it is produced and made available, everyone automatically benefits. Quantitative estimation of the benefits of such expenditures is not possible and cost-effectiveness analysis becomes the appropriate instrument.

III

Recent efforts to introduce systematic analysis into the public decision process began in the Department of Defense. The story of its introduction there is by now well-known and documented. Under the leadership of Secretary McNamara and his Comptroller, Charles Hitch, an Office of Systems Analysis was established and headed by an Assistant Secretary. Cost-effectiveness studies were applied to major departmental decisions. Applying concepts developed at the RAND Corporation and similar research centers, the systems analysis staff was able to furnish the Secretary with the kind of performance and cost information essential for a rational consideration of and choice among alternatives.

The recognized success of this effort stimulated the judgment that a similar analytic approach could and should be applied also to decisions on civilian programs. This judgment was reinforced by the fact that expenditure projects in the water resources budget had been evaluated by benefit-cost measurements since the decade of the 1930's.

President Johnson, on August 25, 1965, incorporated this judgment into an Executive order establishing a comprehensive planning-programing-budgeting (PPB) system throughout the Federal Government. In his words, this system would enable public decisionmakers to—

(1) Identify our national goals with precision and on a continuing basis;

(2) Choose among those goals the ones that are most urgent;

(3) Search for alternative means of reaching those goals most effectively at the least cost;

(4) Inform ourselves not merely on next year's costs, but on the second, and third, and subsequent years' costs of our programs;

(5) Measure the performance of our programs to insure a dollar's worth of service for each dollar spent.

The Bureau of the Budget was given the responsibility for implementing this system in the executive branch and for assisting agencies in the development of methods of analysis and an analytic staff. Ideally, the PPB system was to be the vehicle for generating open, explicit, and comprehensive economic evaluation of all programs and expenditures and establishing a decision process in which choices would be based on the results of analysis.

It has been nearly 5 years since the inception of the PPB system. Many assertions have been made about both the wisdom and the efficacy of the effort. Many failures have been documented and well publicized. Successes have also been claimed. These have received far less attention. Those who fancy themselves "generalists" have viewed efforts to apply sound and systematic thinking about benefits and costs as a narrow "specialist" enterprise. Some, whose traditional power over the Federal budget requires the absence of both knowledge of economic impacts and decisions based on explicit analysis, have viewed it as a threat to their influence.

Advocates of the PPB system argue that the progress which the system has already made has been substantial, especially given the state of the art in applied economics. Decisions have been based on better information and analysis. Parties to the political bargaining process have had to focus on the appropriate concepts of inputs and outputs, gains and losses, benefits and costs. However, even the most enthusiastic admit that, for a number of reasons, substantial progress toward the goal of a rational, consistent, and economic budgetary and decision process has yet to be made.

From the outset, the PPB System has encountered serious obstacles which impeded improvements in the public decision process. Among the primary impediments which have been cited by observers of the system are the following:

- The failure of many agency heads to demand program analysis or to use it in decisionmaking when it was available;
- The lack of interest in (and sometimes opposition to) the system by important congressional committees and congressmen;
- The failure of much legislation to clearly stipulate program goals and objectives and to provide funds for the collection of followup data and other program appraisal information;

- The existence of private interest groups which anticipate that hard and quantitative program evaluation will endanger the size or existence of expenditures which benefit them;
- The constraints on substantive and time-consuming policy analysis imposed by the annual budget cycle and process to which the PPB System is tied;
- A serious scarcity of analytical personnel in the PPB offices of civilian agencies;
- A basic resistance by many Federal employees to economic analysis and the difficult job of program evaluation;
- The lack of professional agreement on certain basic analytical issues, such as the appropriate public interest rate for discounting long-lived public investments, the development of shadow prices when outputs are not marketed, the evaluation of expenditures with multiple objectives, and the evaluation of public expenditures in regions or periods of less than full employment;
- The lack of adequate data from which to develop measures of the social benefits of outputs and social costs of inputs.

From the beginning, the Joint Economic Committee has supported the goals of the PPB system and the efforts of the executive branch in implementing the comprehensive economic analysis of public expenditures. Even prior to the inauguration of systematic analysis into the Department of Defense, the Joint Economic Committee stated its concern with the budgetary process, the budget document, and the analysis of budgets and programs.[4] Since the establishment of the system in 1965, the committee has followed its development with interest. Two years ago, the progress and potentials of the system were appraised by the Subcommittee on Economy in Government in both hearings and a report.[5] Within the past year, this subcommittee studied and issued a report on the appropriate discount rate for evaluating public investments.[6] The set of background papers contained in the present volumes represents the most recent effort of the Joint Economic Committee to stimulate improved economic analysis of public expenditure programs. It is appropriate as a new administration assumes control, and as the PPB system begins its next stage of development, to take a deep and comprehensive look at public expenditure policy and the role of analysis in improving program design and budgetary allocation.

IV

The background papers presented in these volumes form a most comprehensive study of public expenditure economics. The theory of public expenditure analysis is surveyed. Current Federal Government practice of evaluating program alternatives is described and critiqued. Conceptual problems in applying economic analysis to public expenditures are discussed and the institutional factors which affect the Gov-

[4] U.S. Congress, Joint Economic Committee, "Federal Expenditure Policies for Economic Growth and Stability," report of the Subcommittee on Fiscal Policy, Jan. 23, 1958, and U.S. Congress, Joint Economic Committee, "The Federal Budget as an Economic Document," report of the Joint Economic Committee, Aug. 14, 1963.

[5] U.S. Congress, Joint Economic Committee, "The Planning-Programing-Budgeting System: Progress and Potentials," report of the Subcommittee on Economy in Government, Dec. 4, 1967.

[6] U.S. Congress, Joint Economic Committee, "Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis," report of the Subcommittee on Economy in Government, Sept. 23, 1968.

ernment's ability to implement effective public expenditure decisions are described.

In these papers, the objective of economic efficiency in the allocation of national resources receives primary emphasis.[7] The basic presumption of nearly all of the authors is that the level of national economic welfare can be increased if Congress and executive policymakers increase the role of efficiency-oriented analysis in making public expenditure decisions. While the efficiency objective is primary, it should be emphasized that the papers do not reflect a strictly economic point of view. Institutional factors, problems of multiple-governmental objectives, issues of organizational structure, and the unique characteristics of decisionmaking in each of the major functional areas of the budget are all discussed in these papers.

This study has a number of objectives in addition to its general goal of improving the effectiveness of public expenditure policy. A principal objective is to focus the attention of the Congress and the new administration on the need for quantitative benefit-cost-type analysis in the formation of Federal expenditure decisions. Objective analysis, especially within the framework of the PPB system, has not always been welcomed in either the executive or legislative branch. In part, this response reflects the failure of decisionmakers to understand how such analysis can lead to more productive public expenditures and an improved allocation of national resources. It is hoped that these papers and the hearings which are to follow them will go far toward increasing the understanding of and appreciation for the improvements in decisionmaking and policy implementation that PPB-type analysis can bring about. In appraising the role of public expenditure analysis in its policy planning and implementation, the new administration can do no better than to listen to the critique of the past and the proposals for the future presented by the contributors to this study.

A second objective of this study is to stimulate economists and other social scientists to undertake those efforts necessary to increase the efficacy of economic analysis in public sector decisions. These efforts entail basic and, especially, applied research. The literature of public expenditure economics is now relatively rich in theoretical insights. However, the literature which develops empirical methods for estimating social costs and gains and which applies welfare economics to public policy decisions is a good deal thinner. Moreover, attempts to increase the role of PPB analysis in public decisionmaking have been constrained by the lack of appropriately trained people. The difficulties of establishing a new area of study in traditional graduate and undergraduate curriculums have impeded attempts to reduce this scarcity of trained personnel.

In addition to public decisionmakers then, this study is addressed to economists, systems analysts, and other social scientists in universities and research organizations. It is designed to emphasize the importance of developing improved methods of public expenditure analysis, evolving a consensus on the appropriate methods of valuing public sector outputs, and establishing programs to train people to apply these analytic procedures.

A final objective of this study is to assist those in government whose

[7] In the economics literature, efficiency in the allocation of resources is directly related to the level of economic welfare. If resources are allocated efficiently, they are placed in those uses in which the value of their output is the greatest possible. It follows that, if all resources are allocated efficiently, the social value of the Nation's output is maximized.

job it is to do program planning and policy analysis. The papers dealing with analytic problems and those containing suggestions for next steps in applying policy analysis should provide both conceptual and methodological guidance for those analyzing the social worth of public alternatives. The study attempts to bridge the gaps between the theory of public choice and the application of welfare economics to public expenditure decisions. It reflects the judgment that the implementation of public expenditure policy can be more effective if guided by the results of sound and quantitative economic analysis of available options.

V

The papers collected in these volumes span the range of issues in public expenditure economics. Volume 1 contains parts I, II, and III of the study. The issues discussed in this volume extend from abstract, conceptual matters of collective decisionmaking to the institutional factors which constrain effective policy analysis to the troublesome analytical issues which hinder attempts to empirically measure the economic impacts of public expenditures. Volume 2 (Part IV) presents a comprehensive description of the structure and functioning of the PPB system. It analyzes the current status of the system and discusses its future prospects. Volume 3 of the study, which contains parts V and VI, deals in greater depth with the problems of developing an explicit and open public expenditure decision process in the context of the planning-programing-budgeting system. There, the performance of the system to date is critiqued and some lessons learned from experience with the system are presented. A number of papers offer recommendations on the next steps to be taken in implementing PPB analysis in the several functional areas of the Federal budget.

Part I of the collection explores the economic basis for public expenditures and other governmental action. The first two papers survey the basic issues involved in determining the appropriate economic functions of government, and the optimal division of national resources between the public and private sectors. This overview is followed by five papers which deal with the structure and performance of the private or market sector of the economy and the efficiency basis for collective action. They analyze the role of external effects (spillovers), decreasing costs, lack of market information and mobility, uncertainty, and contracting costs in impeding the effective performance of the private economy. Where these conditions are present, the market system fails to generate an efficient allocation of resources. These papers discuss the need for collective action when market failure is present and detail the alternative public sector responses for adjusting the private sector outcome. The final paper in part I deals with the equity basis of government's economic role. It discusses the public sector's responsibility for adjusting the distribution of society's income and analyzes a number of criteria for evaluating attempts to achieve this redistribution through public expenditure policy.

The papers included in part II focus on the prevailing institutional arrangements which affect the ability of the Federal Government to implement appropriate and effective public expenditure policy. The first paper in part II presents a broad and far-ranging discussion of this subject. It argues that sound public policymaking requires careful and explicit consideration of the incentives, penalties, and rewards which exist in the economy. Because these institutional factors influence

the ability of any program to attain its objectives, sound policymaking must take account of the structure of incentives. The remaining papers of part II examine a number of these penalty-reward matters, as well as other institutional considerations which surround public expenditure policy. Among the specific issues analyzed are (1) the role of user charges and other beneficiary cost-sharing arrangements; (2) the complexities entailed by a Federal system in which various levels of government have overlapping jurisdictional responsibility; (3) the obtacles to effective policy implementation imposed by the structure of government and the inflexibility of the budget; and (4) the lack of economic data pertinent to the effective planning of public expenditures.

In part III, several of the most basic analytical and empirical problems encountered in estimating the benefits and costs of various public expenditure alternatives are analyzed. These include the valuation of the social benefits and costs of public expenditures when there are no observable market values or when the Government has objectives in addition to economic efficiency and can use instruments other than expenditures. Additional papers deal with the appropriate Government response to risks and uncertainties which surround their decisions, the correct concept and size of the public rate of discount, and the effect on benefit-cost analysis when there is unemployment in the private economy. The papers in this part analyze both the nature and the causes of these problems of analysis and, in most cases, offer recommendations for improving the practice of Federal Government expenditure evaluation.

Part IV (vol. 2) contains a comprehensive evaluation of the planning-programing-budgeting system and its component parts from the perspective of the U.S. Bureau of the Budget. In this paper, the Assistant Director of the Bureau of the Budget for Program Evaluation presents the current position of the Bureau of the status of the PPB system and its plans for improving the system's structure and performance.

The papers in part V critique the performance of the PPB system in increasing the effectiveness of public expenditure policy and offer suggestions for modification in both the structure of the system and the operation of both the executive and legislative branches. These papers have been prepared either by individuals who have been intimately involved in directing and implementing the PPB system or by persons outside of government who have closely observed the development and performance of the system. Those papers which present "lessons from experience" deal with questions of both the organizational structure of the system and the institutions of government.

The papers collected in part VI deal with the unresolved issues and next steps for policy analysis and program evaluation in the major functional areas of the Federal budget. Nearly all of them have been prepared by nongovernment analysts who are knowledgeable in the structure and implementation of policy and policy analysis in these areas. It is the explicit objective of these papers to place before the new administration the insights and recommendations of an outside expert in each of these areas. It is hoped that these papers will be seriously studied by each of the pertinent agencies and that the next stage in the evolution of the PPB system will be guided by the issues which they raise and the recommendations which they contain.

Part I

The Appropriate Functions of Government in an Enterprise System

Section A

The Public Sector and the Public Interest: A Discussion of Basic Concepts

27-877—69—vol. 1——3

THE PUBLIC SECTOR AND THE PUBLIC INTEREST*

BY PETER O. STEINER

Peter O. Steiner is Professor of Economics and Law at the University of Michigan.

In this paper, Professor Steiner addresses the most basic questions pertinent to discussion of the appropriate role of the public sector. In a society in which most goods and services are provided for in a market system based on individual choices, what should be the role of collective action? "It is necessary to ask what it is that persuades or requires members of a group to seek a collective solution to some problem rather than to rely solely on individual action. It is also necessary to ask whether collective desires merit public support, public indifference, or public hostility." In the discussion of these issues, the economist's notion and definition of "public goods" is criticized and the (primarily economic) reasons for reserving certain activities to the public sector are classified.

Following the discussion of public (or collectively provided) goods, Professor Steiner discusses alternative views of the relationship between individual values and the legitimate demand for governmental (or collective) action. He attempts to define the amorphous concept of the "public interest." Also discussed are a number of views concerning the process by which the political system articulates the "public interest." He concludes his paper by arguing that there is an important role for economic analysis in public decisionmaking. Moreover, he urges that the decisionmaking process be a more open one, involving the explicit statement of objectives and open identification and consideration of those values pertinent to a decision. "I think at present that we conceal so many issues and conflicts, both among objectives and among alternative means, that we increase the discretion of the policymaker beyond that necessary or desirable."

Introduction

If one starts at any point and place in history—say the United States in 1969—it is clear that the society has decided that there exist certain activities that are legitimately performed by governments. Many activities are by long tradition provided by various levels of government and are paid for by using the police powers of the state to raise funds. Others are left to the private sector. Without wishing to disparage the importance of the debate about the proper dividing line between private and public sectors, the fact is there is a large, relatively stable and broadly uncontroversial governmental "sector" of this economy, and of every other economy in the world.

In order to focus on certain critical issues I shall suppress some real distinctions and create some arbitrary ones. The most important simplification is to treat "government" as a single cohesive force, thus neglecting intergovernmental transfers as well as conflicts of authority and philosophy among Federal, State, and local governments. The

*This paper is a modification and abridgement of a larger study to be published by The Brookings Institution. [43] Research underlying that study has been supported by The Brookings Institution, by the Graduate School of the University of Wisconsin, and by the Cook Foundation of the University of Michigan Law School.

most important complication is to pretend that the theory of public expenditure policy is in reality two very different sets of propositions. One of these may be called the theory of the marginal public expenditure. It takes as given the legitimacy of government activity, and is concerned with choicemaking of the public decisionmaker between competing demands for his limited resources. The other set of propositions may be called the theory of the public interest, and concerns the way in which demands for public activity arise, are articulated, and are legitimatized. It is this latter that I wish to discuss in this paper.

This is a separation of convenience, not of fact. For example, every marginal decision to expand some public program into a new area implies a legitimate public purpose in that area; thus proposed discrete extensions often pose the questions of public interest. Similarly, making a marginal decision requires rationally knowing what aspects of the public interest are being served.

There is one decisive reason for treating the theories separately. It is, simply, that as of this date the available theories are of very different levels of adequacy in the two cases. For the marginal decision a well-developed, highly articulated and largely uncontroversial set of theories exists and awaits implementation into practice. In contrast, with respect to the nature of the public interest, it is we theorists who are the primitives in the sophisticated world of public decisionmakers. It is the theorists who know how to choose between two public housing proposals but not whether public housing is right and proper; while the bureaucrats and Senators have less difficulty deciding when public housing is required than in choosing between alternative schemes of public housing.

Definition of the public interest is genuinely difficult because the notion embodies at least two implicit distinctions. One is between collective action and individual action, the second between public (that is, governmental) action and private action. Each is important. It is necessary to ask what it is that persuades or requires members of a group to seek a collective solution to some problem rather than to rely solely on individual action. It is also necessary to ask whether collective desires merit public support, public indifference, or public hostility. Finally, if collective desires are in some sense legitimatized the question remains as to what form of collective action is to be chosen. There is no simple dichotomy between individual private activities and collective public action. Instead there are various kinds of collectivities—clubs, unions, churches, political parties, as well as governments; various degrees of public involvement from outright prohibition of certain activities, to taxes or subsidies, to direct public provision of services.

The *desire* for collective action, which underlies many demands for public provision of goods and services, may arise any time a group feels it cannot achieve its objectives unaided. But mere demands (however genuine) are not enough. Aid to the needy aged (or unattractive prostitutes) may be the only effective device by which this portion of the population may be assured a subsistence level of living; price supports may provide farmers (or retail grocers, or racetrack touts) with protection against excessive competition that they could not achieve without Government action; a program to place a man on the moon by 1970 (or to build a tunnel under Lake Michigan, or to commit geno-

cide) can be visualized only as a collective program. Each of these activities transcends individual solution and thus requires collective solution or no solution, but that per se does not render them legitimate activities of Government. Most will find some of these proposed activities meritorious but some objectionable as spheres of public action.

Moreover, the required use of collective action is not only not sufficient to define public activities, it is not strictly *necessary*. We may have governments provide education, housing, transportation, and recreation even though private alternatives exist. Such a choice might rest upon considerations of efficiency but it might also reflect captious preference or even prejudice. What leads men to choose public provision from among alternative means of meeting particular ends?

Casual observation suggests that the public interest may be served by providing or encouraging provision of a variety of goods or services, and by nonprovision or discouragement of others. Let me loosely define these goods and services as vested with the public interest, or as *public goods*. Let us first look more closely at the nature of public goods.

A. The Nature of Public Goods

Serving the public interest may take many forms, among them providing of goods, subsidizing their private provision, passing laws that impede or prohibit their provision or constrain the form in which they are provided. Because of the focus of this compendium, I shall limit attention to policies that involve public expenditures. But it should be remembered that an important policy issue always concerns choice among alternative available means.

The goods and services provided by public expenditures or encouraged by public policies can be described and classified in a number of different ways. Though we speak of a single category, "public goods," any review of actual public policies shows great diversity and variety. Some kinds of public goods are provided publicly or not at all because there exist no reasonable private alternative ways of providing them. This can happen (as in the case of national parks, national defense, or space research) because there exists no private mechanism to pay for the provision of these goods, or it can happen (as in the case of sewage disposal or justice) because compulsory use of the goods by all is required to permit its enjoyment by any group. Other public goods, such as public housing or public education, may be functionally similar to available private alternatives, but be qualitatively different in ways that society somehow prefers. Still other public goods may differ from private ones in no way other than in the distribution of beneficiaries and costs.

If the proper domain of public expenditure policy is public goods, their definition becomes vital. The concept has been defined in many ways, and for diverse purposes, and it is not surprising that definitions motivated by purposes other than ours—understanding the rationale and process of public expenditure policy—are not wholly satisfactory.

"A *public good* is any good or service which is *de facto* provided for or subsidized through Government budget finance. (Birdsall [5]

p. 235).[1] This definition is neat but unhelpful. It is deficient in that it provides no guidance as to what attributes of a potential good or service a policymaker should look to in deciding whether to provide the goods. In this definition "publicness" is an act of congressional designation, not of any characteristics of the good or service.

In many ways, it would be desirable to have an intrinsic definition based upon technical characteristics of goods or services. One such definition of public goods is of the perfect collective consumption good. An impressive array of economists have so defined public goods including Samuelson [39], Strotz [45], Bowen [7], Breton [9], and most recently Dorfman [14]. Hear Dorfman:

> "There are certain goods that have the peculiarity that once they are available no one can be precluded from enjoying them whether he contributed to their provision or not. These are the public goods. Law and order is an example, and there are many others too familiar to make further exemplification worth while. Their essential characteristic is that they are enjoyed but not consumed [and that their benefits are derived] without any act of appropriation." ([14], p. 4).

This kind of very narrow definition was designed to demonstrate that there may be a type of activity that is socially desirable but that will not be achieved by the unaided private market. It serves well the purpose of showing the existence of public goods. It can prove a hindrance, however, if it leads to the view that such goods are *the only* class of goods which Government can legitimately provide. In fact, examples are hard to find, not ubiquitous, and the great bulk of (nondefense) public expenditures are for goods and services that do not meet the definition. Roads, schools, welfare payments, recreational facilities, housing, public power, irrigation (among others) are important classes of public expenditures that some can be precluded from enjoying, that can be consumed in whole or in part, and that technically can be made subject to user charges. The perfect collective consumption good, while sufficient to justify public expenditure, is neither necessary nor does it embrace much of what public expenditure policy concerns. What it does do is to identify certain characteristics such as nonconsumption, nonappropriation, and the existence of externalities that may give a good public goods aspect.

Externalities are very important, as has been recognized for a long time. Wicksell (himself citing earlier authority) put it eloquently in 1896:

> "If the community or at any rate a sizable part of it has an interest in a particular utility accruing to an individual, then it would clearly be unreasonable to allow the creation of that more general utility to depend solely upon that individual: he might not value the state activity highly enough to make the sacrifice of paying the required fee or charge, or else ignorance may cause him or poverty force him to do without the service. Herein lies the chief justification of the modern demands for free or very cheap process of law, elementary education, and medical care, certain public health measures, and so forth." [2]

---

[1] Similar is Musgrave's definition of public goods as those produced under public management [33], p. 42. It should be noted that Musgrave recognized the need for a more complex classification, which he provides.

[2] Knut Wicksell, "A New Principle of Just Taxation." Reprinted in [34].

While they are important, it would be easy to follow externalities too far. Few goods fail to have some elements of externality in the sense of some benefits or costs that do not require an act of appropriation. But if few goods fail to meet this test it cannot provide guidance as to which goods ought to be candidates for public provision, nor to explain which goods are publicly provided.

Let me venture a definition of my own: *any publicly induced or provided collective good* is a public good. A "collective good" in my definition is not necessarily a collective consumption good. Collective goods arise whenever some segment of the public collectively wants and is prepared to pay for a different bundle of goods and services than the unhampered market will produce. A collective good thus requires (1) an appreciable difference in either quantity or quality between it and the alternative the private market would produce, and (2) a viable demand for the difference.

Collective goods may be privately or publicly provided. Co-ops, unions, vigilante organizations, country clubs, carpools, and trade associations are all examples of private organizations that arise in response to collective demands for private collective goods or services. When the coordinating mechanism for providing a collective good invokes the powers of the state I define the good as a public good. In this definition there is the requirement that a public good must meet the tests of being a collective good. Public provision by itself does not create public goods. This definition is virtually implicit in the discussions of Head [20], Musgrave [33], Olson [35], Weisbrod [50] and Margolis [30]. It provides something of a framework for considering various sources of public goods, as we shall see just below.

A most important aspect of this definition is that it makes "publicness" not an all-or-nothing attribute of a good, but an attribute that may apply merely to particular aspects of a good. While there are cases (for example, national defense) in which the choice is between public provision and no provision, and why we thus argue that the good is entirely a public good, the more common situation is for goods to provide a variety of services only some of which have the attributes of collective goods. When the aspects that are thus of collective interest become sufficiently important, we may be led to public provision either of those aspects, or of the entire good (including its noncollective aspects). Thus provision of smog control or river pollution attacks a particular externality of private production. In contrast public housing provides individuals with services they would otherwise have purchased privately, along with the peculiarly public services that public housing is supposed to entail.

Such mixed goods test and stretch definitions. "Public education" and "public housing" reflect both quantitative and qualitative differences from the privately produced or producible goods, "private housing" and "private education." If the differences are intended and desired, the differences constitute public goods in my definition.

This somewhat vague and embracing notion of public goods can be filled out by a more detailed classification of different types of public goods.

B. A Classification of Public Goods

I have (in effect) defined the vector of public goods as a vector of differences between the goods and services the private economy is motivated to provide and the goods and services the "public" wants, is willing to pay for, and expects its government to assist it in achieving.[3] This is, to an important degree, a normative definition, and much of the debate about the appropriate elements of the public goods vector is a normative debate. But there is a positive aspect as well: What is it about particular goods and services that makes them candidates for public consideration? What is it that makes certain activities the traditional province of governments?

It seems worth while to distinguish three types of public goods: (1) Those arising from intrinsic (perhaps technical) characteristics of specific goods that result in externalities that are not effectively marketed; (2) those arising from imperfections in the market mechanisms (rather than in the nature of the goods or services themselves); (3) those arising not from specific goods or services but from aspects concerning the quality or nature of the environment. These become increasingly elusive as we proceed from (1) to (3), but it is impossible to capture the flavor of actual government expenditure programs without all of them.

1. Public goods arising from nonmarketable services of particular goods*

A discrepancy between public wants and private supplies often arises from externalities (or as they are alternatively called, spillover or third-party effects). Any time provision of a good or service provides side effects whose value is not reflected in the prices of the outputs sold or the resources used, external economies or diseconomies are produced. There can be many reasons for such externalities: Private producers may use resources they do not consider scarce, or produce byproducts that they do not consider valuable because they cannot control and market them. Familiar examples are discharges of noxious wastes into water or air, downstream navigational or flood control consequences of a private power dam, civic beautification or uglification incident upon building of private golf courses, factories, or slaughterhouses. Because some of the resources used or outputs produced are not correctly valued by the market there is every reason to expect the market to misuse them. Thus collective concern and public action may be required on simple efficiency grounds to allocate resources in accord with "true" valuations.

Whether the existence of such externalities (which must surely be present to some extent in every good) justifies public notice and action depends upon the benefits to be achieved measured against the costs of interference. Different people will have divergent views of the costs of interference but debate as to what the cutoff level should be is a different matter than debates about the nature and size of the externalities.

---

[3] The identity of who the "public" is, is deferred for the moment.

*Further discussion of this issue is found in the papers by Arrow, Davis & Kamien, and Kneese & d'Arge in this volume.

The perfect collective consumption good is really an extreme case of externalities: all of the output is regarded as individually unmarketable; all of the benefits are external. The outputs of those goods from which one cannot be excluded as a consumer—and thus for which one cannot be compelled to pay his share of the cost of provision—play a large role in the thinking of those who have been concerned with deriving a legitimate role for public activity. Defense, public health, law and order, hurricane watches, are familiar examples. It is common to list a few examples (and not press them very hard) and then say, "there are many other examples." This is close to fraudulent. If excludability implies no one can conceivably be excluded, the list of such goods is short indeed. One need not police the ghetto, nor defend Alaska. Television signals can be scrambled so as to exclude buyers who will not pay for the unscrambler. Movies, concerts, hospitals, and colleges all use walls to exclude those who will not meet the requirements placed upon their use.

Collective goods may arise because it would be relatively costly to exclude free riders rather than because it is impossible. If at any moment this cost is above a certain level there may be no effective supply of the good privately provided. But in other cases the cost of exclusion may be annoying rather than prohibitive and potential consumers may urge public action merely to avoid bearing the costs. Put differently, the cost of arranging exclusion may be an avoidable externality.

Implicit in this discussion is an important attribute of the public collective good: the willingness to appeal to the police power of the state. One can slide in imperceptible steps from situations where there is no viable alternative means of providing the good, to cases where the alternative seems unnecessarily costly, to cases where the alternative while not very costly is simply judged to be less desirable, to cases where the alternative differs only in who pays for it.

There is real purpose in downgrading the distinction between *inability* and *unwillingness* to provide a good privately.* If there is to be a practical definition of specific collective consumption goods and services, it seems difficult to escape the view that a judgment is required about reasons for turning to the political process and the coercive power of the state, rather than dealing with the second best solution. These reasons must be judged meritorious by the social decision process. *If this is so, collective consumption goods are defined by the exercise of legitimate governmental decision processes as much as they define them.*

Among the positive issues that underlie the normative debate about whether a particular collective good ought to be publicly provided are (1) whether private market alternatives to public provision are impossible, impractical, merely costly or simply unwanted; (2) why the market solution is unsatisfactory to members of the group and to society as a whole; and (3) the identity of the group of beneficiaries. The last deserves a bit of comment.

A collective good need not provide joint benefits to all of a society's members, only to some subgroup. But which group? The larger the

*Further discussion of this issue is found in the paper by Krutilla in this volume.

group the more persuasive its demand for public action is likely to be, or (put differently) the less willing will its members be to accept a costly alternative. There are bases other than size for weighing the merits of the demands of any group, and these may vary over time. Domestic producers of good subject to foreign competition, farmers, labor unions, small businessmen, and minority groups are among the identifiable groups that have asked and received special treatment. Today, for example, we seem more responsive to the demands of the underprivileged than the wealthy; a half century ago it was clearly otherwise.

One reason many collective consumption goods lead to demands for public provision is because the potential willingness to pay of different consumers cannot be tapped by private suppliers. Weisbrod [50], in an important paper, suggests a further source of values for which there is no market: option demands.[4] Some examples: I value Yellowstone Park being there, though I hope I never have to visit it again; I value the Everglades because I may want to visit them (but probably will not); I value the existence of a first-rate tuberculosis sanatorium, through in all probability I shall never need its services. Were any of these threatened with extinction I should be the loser, but there is no market whereby my willingness to pay for the option of being able to use them can be translated into revenue to the providers.[5]

Weisbrod's most suggestive example concerns the standby availability of transport. How much is it worth to the New York-Washington air travelers to have a good rail alternative in case of snow or strike? Suppose it is worth enough to justify the rail service, but that the railroad has no way of being reimbursed by those whose option demands are critical to continuation of the services. In these circumstances, the public good may be provided by the government's insistence that the railroad's passenger service be maintained with or without subsidy. In this view governments may not have been irrational in trying to preserve passenger train service even in the face of inability of the carriers to find user charges equal to costs.

2. PUBLIC GOODS ARISING FROM MARKET IMPERFECTIONS

There can be in practice no sharp distinction between market failure caused by technical characteristics of particular goods, and market failure caused by market imperfections. Inability to handle externalities, for example, may be regarded as a shortcoming of existing markets rather than as the absence of markets for specific services. But a distinction suggests additional sources of unsatisfactory private market performance that generate demand for public collective action. Efficient markets frequently suppose adequate information, sufficient competition, timely adjustment, and modest transaction costs. The absence of any of these may create motives to replace market determination by nonmarket provision, or to supplement markets with ancillary public goods.

---

[4] Millard F. Long [24] has recently challenged Weisbrod's concept.

[5] Option demands are in a sense much like consumers surplus: they arise because the price charged for the good or service is below the maximum each buyer would be willing to pay. Thus the option to buy at a low future price has present value. Weisbrod's insight, I think is that option demands are a significant source of demand for public action. [See aslo the paper by Zeckhauser, in this volume.]

*a. Information**

Suppose all conditions for ideal resource allocation are satisfied except that market signals are systematically not read or are misperceived by economic actors. An allocation of goods and resources will occur, but it will, in general, differ from the allocation that would occur without ignorance or misinformation. Information may be a collective good (and thus generate a demand for its public provision) because even if there is a well-articulated private desire to have information, there may be no effective market in which to buy it efficiently.[6] It may be a public good in addition because the externalities of misinformed traders may be judged to be socially undesirable.

b. Time lags

If resources respond to market signals surely but slowly, the market process may prove an expensive way to achieve resource shifts. If physicists are in short supply, their price may be expected to rise and this may motivate additional youngsters to undertake education leading to careers as physicists. Since education is a slow process, existing physicists may earn high rents over long periods due to the long supply lags. It may well be that public policy can increase the supply of physicists more quickly and more cheaply by fellowships, by research grants, etc.[7] than the unaided market. If increases in supply of physicists, but not increased incomes of existing physicists, are desired results, then such programs supply public goods.

There is a large and growing literature that is concerned with the extent and causes of factor immobility. Education is but one of the sources; others include unemployment rates, prejudice, institutional barriers such as seniority and pension laws and State laws affecting eligibility for relief. Whenever markets work to reallocate resources sufficiently slowly, there may be a collective demand to supplement the market mechanism, or to replace it. Retraining programs, moving allowances, public employment services and even attacks on prejudice may be public goods if they serve to reduce the lags that the market economy accepts.

*c. Monopoly power***

Noncompetitive imperfections need little comment here. Public activities to encourage or compel competitive behavior, or to replace monopolistic private by public provisions are further sources of public goods.

*d. Transaction costs****

We have seen that an important aspect of collective goods concerns the inability of the market to translate potential willingness to pay into revenues. Related is the situation where the private market is technically able to collect revenues, but at a high cost. Collection of

---

[6] Stigler, [44] provides a conceptual analysis of the costs and benefits of obtaining information. Telser [46] deals with the problem of buying information in the form of advertising as a joint product with news, entertainment, etc.

[7] A Department of Health, Education, and Welfare study, [49] supplies some evidence on the incentive effects of subsidies to scientists and other academic personnel.

*Further discussion of this issue is found in the paper by Davis & Kamien in this volume.

**Further discussion of this issue is found in the paper by Vickrey in this volume.

***Further discussion of this issue is found in the paper by Demsetz in this volume.

user tolls on interurban roads and urban bridges may or may not be both feasible and efficient, but intraurban toll roads would surely involve even larger collection costs and time losses. Prohibitive transaction costs of collecting revenues for intraurban toll roads make high-speed roads of that kind public goods. Metering costs may be justified for high unit value commodities such as gas and electricity, but not for sewage (and, in some high population density areas, for water).

Where these high transaction costs inhere in the particular service they are simply an externality; where they reflect the institutional arrangements of the market they are a potential additional source of collective concern. The higher cost of attempting to gear a pricing system to an individual's willingness to pay is a repeated source of turning away from the market. For many goods, willingness to pay may be regarded as rising at least proportionally with income. Most private services are not provided on a basis that reflects income, because of the enormous costs of administration that such pricing would entail. If such a basis of payment is appropriate, reliance on the income tax, and thus on State provision, may appear desirable.

3. PUBLIC GOODS ARISING BECAUSE OF CONCERN WITH THE QUALITY OF THE ENVIRONMENT

To this point public goods have been discussed in terms of market failure; failure either because of the absence or the imperfections of private markets. This is the grand tradition of classical economics. But even perfectly functioning markets for *all* goods and services would not eliminate the desire for market interference. Men may choose to reject the market solutions to allocative problems with respect to the distribution of income, the nature or quality of goods produced, or the patterns of consumption that markets produce.

The most compelling examples of collective public goods have always seemed to me to be national defense, law and order, and public health. What is their particular appeal? Is it that they are collective consumption goods? So is television. It is not in the specific planes, rockets, soldiers, policemen, vaccines, or nurses that are their elements—each of which can be readily provided as private goods to private users—but rather in the fact that they are part of and condition the *environment* of the society. Even the criminal who detests the legal framework is affected by it. Looked at this way they suggest other things that affect the environment and thus create externalities not linked to particular goods: for example, the literacy rate, the level of unemployment, the crime rate, the rate of technological progress, and importantly, the pattern of distribution of income and wealth.

a. Distribution of income [8]

Accept this assertion: it is fully feasible to charge users for use of parks and playgrounds, to charge parents for school bus service and school lunches, to charge fishermen for fishing privileges. Suppose in each of these cases that there is sufficient willingness to pay and ability to collect so as to assure private provision of parks, play-

[8] See Freeman [17] for a recent effort to work out the implications of income distribution for public investment planning. [See also the papers by Weisbrod, Freeman, and Bonnen in this volume.]

grounds, school buses, school lunches, and fishing opportunities. Should these functions be left to private provision?

To answer this question we must decide whether we are concerned merely with the distribution of income or instead with the pattern of consumption. When we provide subsidized public housing to the urban poor do we desire to provide more or better housing to users who would be excluded by private provision (or who would exclude themselves) or do we simply choose to increase their share of national consumption and use public housing instead of a cash income supplement for some obscure reason? (One might use indirect means, for example, in order not to impair the feelings of self-respect of the recipients.)

It is sometimes argued that purely redistributional objectives which reflect a dissatisfaction with the initial situation of ownership of wealth and resources ought to be satisfied by income transfers rather than by provision of goods and services in order not to distort resource allocation. This familiar argument is unpersuasive if one regards as legitimate a desire of a society to interfere with the *pattern of consumption* that would result from market determinations. A society may choose to affect jointly both income distribution and the pattern of consumption. Provision of housing, education, milk, and recreation for underprivileged children may be public goods because of the externalities which children so treated bestow upon others. Public policies designed to provide aid for small business, for the family farm, for the needy aged, and for the slum child all reflect rejection of market determination, rather than denial of the possibility of market determination.

It is, of course, not clear that all actual interferences reflect a positive intention both to redistribute income *and* to change consumption patterns. In the United Kingdom (by way of contrast to the United States), fishing rights are sold, and it is an upper class form of recreation. On the other hand, virtually all Scottish golf courses are subsidized municipal ones, and in Scotland golf (but not fishing) is a working class recreation. But if some consumption distortions are fortuitous, others are intended.

b. Nature and quality of output

For some goods and services the quality and nature of the goods produced is of public concern, quite independent of any distributional considerations. Often the nature of the goods or service is affected by who provides it—for better or for worse. Government newspapers differ from private ones, public television and radio from commercial broadcasting, a system of public schools from a private school system, private research and development from public. In some of these examples both kinds of goods may coexist, in others an exclusive choice is made. But in all cases a choice among qualitatively different outputs may and can be made; the qualitative difference of public from private constitutes a public good or a public bad.

4. PUBLIC GOODS: A SUMMARY VIEW

I have stressed the pluralistic nature of the sources of collective demands, as arising from technical characteristics of particular goods, from market imperfections and failures, and also from other diver-

gences between collective and individual values. This time is long since past when we need to define public goods merely in order to establish the prima facie case for some public interference with private markets; instead we seek a framework to structure debate about whether particular activities merit inclusion in the public sector.

It seems to me worth identifying in each case what is the alleged source of collective concern. Does it depend upon a major qualitative difference between public and private provision, or does it merely seek some incremental output, perhaps by considering a particular neglected externality? In this distinction often lies an important policy choice between public provision on the one hand or a less fundamental public restructuring of private incentives. Similarly one wants an indication of whether public concern is specific to the particular good or service or is of a general environmental type. There are more ways to reduce overall unemployment than there are ways to retrain Appalachian miners. Again the relevant alternatives are affected by what are the real objects of policy. A frequent issue concerns whether redistributional policies achieved by provision of specific goods and services are intended to bring about changes in consumption patterns or do so incidentally.

Next, having established the basis of *collective* concern, it is worth establishing the basis of *public* concern. Who are the alleged beneficiaries, and what is their claim to recognition? What is the "second best" that they face if their claims are rejected?

I believe that defining as specifically as one can the vector of differences between a private good and its public alternative to be a critical part of the public decisionmaking process. Neither de facto definitions (such as Birdsall's), nor neat but narrow ones (such as that of the perfect collective consumption good) prove very helpful for the critical problem of defining the scope of the public sector.

C. Sources of a Public Interest: Alternative Views

To convert a collective interest of some group into the public interest requires a distinct act of legitimization. How does it occur? Views differ both with respect to what it is that is aggregated, and with reference to the requisite degree of consensus.

My dominant reaction to rereading the discussion among economists about the "public interest" is surprise at its defensive tone, as if we are somehow disloyal when we find a role for extramarket forces in the economy. Perhaps because economists have felt defensive, much of the economic discussion has revolved not around the issue of how to define the public interest, but rather how to demonstrate that there is a de minimis role of government activity that clearly benefits everybody. Much of welfare economics consists of such a possibility theorem. Possibility theorems are fine in their own way. If, however, they are misinterpreted they may greatly limit the scope of the phenomenon whose possibility they are concerned to establish. This has, I believe, been the problem and the fate of formal welfare economics. To document this view or to discuss the alternatives in any detail involves technical issues best avoided in this paper.[9] But it is possible to identify a variety of different points of view and the different implications that they may lead to.

[9] They are discussed in [43].

1. THE POINT OF VIEW OF INDIVIDUAL UTILITY

Those who hold this view consider the public interest of a society as being simply an appropriate aggregate of the private interests of the individuals who comprise the society. Each individual is assumed to seek his own utility (or satisfaction) by pursuing all avenues open to him. Assume as a rough distinction that he draws satisfaction from the consumption of two kinds of goods, private goods and collective goods. (He may, of course, derive utility not only from his own consumption but from the status of others, rejoicing in either their good fortune or their bad.) Let us distinguish initially the individual's wants reflected by his tastes and preferences and his effective demands, determined by his utility function *and* the constraints, such as income, that bind him.

For private goods there is a market through which individuals can make their effective demands for goods felt, and in the context of an enterprise system whenever the aggregate of such individual demands warrant there is incentive for private producers to meet these demands. Collective goods differ in that (as we have seen) private markets fail to respond to real effective demands; thus *collective action* is required to satisfy *individual demands*. The devices of government can provide the form of collective action that substitutes for private markets in channeling resources to meet the aggregate of individual demands. (So far as the individual's utility depends upon others' consumption, it is likely to be outside of his individual ability to do much about; thus here is further motive for looking to the government.) In this attractively symmetrical view, government activity permits individuals better to achieve their individual objectives.

These assumptions have settled an issue of principle (that there may be a legitimate role of government) without having answered the practical question of which public goods the society should provide. Granted that individuals have demands for collective goods, how should individual preferences be aggregated to determine whether the aggregate welfare is sufficient to justify the total cost? Is this answer affected by the system of taxes used to raise the funds? These are among the critical questions of welfare economics.

A simple and uncontroversial solution is available in cases where unanimity prevails. Clearly, if some quantity of a particular new public good had the happy property that for every individual it added at least as much to his utility as his contribution to its cost detracted from his utility, such a good would be desirable. A totally de minimis view of the public interest would limit it to cases that fit this requirement. It is de minimis because, while it is conceivable that there are goods that satisfy this criterion, there cannot be many. In a many-person economy, if even one person had less use for the public good than his contribution to its costs no positive quantity of output would achieve unanimous consent.

Most public activities imply (often intentionally) a redistribution of income which leaves some individuals worse off and some better. Whatever the merit of establishing *at least* the quantity that might achieve unanimous consent as desirable, any implication that exactly that quantity is desirable seems quite unwarranted. Few economists, welfare or other, are content to rule out any change merely because

it has redistributional consequences. To do so invests the initial distribution of society's resources with an overriding sanctity. Nor can we avoid this difficulty by assuming that an individual's utility function includes as an argument preferences about the distribution of income. For, unless an implied redistribution were to be regarded as desirable by everyone, there would be no unanimous consent to it.

The struggle of formal welfare economics to escape this dilemma has been tortured and fruitless. The Kaldor-Hicks "compensation principle" wherein a change that benefited some but not all was justified if the gainers *could* compensate the losers even if they *did* not only encounters technical ambiguities [10] but implies an inherent neutrality toward redistributional consequences.

Neither am I impressed with the argument that we ought to be neutral toward redistribution in allocational decisions, on the ground that redistribution of income can be accomplished directly if it is desired. The inertia of social change is such that this would overweight the current distribution of income, and would serve in practice to limit the possibilities for redistribution.

Clearly we would be paralyzed by an operating rule that said: "provide no public good if it changes the distribution of income." But if we permit redistributions there is no reason why we cannot regard some as desirable and others as undesirable.

Clearly some reckoning of redistribution gains and losses is required if all proposals are not to be rejected. If one insists on basing decisions on individual utility functions this implies either assigning weights to the utility of individuals or being able and willing to measure and compare individual utilities. Making interpersonal comparisons is potentially unattractive,[11] unless it is itself subject to well-defined rules of procedure. Such rules might be formulated. For example, one may take a strict majoritarian view. But there is nothing inherently just or appealing about a rule that leads to the median effective demand of the society being dominant: indeed, without protections of minorities it would seem offensive to most people.

The major objection to a utility-consensus view of social welfare functions is that it is nonoperational and does not seem to provide guidance to the decisions of real societies. Certainly we *do* take decisions with less than unanimous consent. Certainly too, many public goods provide benefits in excess of their contributions only to very small minorities of the society, but with the evident acquiescence of sizable majorities. One can argue that, ex post, individuals are thus revealed to value the benefits which accrue primarily to others. But this rationalization leads us back to a de facto definition: whatever the Government does is revealed to be desired by the people.

---

[10] Scitovsky [41] demonstrated an inconsistency in the principle in the case in which price changes occurred: Baumol [4], Little [23], and Kennedy [21] questioned the use of potential compensation.

[11] Banfield ([2], p. 11) suggests the following as an extreme parable about a society that can make accurate interpersonal comparisons of the subjective states of individuals A and B:

> "If A's preference is for putting B into a gas chamber, then 'bliss for the whole universe' is served by his putting him there, provided only that B's loss of satisfaction at being put there is less than A's at putting him there. Even if B claims that his loss of satisfaction will be at least as great as A's gain of it, the just and equitable society will tell him that he is mistaken and put him there anyway in the 'good cause of adding more to the rest of mankind's (that is to say, A's) well-being.' If perchance A and his friends constitute 51 percent of the population and B and his friends only 49 percent, the matter will be simple indeed."

Thus if formal welfare economics does not go beyond individual utility functions, it fails either because it justifies too little or because it justifies too much of public expenditure. Viewed from the vantage point of welfare economics, public decisions about public goods appear to be impossible to make.[12] Fortunately, other economic views are possible: economists are saved the humilitation of abandoning as barren a fertile field. It is the wasteland of welfare economics, not the reality of public decisionmaking that is the mirage.

A partial escape from the wasteland can come from a pluralistic view of the individual. Suppose that each individual in addition to his personal evaluation of any proposed activity will also view it from the point of view of any one of a number of groups he belongs to, be it social club or trade union.[13] If he is willing to be bound by the consensus view of the members of the group, there is a much greater possibility of consensus, first because a significant clustering of views is likely to emerge, and second because logrolling between groups can create collections of activities that command dominant majorities. Suppose individuals are prepared to accept and to be taxed for things they consider socially worthwhile, such as (say) foreign aid, wars on poverty, and higher pay for Senators, even though they cost many individuals more in income foregone than they contribute to that individual's utility. They accept them as part of a package which they find adds to their own utility on an all-or-nothing basis.

The view that social choices may rest on collective values arrived at by caucus rather than by simple aggregation is more than an escape from the general impossibility of deriving a social welfare function from individual values. It has positive merit in that it embraces a view of the individual which many find descriptively accurate and analytically helpful. In this view an individual functions in a pluralistic sense with loyalities, commitments, and valuations at many levels: to himself, his family, his church, his neighborhood, his employer—and possibly also to his race, religion, class, country, and political party. The pluralistic view is the heart of sociology, social anthropology, and much of economics. If it is accepted it suggests that individuals will be prepared to act on collective issues without inevitably tracing back to the explicit question: "What's in it for me?" They may ask instead: "What's in it for the Negro?" or the farmer, workingman, etc. If they do, they invite an analysis of the views of political pressure groups, which usually have highly articulate spokesmen and well-defined programs they are seeking to enact.

2. THE POINT OF VIEW OF WILLINGNESS TO PAY*

The difficulties of deriving an aggregate preference ordering based on individual utilities arise from the incommensurability of individual's utility indexes. Without unanimity of views or well-defined rules for assigning weights, consistent decisions become impossible to make.

[12] This is, of course, a conclusion of Arrow's "General Impossibility Theorem": "if we exclude the possibility of interpersonal comparisons of utility, then the only methods of passing from individual tastes to social preferences which will be satisfactory and which will be defined for a wide range of sets of individual orderings are either imposed or dictatorial." [1], p. 59. For discussion and criticism of this famous proposition, see especially: Little [23], and Tullock [48].

[13] See Truman [47] for an extended discussion of groups in the political process.

*Further discussion of this issue is found in the papers by Arrow and Margolis in this volume.

A second view is to ask not how much an individual values a given collective good, but how much he is willing to be taxed to provide it. The differences between this view and the previous one may seem small but they may make a substantial difference in one's view of the scope of the public interest. In the first place a shift from every individual having a positive preference for some activity, to his being willing to tolerate it, is practically quite substantial, as the well-known political literature on the use of veto power makes clear. Second, the metric of value of the activity has been shifted from the inherently unmeasurable "utility," to the discernible, and interpersonally comparable, willingness of individuals to pay. To be sure we pay a price for this, in that an individual's willingness to pay reflects the status quo, and particularly his income and wealth. We must therefore, be all the more wary about distributional biases toward the status quo.

When applied to the collective consumption good this approach leads to the so-called *pure theory of public expenditure.* This theory has a long life and is well articulated by Bowen [7] and Samuelson [38].[14] In barest essence the argument is that since demands of different individuals for a collective good are complementary rather than competitive we can add the willingness to pay of different individuals and if the aggregate sum exceeds the costs, the good is worth producing. For all quantities for which this is true, there exists a tax policy which would collect levels of taxes sufficient to cover marginal costs, and still leave all citizens satified.

Appeal to the "existence of tax policy such that * * *" sounds very much like the compensation principle once again. And it is subject to a similar objection. If the tax structure is not (or should not be) malleable, there is no bliss in this solution. At independently determined tax rates some taxpayers are getting more than they are willing to pay for and are being coerced to provide what others want. If these others' needs are in some sense more meritorious than those of the reluctant taxpayers, this may be appropriate, but there is then no automatic stopping point. One cannot escape the distributional question, unless one insists on regarding it as irrelevant. If the collective good involves aid to the needy aged one may take a different view of the effects of the coercion than if it involves providing a civic yacht harbor.

Nor (in my view) are these difficulties with the optimal solution merely symptoms of imperfections in the taxing schema. Citizens have social values about appropriate tax policies too. Suppose we are building a public playground to be used by underprivileged children. If A is a rich misanthrope and B a poor Samaritan, there is no compelling reason why B should carry more of the tax burden, even though he may be willing to do so.

Notwithstanding these and other difficulties, some have argued that this approach offers the following rule as a usable rough guideline to public decisionmaking: if, in aggregate, effective demands exceed cost, the service should be provided whether or not payment is exacted. Unfortunately this opens the door to game strategic behavior. Suppose payment proportional to demand is not to be required. Any group that

[14] A better historian of thought might define it as the Sax-Wicksell-Lindahl-Musgrave-Bowen-Samuelson * * * tradition, to recognize the apparently valid theorem that no one ever has an original idea. Samuelson is diligent in identifying his predecessors and Musgrave provides admirable summaries of earlier views. In an earlier day theories of this type were called voluntary exchange theories.

knows it will not be asked to pay more than a fixed share can exaggerate its valuation of a service it desires. If (on the other hand) proportional payment is to be required, every group will have an incentive to understate its real valuation as long as others value the service enough to get it provided.

3. THE POINT OF VIEW OF AN AGGREGATE SOCIAL WELFARE FUNCTION*

Society is necessarily made up of individuals, but it need be no simple aggregate of them. In looking at their behavior it may be that the interdependencies between people and their interactions are more important than the individual value structures. Nations, races, even football teams, acquire personalities and modes of behavior. While a search for collective values has until recently been more congenial to political scientists and sociologists than to economists, an increasing number of economists are moving toward the view that individuals voluntarily yield certain coercive power to a government which is somehow charged to discover, articulate and implement social priorities, or collective wants.

This, collective, view is broader than, not necessarily competitive with, the previous ones. It may be that social priorities are indeed arrived at by some aggregation process. But there are other possibilities as well. Political theory has long been concerned with the legitimacy of government and with the nature of the social contract among citizens or between a government and its citizens.

In this view a collectivity, *society* may be fruitfully viewed as an independent entity possessing its own value orderings. In Rothenberg's view [36], if I understand it, social valuation as opposed to solely individual valuation is an existent reality. This view has the great pragmatic value that it invites the search for revealed social priorities without insisting on a single source of them. If priorities with respect to income distribution (for example) are established, and consented to by the citizens, then the distributional consequences of particular public decisions become "benefits" or "costs" instead of barriers to either clear thinking or clear action. The formidable measurement issues—of quantifying the benefits, and assessing the costs—remain, but a major hurdle has been crossed. In this view one is aggregating not across individual utility, nor willingness to pay, but rather across individual political influence and tolerance. In so doing we vest to some extent the existing distribution of political power and influence, but possibly it can be argued that this is more nearly the result of a social contract than is the distribution of wealth.[15] The degree of consent that is required is whatever the political process demands.

Personally, I find this view of the problem both congenial and fruitful. It does not dispense with the individual and individual values, for one must ask how the political process articulates the public interest, but it does recognize that most individuals have a large range of things they will accept without opting out of the system, and will vent their approval or disapproval in some sort of orderly political process. It does not lead, inherently, to either a minimal or maximal

[15] Buchanan and Tullock [10] develop an ingenious "economic theory of constitutions" in which individuals find it advantageous to agree in advance to certain rules even though they may work to the individual's disadvantage on occasion.

*Further discussion of this issue is found in the paper by Arrow in this volume.

role for government. It does not exclude distributional questions from policy, nor does it vest the existing distribution of income with a special status. What it lacks is any clear indication that one situation is superior to another in a wholly unambiguous sense. This does not mean that whatever society does is desired, rather it means that particular public decisions can be shown to be valid only in terms of particular value judgments. It tends to pose issues of public policy in terms of whether society does in fact hold certain value judgments rather than in terms of the demonstrable inherent legitimacy of certain activities. Some will regard this as retrogress. I am not so inclined. Economists have long sought a calculus of consent, but in the search have found it easier to derive a lower bound to public activities than to define their proper domain. If we are fruitfully to discuss public expenditure policy we must be prepared to go beyond this.

The central issue in this debate does not concern the logical correctness of looking at social choices on the one hand as an aggregation of individual values or on the other as a two-step procedure in which we first agree (some way, any way) on collective values and then use them to make social choices. Any perfectly understood aggregate behavior can be decomposed into its disaggregated elements. It is instead a matter of research strategy. Are we likely to achieve greater insight one way or the other? If there is stability in aggregate social values then basing policy on such aggregate values is likely to work. My own view is that at this stage in the development of our science there is more insight and less bias in the third view of the public interest than in either of the others. Many will disagree.

If one accepts the notion that aggregate social views must be discovered there remain two important questions to discuss. First, how these views get articulated and, second, how competing objectives get reconciled. It is convenient to treat them in reverse order.

D. The Problem of Weighting*

It is increasingly the practice to treat the public interest (or "social welfare") as a function with several arguments. For example following Marglin [28] we might write:

$$U = U(Z_1, Z_2, Z_3, \ldots Z_i)$$

where the Z's are different aspects of the public interest such as economic efficiency (or contribution to national income), the pattern of income distribution, the rate of economic growth, balance of payments equilibrium, economic stability, national security, and freedom. Whether one considers these as different aspects of a one dimensional index of utility, or as different dimensions of a multi-dimensional concept is more than a semantic issue. If for example Z_1 and Z_2 are readily comparable in terms of a cardinal measure of their contribution to utility (e.g., market value of electric power and market value of irrigation water) they can easily be elements of a scalar measure of welfare. Their relative weights are given by the common yardstick used for measuring them. If on the other hand Z_1 and Z_2 represent efficiency

*Further discussion of this issue is found in the paper by Freeman in this volume.

and freedom each of which (let us suppose) may be meaningfully defined and ordinally measured we may have no simple yardstick for comparing them. Thus the problem of trade-offs between these "separate dimensions" of a public interest vector remains to be solved. Because noncomparability, of difficulties in comparison are a feature of many public choice problems, it seems to me convenient to regard the public interest as genuinely multidimensional, and thus to consider explicitly conflicts among objectives.

The number of dimensions and their definition is a matter of analytic and operational convenience. Because there are many different sources of a public (as distinct from private) interest there will be many forms of a proximate contribution to public welfare. Whether it makes sense to combine different effects into a single dimension or to treat them as separate dimensions of the public interest depends upon whether one believes there is an acceptable cardinal measure of their contribution to public welfare. If so it makes sense to combine; if not, it makes sense to keep them separate.

With a one dimensional objective there is a simple decision rule. If (for example) one is concerned solely with contribution to national income it is conceptually easy to choose between a dam in Oregon and a retraining program in West Virginia. But if one cares as well about who gets the income, such a simple rule can become simplistic. Given multiple objectives, it is inevitable that individual proposed actions will affect more than a single dimension. Inevitably also the objectives will often conflict. The definition of a multidimensional objective function neither creates nor resolves the conflicts, instead it identifies them.

The central aspects of choosing policies when faced with multiple objectives are how to define an appropriate measure of each objective, and how to resolve conflicts among objectives.

A very simple, indeed trivial, theorem says that if weights are left unspecified, any policy A may in general be made to appear less or more desirable after the fact than an alternative B (which may easily be "not A"), by specifying which objective is implicity the important one. This theorem is important only because it is so frequently neglected.

Many forms of "implicit" weighting exist in practice. It is almost routine in lay discourse to argue that because a proposed policy advances *some* object of social policy, it is desirable: "The war on poverty will improve the distribution of income." Or to argue that because it retards some other objects it is undesirable: "The war on poverty extends the role of government and thus reduces individual freedom." Neither of these statements tells us whether the war on poverty is desirable.

Somewhat subtler is the implicit neglect of certain objectives by assuming the dominance of others. Much of the economist's traditional emphasis on efficiency has had the effect of giving it a very high weight relative to growth or distribution. Interestingly, Joseph Schumpeter, always treated as a giant with respect to his theories of development and cycles, is still regarded as a crank with respect to his views on monopoly, because he challenged orthodoxy by arguing that static efficiency considerations are overweighted relative to growth.

If objectives are genuinely multidimensional and not immediately comparable, some solution to the weighting problem is implicit or

explicit in *any* choice, and that solution reflects someone's value judgment. Put formally, we now accept in principle that the choice of weights is itself an important dimension of the public interest. This choice is sometimes treated as a prior decision which controls public expenditure decisions (or at least should), and sometimes as a concurrent or joint decision—as an inseparable part of the process of choice.

1. WEIGHT SELECTION VIEWED AS A PRIOR DECISION

Several widely divergent views of the public decision process have in common a view that important aspects of the weighting decision should be regarded by the decisionmaker as given to him. Two of these virtually assume lexicographic ordering of certain objectives. In lexicographic ordering, objective "1" is dominant, but in case two choices are equivalent in terms of objective "1," choice is made upon the basis of objective "2," etc. (Listing in alphabetical order is the best known lexicographic procedure.)

One view of this kind is that efficiency, as measured by private market allocations, is the dominant criterion; if a project is efficient in this sense, it (or some substitute) is worth undertaking; otherwise not.* Once a project is so legitimized, the decisionmaker is welcome to examine other, secondary objectives in project selection or design. This appears to me to be the view of McKean [27], Harberger [18], and Mishan [32] when they insist that the correct (really the only correct) discount rate for discounting future benefits and costs and for assessing the opportunity costs of public funds is some specified measure of the marginal productivities of capital in the private sector.[16] This view "solves" the weighting problem of assumption and makes many otherwise difficult decisions easy. Those who hold this view are repeatedly appalled at the obvious outrages performed by the public sector, and the apparent acquiescence therein of otherwise sensible men.

A different but no less arbitrary lexicographic view is that public budgets for particular activities reflect dominant social choices, and that while efficient allocation of the funds within such budgets is appropriate, efficiency considerations do not reflect sensibly on the size of the budget.[17]

While neither of these forms of solution-by-assumption of the weighting problem is likely to prove literally satisfactory in all situations, either might provide insight into how the public decisionmakers regard their actions as constrained by a society's underlying consensus view of key issues they face.

An alternative to taking weights as inherently given is to regard them as an explicit prior decision. It is conceivable to imagine the political system having a procedure whereby we decide upon and then announce a fixed set of weights that will be controlling in choosing among income, income distribution, and so on. No one suggests this is the procedure used, although both Chenery [12] and Marglin [29] urge this as a real possibility for economic planners. Chenery,

[16] The discussion of the discount rate as an implicit weighting scheme involves technical issues best omitted from this paper. They are discussed in [43].

[17] A paper of my own [42] is a good example of use of this approach.

*Further discussion of this issue is found in the paper by Baumol in this volume.

for example, suggests the planners announce a national income equivalent to balance-of-payments effects, thus making for a fixed tradeoff.[18]

Eckstein [16], Haveman [19] and Weisbrod [51] among others, while more or less accepting the view that weighting decisions are relatively stable and pre-existing, do not wish to assume them nor to expect political leaders to articulate them. Instead they suggest attempting to infer the weights by an analysis of past choices. Eckstein [16] suggests we look at an issue such as differential tax rates in which a decision on distribution is at the heart of congressional intent in order to discover implied values Congressmen hold. Haveman [19] applies this approach to evaluating water resource investments with respect to a multidimensional objective. These approaches are suggestive, even if one is unwilling to go all the way and take congressional actions as perfectly revealing social consensus.

Weisbrod deals with a two-dimensional objective function covering efficiency and income distribution. He suggests that every example of choice of a less efficient over a more efficient alternative implies a minimum implicit weight to the redistribution that is involved. His hope is that analysis of many decisions would reveal a weighting scheme. The advantage of this procedure is the potentially large sample available. The hazard is that any irrational choices, any misestimates of efficiency, and any nonincluded objectives would all be imputed as distributional benefits. But the real test of Weisbrod's suggestion will depend upon whether it produces a clear and consistent pattern of implicit values. To my knowledge, it has not been tested as yet.

Yet another approach to weighting of objectives is to regard certain objectives as constraints. Suppose we would like to attain one objective, with the constraint that we cannot concentrate on that objective until certain other basic goals have been secured, at least to a minimum extent. Initially, then, these latter goals assume high priority, for we cannot turn our major attention to the former objectives without having satisfied the latter. Once the latter goals have been satisfied, they can be given shadow prices which reflect the price paid in assuring their achievement. This is often a useful (as well as a traditional) approach to allocational problems. It may, however, prove difficult to define which objectives are genuinely to be regarded as constraining.

2. WEIGHTS AS THE OUTCOME OF POLITICAL PROCESS

The discussion above treated the appropriate weighting scheme as a preexisting condition for the decision process. Dahl and Lindblom [13], Braybrooke and Lindblom [8], and Maass [25], Major [26], Banfield [2], Eckstein [16] among many others regard weights as generated by the process of decision. In this view the political process addresses weighting problems not abstractly but as a case-by-case confrontation. In each case need for a decision about what to do forces a discussion, or compromise, or struggle between competing objectives. To Maass, the essence of the *legislative* process is the making of choices between conflicting objectives. The decisions will be made upon the

[18] One is not limited in principle to linear relationships. Chenery's tradeoff at time t could be a variable function of the size and sign of balance of payments disequilibrium at time t.

basis of the information (or prejudices) available, and the scope for the analyst in affecting the decision will be reasonably to identify the choices. Eckstein expresses the strong view (to which I find no academic dissent) that administrators and project analysts should not arrogate the weighting process and bury the choices within a single measure of benefit.

This whole area strikes an analyst as untidy. It is clear that choices among objectives must be made, and that the political process must somehow make them. But we appear to be undecided about whether it does so within narrowly confined limits of underlying consensus, or with substantial discretion. Is the alternative to the invisible hand the responsible arm or the visible paw? Perhaps more important is the source of such political discretion as exists; is it simply variance in underlying views, is it ignorance or indifference on the part of the citizen, or is it an explicit delegation of authority by the electorate? Answers to these questions critically affect subsequent research strategy. Banfield [2] appears to espouse the nihilistic view that discretion arises from such deep underlying ignorance and/or indifference that there are no effective limits on the wondrous ways of politicians. At the opposite extreme, if one assumes that politics is a mere veil that masks a variety of different underlying views, a key to understanding the outcome is to study the variance of underlying views. Downs [15] and to a lesser extent, Lindblom, embody this view. It has important precursors in an older political science literature on interest groups.[19]

What is particularly disturbing is that I can find virtually no disposition on the part either of economists or political scientists to engage in an empirical study of the decision process that will resolve these areas of debate. The large literature is almost entirely theoretical and assertive. Yet, survey data about public attitudes on issues exist and provide some sort of a base. Similiar data about ex-post public reaction to political decisions and procedures might be developed. In fact we now ask such evaluative questions but only about things like a major war or the overall evaluation of a president or a party. Even to say more would take me further along the road than I care to go to design a study of how decisions are reached and the extent to which they are responsive to public opinion. Such a study seems worth somebody's undertaking; it seems more promising than another decade of assertions in resolving our differences.

E. Articulation of the Public Interest

There are divergent views about how the political structure articulates social priorities and the extent to which *process* determines outcome. It is helpful to start with a dichotomous classification that overstates differences in points of view. In one view the political process is a market-like mechanism that coalesces views that inhere in the members of the society. Here the political process is a facilitating and implementing one, not inherently a formative one. An efficient government, like an efficient market, quickly and accurately translates inherent preferences into explicit consensus. Just as a market may perform a mapping function (since people discover their preferences best

[19] An admirable summary with an extended bibliography is found in Truman [47].

by confronting real alternatives) so an efficient government serves to help people discover as well as fulfill their their collective preferences. In principle (though, of course, not necessarily in practice) one should be able to simulate such a process, and simply add a government sector to a general equilibrium model of the society. Government (in this view) is a decisionmaker only in the limited sense of a reactor and processor of signals it receives. A properly functioning government will arrive at an optimal decision set without exercise of independent judgment.

The second view is that while individual social preferences clearly exist and play a role, they are sufficiently inchoate, ambiguous, or conflicting that the political process is required to force a public interest and it does so with substantial discretionary choice. Without knowledge of the motives of the governors, and of the political process itself, there is no indicated solution, nor accurate prediction of governmental action. In this view, to create a government sector it must be given objectives, procedures, and decisionmakers. However, the way in which individuals' preferences constrain or otherwise affect public decisionmakers must be made explicit.

The differences between the approaches are important not only in terms of the information each requires in order to permit prediction of outcomes, but in whether comparative static analysis is possible or must be replaced by a genuinely dynamic model of political decisions. I limit myself to a few dicta here, and refer the interested reader to the fuller discussion in [43].

1. THE GOVERNMENT AS A QUASI-MARKET

In all market-type analyses of the political process, voting is the means by which individual values are translated into action decisions. Bowen's pioneer article [7] is of this type. If we assume that the tax-burden of any public expenditure on each individual is known to him, and if we assume everyone votes under simple majority rule we can predict the outcome of elections.

Abstracting from the effect of the tax structure on political choices, it is clear that so far as we can reduce preference to a single dimension, it is the median preference that dictates the outcome of majoritarian voting. The result thus depends wholly upon the pattern and variance of voters' preferences and the structure of taxes. The result may, but need not, yield the pure theory of public expenditure solution. That is based upon aggregate willingness to pay and thus, if tax burdens are equal, it depends upon the mean rather than the median willingness to pay. Bowen believes that the institutional facts are such that median and mean will tend to coincide because voters' preferences are symmetrically distributed. I know of no evidence with which to confirm this conjecture. Bowen does not suggest that a referendum on each proposed decision actually is the decision procedure, but rather that it is a decision procedure that might be used, and that perhaps roughly is represented by the institutions of democratic government.

A much more elaborate but broadly similar theory is presented by Downs [15] to whom government consists of men who like the emoluments and perquisities of their jobs and whose goal is reelection or reappointment by elected officials. The government is thus motivated

to maximize its political support. The government is interested in a citizen's vote, not his welfare; it must, however, cater to his view of welfare to get his vote. Were it not for uncertainty, Downs' model would be fully mechanistic and predictable.[20]

A less mechanistic model is offered by Maass. He offers an explanation that to a major degree is a two-stage political market theory. In the first stage the voters choose men, who in their personal capacity, and in virtue of their character, are fitted to discharge the task of deliberation and discussion at the parliamentary stage.[21] In the second stage, these officials are held accountable for their political acts by the need to seek periodic reelection.

Maass' model gives rather more freedom to the politician than does Downs': officials are not worrying about the probability of reelection in every move, only in their overall performance. At the very least they are only constrained to be aware of their constituents' sensibilities. Indeed if they regard their mandate as sufficiently general they have the need as well as the opportunity to crystalize their constituents' values. Here Maass verges into a creative or formative view of government and the definition of public interest.

Market theories can be criticized from within or from outside the market framework. Most critics of the votes-as-market-signals approach to analysis of the political process have sought a more substantial discretionary role of government; we shall consider these just below. Within the market-analogy framework, Arrow [1], Black [6] and Buchanan [11] have questioned the ability of such a system to translate inherent preferences into rational social priorities, and the efficiency of such a quasi-market system.

Arrow's well-known demonstration of the paradox of collective choice is so simply illustrated that it bears repeating.

Individual	*Individual's preference ordering*
I	A→C→B
II	C→B→A
III	B→A→C

Even if each of the individuals has no difficulty ordering his preferences among three competing possibilities—A, B, and C—there may be no clear collective preference. In the example clearly two-thirds prefer A to C, and two-thirds prefer C to B. Thus, if they first choose between B and C and then between A and C, A will command a majority. But two-thirds prefer B to A and a different order of choice can produce any one of the choices.[22] Black [6] and Rothenberg [37] among others have explored the theoretic consequences of alternative voting and balloting schemes in producing outcomes, and both dis-

[20] Uncertainty creates some scope for leadership, and for errors that give politics an interesting dynamics. It is not possible to pursue them now. See Downs [15], esp. chapters 5–8.

[21] The words are Barker's [3], quoted with evident approval by Maass [25], p. 569.

[22] The theorem proved is that transitivity in all individual orderings is not sufficient to assure a collectively transitive set of choices.

cover a purely political dimension to the politics of consent. To this we shall return.

Buchanan [11] in a paper stimulated by Arrow, identifies some important weaknesses in the market analogy. He accepts the voting process as analogous to the market mechanism, but is wary about drawing welfare implications from the analogy. Buchanan believes in the efficiency of decentralized market decisions and is concerned lest the loose analogy with voting give a similar blessing to voters' decisions. He notes:

(1) Voting involves an extra dimension of uncertainty: consequences follow the collective vote, not the individual vote; therefore, the voter may not really vote his own best interests because he underestimates the possibility of the decision impinging on him. Indeed since in a collective vote there is a diffusion of responsibility for the collective decision, the individual may act in the mass as he would never act individually. A man may vote for prohibition, for capital punishment, or for a war policy while at the same he would not abstain from alcohol, invoke the death sentence, or opt for military service for himself or his son.

(2) In voting, the individual is influenced by his sense of participation in social choice. A vote for open housing need not imply willingness to live in a racially mixed neighborhood: indeed many of the most ardent supporters of such laws have exercised their option to move further away from integrated neighborhoods. Men may be willing to do collectively unto others what they would not do individually nor consent to have done to them unless done to all.[23]

(3) In voting, the individual is often faced with indivisible votes for mutually exclusive choices. He cannot make marginal choices, or influence very much the definition of candidate or choices. Often he votes for candidates some of whose policies he disapproves. Thus the mandate of a winning candidate is readily misinterpreted.

(4) Minority votes are wasted, whereas even minority preferences exert influences in the market. If fear of wasting their votes leads voters to vote for their second choices, even actual votes for candidates may fail to reflect the strength of the support their views have.[24] Nonvoting, an alternative form of expression, is not often easily interpreted.

(5) Typically, voting provides equality of influence of individuals, instead of reflecting command over resources. Bowen is wrong: the weighting of individual choices is different in the marketplace and the polling booth. Buchanan here goes well beyond the point that it would require some remarkable coincidences to assure that the distribution of political and economic influence were perfectly correlated. He adopts (I think) a view of Frank Knight's [22] that market votes are in some sense superior to political votes.[25]

These critiques by Arrow and Buchanan serve to warn against too quick acceptance of an analogy of the political process with the market process. The practical question is whether the two situations are suf-

[23] Margolis and Marglin offer yet a different view of this problem. A man may be willing to pay his share of a joint venture only if he can force his reluctant neighbors to do the same [31]. In this view a vote is an offer to sign a social contract if enough others also sign.

[24] This phenomenon may be found in the market too. If buyers can be persuaded to accept second best, their true preferences may not be effective.

[25] Knight points out that individuals may be unequally constrained by voting from utilizing their "normally available capacities for action." Evidently Knight feels that unequal constraints are more unfair than unequal initial distributions of wealth.

ficiently similar that the economist's techniques of analysis and his theorems about market behavior can be applied directly and fruitfully to political decisionmaking. My own answer after reading in the political-market literature is "No." I certainly believe that individual citizen voters both influence and constrain political choices; but they do so within limits sufficiently broad that attention to choices *within the limits* needs, deserves, and repays analysis.

The important issue is, of course: how large are the limits? In the political choice arena, the limits seem to me sufficiently large that we cannot merely pay attention to voters' preferences. All of the following contribute to a substantial discretionary role of government: the variance in individuals' views; the fact that voters choose infrequently and among bundles of policies; the fact that many views are not held so strongly that a political leader who violates them is at once anathema; that many voters' preferences are inchoate, uncertain, and subject to change; that pressure groups can and do negotiate with governments; and finally that political leaders can, in fact, lead their followers on many issues. If these things are true, the choice set of the government while constrained by voters' preferences may be far from singular.

2. GOVERNMENT AS AN ORGANIC CHOOSER OF ENDS*

Dorfman [14] presents a suggestive model of the public sector that is close to the market analogy models while yet explicitly introducing governmental choice. He returns to an older view of the world in which individual views are oriented to and expressed by socioeconomic pressure groups. The government is in a sense a coalition of such blocs that cooperate in order to provide public goods. The constraint of coalition requires that no group be sufficiently badly treated that it is motivated to withdraw its support. In Dorfman's terminology each group has a potential voters' surplus: the excess of its self-perceived benefits from provision of the public good over the group's contribution to its provision. The imposed constraint is that this voters' surplus must be nonnegative for each group. But the government has multiple possibilities beyond this since it can weight the interests of the different groups unequally. Dorfman notes that the differences in political parties constitute, in effect, different weighting schemes, and members of the party in power have substantial freedom to pursue their own preferences, and to compete for the right to govern by catering to those groups that can generate sufficient support to keep them in power.

What is refreshing in this formulation is first the explicit statement of the constraint; second, the recognition of Governors as a group with ends of its own and some ability to pursue them; and third, the recognition of intermediate groups as a focus for articulating and coalescing individual values. What the behaviorial implications of this view are, are in dispute. Dorfman believes that the need to satisfy all (or most) groups imposes a downward pressure on public expenditure; I suspect that this same need is more likely to lead to logrolling and expenditure increases than to limit expenditures.

---

*Further discussion of this issue is found in the paper by Dorfman & Jacoby in this volume.

A more direct view of governmental search procedures can be represented by work of Maass [25] and Major [26]. Here the legislative process is centrally concerned to discover what agreement on objectives can be reached. Partly this is a matter of bringing together elected spokesmen for individuals and groups, but more basically it is a genuine search. The process of discussion is critically involved in a mapping function—it poses issues that permit both legislators and their constituents to discover their views about objectives. It may thus permit (via compromise and persuasion) the development of agreement on public objectives, which can serve for a time as the social objective function. This view continues to disquiet many economists, for it leaves a large element of slack within the governmental decision process that is not readily understood in terms of inputs into governmental decisions. Thus we are limited in our ability to predict Government decisions.

Rothenberg [37] has an enormously elaborate model whose purpose is to remove this element of slack. Unfortunately it defies concise description. To quote his summary, in part, "The legislative process is seen as an n-person, nonzero sum, repeated cooperative game of strategy, for which no general solution exists * * * unfortunately, manipulation of the model to elucidate its complications is beyond the scope of the present paper." I confess to being skeptical. Game theory, here as elsewhere, seems to me to provide a vocabulary for discussing a multiplicity of outcomes rather than proving a tool for predicting particular outcomes.

One aspect of Rothenberg's model picks up a strand that has characterized the work of Duncan Black [6] over two decades: the influence of the institutional rules of legislative decision on the outcome of the decision process. I shall neglect this fascinating literature only because the chief question this paper addresses is the choice of policies within a relatively fixed institutional framework, not with the influence on choices of changes in the institutional framework.

Banfield [2], like Maass and Black, is a political scientist, but unlike them he is profoundly skeptical of the utility of the type of analysis that characterizes economics: the solution of constrained maximization problems. His central point is that in the articulation of social values, the techniques of economics do not merely fail to predict behavior (because we need more information about individuals' values, or the nature of the constraints, etc.) but rather that they are inherently biased and are bound to mispredict. As I understand Banfield's view, the alternating and intertwined activities of discussion, of struggle, and of arbitration, constitute the heart of the decision process and exhibit such variance in possible outcomes that they dominate the problem of explaining political behavior. To neglect these in favor of the inputs into the political hopper is to neglect the major sources of variance in favor of the minor ones. In effect he argues that the limits placed on decisionmakers by individual preferences are so wide as to be of no real interest.

Banfield's view might be expected to lead him toward the Black and Rothenberg analyses of process, or toward the literature by Schelling [40], and others on the strategy of conflict, or perhaps toward the organizational theorists. Here too, however, Banfield is pessimistic almost to the point of nihilism. His pessimism seems to me

extreme, but it underlines the absence as yet of a compellingly effective set of predictive theories. Banfield's view that politics and thus also the prediction of its outcomes, is an art rather than a science, may prove right; at this stage neither the economists nor the bulk of the political scientists seem prepared to abandon the search for an explanation of public decisionmaking.

3. EVALUATION OF ALTERNATIVE VIEWS

This extended discussion of alternative views as to the nature of and means of articulation of the public interest clearly reveals no consensus. Indeed it is relatively easy to demonstrate why each approach is deficient. While one of these theories may with slight changes prove adequate, it is more likely we will have to await a more profound insight. It seems to me of prime importance to distinguish between the present inadequacy of our theories, and the presence of a phenomenon. Things need not be understood to exist; substantial agreement on certain social priorities may exist despite our inability (at any point in time) adequately to derive them from basic principles. Suppose we are today unable to *derive* from individual values a consistent set of social valuations that enable us to say "there is a clear collective demand for this activity." Are we constrained to act as if it does not exist, or to settle for the logic of the lowest common denominator of acceptable action, that which will command unanimous consent? The answer seems to me to be "No." To answer otherwise would make us prisoners of our ignorance. It is less elegant, but not less scientific, to take as a starting point for evaluating social actions, the revealed objectives of society instead of the derived ones.

Let me put the matter more strongly. Suppose one could prove that in some fundamental sense the prediction of social values is impossible from basic information about individuals and their political representatives. Would it then be necessary to quit the analysis of public decisionmaking? I think not. One might take the nature of social valuations as revealed by past actions and assume that such preferences have some stability. In other words it might be possible to infer dominant collective social priorities from social actions and the repudiation or nonrepudiation of them by the electorate.[26]

In the United States today it is hard to avoid a strong presumption in favor of believing there is a strong collective preference for certain public goods such as public aid to education, for improved highways, for redistribution of income in favor of the elderly, and in favor of farmers. Are these only today's choices of today's individuals or are they a reliable indicator of how Americans are likely to feel next year or even 10 years hence? I think there is evidence of stability in some choices, and gradual change in others. In any period there are highly debatable issues that get resolved and stay resolved for generations, not unanimously, but sufficiently that legislators of each party are content to let them lie. Today the debates about minimum wages, and social security (even medicare), seem remote and (in this sense) resolved. We may ask about the appropriate level of the programs, but their existence is not likely to be subject to any serious challenge. On the other hand, I would not argue very strongly in favor of the pre-

[26] See Birdsall [5] for such an attempt.

sumption of similar preferences for or against integrated housing or schools, for foreign aid of nonalined or Eastern countries, for domestic gun control, or for international disarmament. At any time some issues are genuinely unresolved, for others the degree of consensus is uncertain.

Evidence on these matters is available even in the absence of analytic solution. Not only do elections provide some information about revealed preferences but enormous quantities of attitudinal information can be collected by the techniques of survey research. As theorists, we tend to denigrate stability and regularity in the absence of comprehensive theory. We should not. Men successfully and repeatedly circumnavigated the globe using navigational theories now regarded as naive and wrong. Today we can predict with enormous accuracy the tides along the Bay of Fundy but we do not begin to understand their differences from place to place. Closer to home, we can more accurately predict the aggregate effects of a tax cut, than the incidence of it.

In arguing so I do not mean to minimize or disparage the progress of the purely theoretical debate. Even our failures are more promising now, and we have come a long way beyond the emptiness of the new welfare economies. Most encouraging is the genuine joint dialog among social scientists of different fields as well as different persuasions.

F. Concluding Comments: The Role of Efficiency

Economists traditionally place major emphasis on the efficient allocation of resources, and much of the public interest debate is also so phrased. This is particularly, but not exclusively, the case when dealing with the portion of the debate concerning the discount rate. In this paper I have stressed a multidimensional objective function rather than one involving only efficiency. Economists who disagree with this approach are not likely to do so on the grounds that efficient use of resources is the only sensible objective of social policy, nor on the grounds that efficient use of resources will never conflict with other objectives. Instead they may argue first, that any worthwhile objectives can be incorporated within an efficiency framework by an appropriate set of measurements of benefits and costs. Thus efficiency can embrace maximizing a utility function that may have several arguments. Or they may argue, second, that objectives like income redistribution not conveniently and conventionally included within the efficiency framework can and should be satisfied by lump sum taxes and transfers that do not serve to distort resource allocation.

Since either one of these arguments is, if correct, sufficient for use in the efficiency solution, the case for efficiency seems powerful and it has many adherents. Nevertheless, many economists, including this one, do not find these arguments persuasive.

The first argument has already been discussed at some length and constitutes a highly important question of research strategy: if there is a multi-dimensional objective structure, can it effectively be compressed into a single dimension by assignment of a measure of benefits? My view is that while it is possible so to compress it, it is not desirable to do so because it leads to submerging real issues behind a facade of faulty measurements. Bias can run either way: by over-

valuing intangibles (see for example the characteristically shoddy imputation of secondary benefits), or by neglecting as benefits those differences in the vector of public and private goods that are not readily measured. There is, of course, no inherently correct answer to the question of what is the best form of response to difficulties in accurate measurement. I would rather measure only what I have confidence in measuring with some accuracy and leave "incommensurables" to be decided by explicit choice. It may be easy to choose between defense and education at a given time even though it is hard to express a price that equilibrates them.[27] If one does not use a uniform system of valuations (and I am arguing that one need not) there is a danger that one will make some inconsistent decisions. But if one uses badly biased data there is a danger of making consistent but faulty decisions. As has been said in another connection, it may be better to be vaguely right than precisely wrong.

The second defense of efficiency fails to prove compelling if one believes that approximate lump sum transfers cannot be made, that they cannot be made without high transaction costs, or that they will not be made. These are factual questions but I have no real doubt about the facts. It might for example be possible to achieve indirectly for under-privileged urban residents what direct public action is today providing via urban redevelopment, and other poverty programs. But it is unlikely that in a political context the same results would have been achieved. The current effort to replace a myriad of welfare programs by a negative income tax provides a contemporary test of how easy or difficult it will prove to achieve by taxes or transfers what might otherwise be provided by direct public action. Whether for good or for ill it is frequently easier to do things one way than another. To limit the public policymaker to allocationally neutral tools constrains him, and thus changes the nature of the results he achieves. Whether the change is (as I believe) large or whether it is small is a factual matter on which one day we may have facts.

Obviously the question "What is the public interest?" has no simple answer. Indeed, asking the question invites the sort of smile reserved for small children and benign idiots. Let me end this discussion by some wholly personal assertions. There is a role for measurement, a role for analysis and some need for explicit decisionmaking. We do our decisionmakers a disservice if we blur the roles. One such blurring occurs if we submerge real decisionmaking among competing objectives into a mere measurement problem by giving the advice: "assign benefits and costs and then pick the optimal set of projects." This provides too little help. One of the economist's most potent functions is honestly to identify what can be accurately measured and compared and what (on the other hand) involves such heroics of assumption that actual measurements are but concealed preferences. The advantage of articulating real choices over assigning measures that appear to obviate them is to make the decision explicit and subject to review. But having identified the scope for explicit choice does not mean public administrators have unconstrained choice. Within particular

[27] Obviously once the dividing line between them is set, an implicit (shadow) price exists. But the price may be very different a few years hence. In any case I am arguing that it may be easier to make the decision (and thus imply the price) than in some objective sense to assign the price and thus determine the decision. Only at a purely formal level are these equivalent.

dimensions, departures from the efficient solution ought to be identified and justified.

Clearly all sorts of decisions do get made and not all of them are sensible. My conception of the analyst's role is to force an articulation of the proximate objectives served and of the conflicts between such objectives. I should be willing to regard open decisions so arrived at by elected (or otherwise responsible) public officials as a reasonable approximation to the collective values that we call the public interest. I think at present that we conceal so many issues and conflicts, both among objectives and among alternative means, that we increase the discretion of the policymaker beyond that necessary or desirable.

REFERENCES

[1] Arrow, Kenneth J., *Social Choice and Individual Values*, second edition, New York; John Wiley & Sons, 1963.

[2] Banfield, Edward C., "'Economic' Analysis of 'Political' Phenomena: A Political Scientists' Critique," Harvard Seminar on Political-Economic Decisions, March 1967 (ditto).

[3] Barker, Ernest, *Reflections on Government*, London: Oxford University Press, 1942.

[4] Baumol, William J., *Welfare Economics and the Theory of the State*, second edition, London: G. Bell & Sons, 1965.

[5] Birdsall, William C., "A Study of the Demand for Public Goods," in Richard A. Musgrave, editor, *Essays in Fiscal Federalism*, Washington: The Brookings Institution, 1965, pp. 235–292.

[6] Black, Duncan, *The Theory of Committees and Elections*, Cambridge: Cambridge University Press, 1958.

[7] Bowen, Howard R., "The Interpretation of Voting in the Allocation of Economic Resources," *Quarterly Journal of Economics*, 58:27–48, November 1943.

[8] Braybrooke and Lindblom, *A Strategy of Decision* (New York, 1963).

[9] Breton, Albert, "A Theory of the Demand for Public Goods," *Canadian Journal of Economics and Poltical Science*, 32:455–467, November 1966.

[10] Buchanan, James M., and Gordon Tullock, *The Calculus of Consent*, Ann Arbor, University of Michigan Press, 1962.

[11] ———, "Individual Choice in Voting and the Market," *Journal of Political Economy* 62:334–343, August 1954.

[12] Chenery, Hollis B., "The Application of Investment Criteria," *Quarterly Journal of Economics* 67:76–96, February 1953.

[13] Dahl, Robert A., and Charles E. Lindblom, *Politics, Economics, and Welfare*, New York: Harper & Row, 1953.

[14] Dorfman, Robert, "General Equilibrium with Public Goods," presented to International Economics Association Conference on Public Economics, September 1966 (in press).

[15] Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper & Row, 1957.

[16] Eckstein, Otto, "A Survey of the Theory of Public Expenditure Criteria," in James M. Buchanan, editor, *Public Finances: Needs, Sources and Utilization*," National Bureau of Economic Research, Princeton: Princeton University Press, 1961.

[17] Freeman, A. Myrick, III, "Income Distribution and Planning for Public Investment," *American Economic Review*, 57:495–508, June 1967.

[18] Harberger, Arnold C., "Survey of Literature on Cost-Benefit Analysis for Industrial Project Evaluation," *Inter-Regional Symposium in Industrial Project Evaluation*," sponsored by the Economic and Social Council of the United Nations, Committee for Industrial Development, Prague, October 1965 (mimeograph).

[19] Haveman, Robert H., *Water Resource Investment and the Public Interest*, Nashville: Vanderbilt University Press, 1965.

[20] Head, J. G., "Public Goods and Public Policy," *Public Finance*, 17:197–220, No. 3, 1962.

[21] Kennedy, Charles F., "The Economic Welfare Function and Dr. Little's Criterion," *Review of Economic Studies*, 20:137–142, No. 2, 1953.

[22] Knight, Frank H., "The Meaning of Freedom," in Perry, C. M., editor, *The Philosophy of American Democracy*, Chicago, 1943.

[23] Little, Ian Malcolm David, *A Critique of Welfare Economics*, second edition, Oxford: Oxford University Press, 1957.

[24] Long, Millard F., "Collective-Consumption Services of Individual-Consumption Goods: Comment," *Quarterly Journal of Economics*, 81:351–352, May 1967.

[25] Maass, Arthur, "System Design and the Political Process: A General Statement," in *Design of Water-Resource Systems*, Cambridge, Mass.: Harvard University Press, 1962, pp. 565–604.

[26] Major, David C., *Decision-Making for Public Investment in Water Resource Development in the United States*, Harvard University, Graduate School of Public Administration, Harvard Water Program, August 1965 (mimeograph).

[27] McKean, Roland N., "Cost-Benefit Analysis and British Defense Expenditures," *Scottish Journal of Political Economy*, 10:17–35; February 1963.

[28] Marglin, Stephen A., "Objectives of Water-Resource Development: A General Statement," in *Design of Water-Resource Systems*, Cambridge, Mass.: Harvard University Press, 1962, pp. 17–87.

[29] ———, *Public Investment Criteria*, Cambridge, Mass.: Massachusetts Institute of Technology Press, 1967.

[30] Margolis, Julius, "Secondary Benefits, External Economies, and the Justification of Public Investment," *Review of Economics and Statistics*, 39:284–291, August 1957.

[31] ———,"The Structure of Government and Public Investment," *American Economic Review, Papers and Proceedings*, 54:236–242, May 1964, with discussion, pp. 250–257, *ibid.*

[32] Mishan, E. J., "Criteria for Public Investment: Some Simplifying Suggestions," *Journal of Political Economy*, 75:139–146, April 1967.

[33] Musgrave, Richard A., *The Theory of Public Finance*, New York: McGraw-Hill, 1959.

[34] ———, and Peacock, Alan T., *Classics in the Theory of Public Finance*, Macmillan, 1958.

[35] Olson, Mancur, Jr., *The Logis of Collective Action*, Cambridge, Mass.: Harvard University Press, 1965.

[36] Rothenberg, Jerome, *The Measurement of Social Welfare*, Engelwood Cliffs, Prentice-Hall, 1961.

[37] ———, "A Model of Economic and Political Decision Making," Harvard Seminar on Political-Economic Decisions, March 1967 (ditto).

[38] Samuelson, Paul A., "Diagrammatic Exposition of a Theory of Public Expenditure," *Review of Economics and Statistics*, 37:350–356, November 1955.

[39] ———, "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, 36:387–389, November 1954.

[40] Schelling, Thomas C., *The Strategy of Conflict*, Cambridge, Mass.: Harvard University Press, 1960.

[41] Scitovsky, Tibor, "A Note on Welfare Propositions in Economics," *Review of Economic Studies*, 9:77–88, November 1941.

[42] Steiner, Peter O., "Choosing Among Alternative Public Investments in the Water Resource Field," *American Economic Review*, 49:893–916, December 1959.

[43] ———, "Public Expenditure Budgeting," The Brookings Institution, forthcoming.

[44] Stigler, George J., "Economics of Information," *Journal of Political Economy*, 69:213–225, June 1961.

[45] Strotz, Robert H., "Two Propositions Related to Public Goods," *Review of Economics and Statistics*, 40:329–331, November 1958.

[46] Telser, Lester G., "How Much Does it Pay Whom to Advertise?", *American Economic Review*, 51:194–205, May 1961.

[47] Truman, David B., *The Governmental Process*, New York: 1962.

[48] Tullock, Gordon, "The General Irrelevance of the General Impossibility Theorem," *Quarterly Journal of Economics*, 81:256–270, May 1967.

[49] United States, Department of Health, Education, and Welfare, *A Survey of Federal Programs in Higher Education*, 1962.

[50] Weisbrod, Burton A., "Collective-Consumption Services of Individual-Consumption Goods," *Quarterly Journal of Economics* 78:471–477, August 1964.

[51] ———, "Income-Redistribution Effects and Benefit-Cost Analysis of Government Expenditure Programs," in Chase, Samuel B., Jr., editor *Problems in Public Expenditure Analysis*. Washington: The Brookings Institution, 1968.

THE ORGANIZATION OF ECONOMIC ACTIVITY: ISSUES PERTINENT TO THE CHOICE OF MARKET VERSUS NONMARKET ALLOCATION

BY KENNETH J. ARROW

Kenneth J. Arrow is Professor of Economics at Harvard University and is associated with the Project on Efficiency of Decisionmaking in Economic Systems at that institution.

Before beginning a discussion of the role of economic analysis in Government expenditure policy one must determine that set of activities in which the public sector should properly be engaged. While the price system of the private sector is an efficient resource allocating mechanism under many conditions, it fails to function appropriately when certain conditions prevail. In this paper, Professor Arrow presents the conditions under which the private competitive market system will lead to an efficient allocation of resources and deals with those conditions in which nonmarket allocating mechanisms appear superior to the price system. The concepts of public goods, externalities, increasing returns, and transaction costs are presented as pertinent to the market versus nonmarket allocation debate. A variety of social institutions such as bargaining, the political process, and prevailing social customs are discussed as nonmarket alternatives to the price system.

Introduction

The concept of public goods has been developed through a process of successive refinement over a long period of time. Yet surprisingly enough there does not seem to exist anywhere in the literature a clear general definition of this concept or the more general one of "externality." The accounts given are usually either very general and discursive, difficult of interpretation in specific contexts, or else they are rigorous accounts of very special situations. What exactly is the relation between externalities and such concepts as "appropriability" or "exclusion"?*

Also, there is considerable ambiguity in the purpose of the analysis of externalities. The best developed part of the theory relates to only a single question: the statement of a set of conditions, as weak as possible, which insure that a competitive equilibrium exists and is Pareto efficient.[1] Then the denial of any of these hypotheses is presumably a sufficient condition for considering resort to non-market channels of resource allocation—usually thought of as Government expenditures, taxes, and subsidies.

At a second level the analysis of externalities should lead to criteria for nonmarket allocation. We are tempted to set forth these criteria in terms analogous to the profit-and-loss statements of private business; in this form, we are led to benefit-cost analysis. There are, more-

[1] A competitive equilibrium is defined below. An allocation of resources through the workings of the economic system is said to be Pareto efficient if there is no other allocation which would make every individual in the economy better off.

*Further discussion of this issue is found in the paper by Steiner in this volume.

over, two possible aims for benefit-cost analysis; one, more ambitious but theoretically simpler, is specification of the nonmarket actions which will restore Pareto efficiency; the second involves the recognition that the instruments available to the Government or other nonmarket forces are scarce resources for one reason or another, so that all that can be achieved is a "second-best."

Other concepts that seem to cluster closely to the concept of public goods are those of "increasing returns" and "market failure." These are related to Pareto inefficiency on the one hand and to the existence and optimality of competitive equilibrium on the other; sometimes the discussions in the literature do not adequately distinguish these two aspects. I contend that market failure is a more general category than externality; and both differ from increasing returns in a basic sense, since market failures in general and externalities in particular are relative to the mode of economic organization, while increasing returns are essentially a technological phenomenon.

Current writing has helped bring out the point that market failure is not absolute; it is better to consider a broader category, that of transaction costs, which in general impede and in particular cases completely block the formation of markets. It is usually though not always emphasized that transaction costs are costs of running the economic system. An incentive for vertical integration is replacement of the costs of buying and selling on the market by the costs of intrafirm transfers; the existence of vertical integration may suggest that the costs of operating competitive markets are not zero, as is usually assumed in our theoretical analysis.

Monetary theory, unlike value theory, is heavily dependent on the assumption of positive transaction costs; the recurrent complaint about the difficulty of integrating these two branches of theory is certainly governed by the contradictory assumptions made about transaction costs. The creation of money is in many respects an example of a public good.

The identification of transaction costs in different contexts and under different systems of resource allocation should be a major item on the research agenda of the theory of public goods and indeed of the theory of resource allocation in general. Only the most rudimentary suggestions are made here. The "exclusion principle" is a limiting case of one kind of transaction cost, but another type, the costliness of the information needed to enter and participate in any market, has been little remarked. Information is closely related on the one hand to communication and on the other to uncertainty.

Given the existence of Pareto inefficiency in a free market equilibrium, there is a pressure in the market to overcome it by some sort of departure from the free market; i.e., some form of collective action. This need not be undertaken by the Government. I suggest that in fact there is a wide variety of social institutions, in particular generally accepted social norms of behavior, which serve in some means as compensation for failure or limitation of the market, though each in turn involves transaction costs of its own. The question also arises how the behavior of individual economic agents in a social institution (especially in voting) is related to their behavior on the market. A good deal of theoretical literature has arisen in recent years which seeks to describe political behavior as analogous to

economic, and we may hope for a general theory of socioeconomic equilibrium. But it must always be kept in mind that the contexts of choice are radically different, particularly when the hypotheses of perfectly costless action and information are relaxed. It is not accidental that economic analysis has been successful only in certain limited areas.

Competitive Equilibrium and Pareto Efficiency

A quick review of the familiar theorems on the role of perfectly competitive equilibrium in the efficient allocation of resources will be useful. Perfectly competitive equilibrium has its usual meaning: households, possessed of initial resources, including possibly claims to the profits of firms, choose consumption bundles to maximize utility at a given set of prices; firms choose production bundles so as to maximize profits at the same set of prices; the chosen production and consumption bundles must be consistent with each other in the sense that aggregate production plus initial resources must equal aggregate consumption. The key points in the definition are the parametric role [2] of the prices for each individual and the identity of prices for all individuals. Implicit are the assumptions that all prices can be known by all individuals and that the act of charging prices is not itself a consumer of resources.

A number of additional assumptions are made at different points in the theory of equilibrium, but most are clearly factually valid in the usual contexts and need not be mentioned. The two hypotheses frequently not valid are (C), the convexity of household indifference maps and firm production possibility sets,[3] and (M), the universality of markets. While the exact meaning of the last assumption will be explored later at some length, for the present purposes we mean that the consumption bundle which determines the utility of an individual is the same as that which he purchases at given prices subject to his budget constraint, and that the set of production bundles among which a firm chooses is a given range independent of decisions made by other agents in the economy.

The relations between Pareto efficiency and competitive equilibrium are set forth in the following two theorems:

1. *If (M) holds, a competitive equilibrium is Pareto-efficient.* This theorem is true even if (C) does not hold.

2. *If (C) and (M) hold, then any Pareto-efficient allocation can be achieved as a competitive equilibrium by a suitable reallocation of initial resources.*

When the assumptions of proposition 2 are valid, then the case for the competitive price system is strongest. Any complaints about its operation can be reduced to complaints about the distribution of income, which should then be rectified by lump-sum transfers. Of course, as Pareto already emphasized, the proposition provides no basis for ac-

---

[2] By "parametric role" is meant that each household and firm takes the market prices as given, not alterable by its consumption or production decisions.

[3] For households, "convexity" means that if we consider two different bundles of consumption, a third bundle defined by averaging the first two commodity by commodity is not inferior in the household's preferences to both of the first two. For a firm, "convexity" means that if we consider two different specifications of inputs and outputs, either of which is possible to the firm (in that the inputs suffice to produce the outputs), then a third specification defined by averaging the inputs and outputs of the first two is also possible for the firm to carry out.

cepting the results of the market in the absence of accepted levels of income equality.

The central role of competitive equilibrium both as a normative guide and as at least partially descriptive of the real world raises an analytically difficult question: does a competitive equilibrium necessarily exist?

3. *If (C) holds, then there exists a competitive equilibrium.* This theorem is true even if (M) does not hold.

If both (C) and (M) hold, we have a fairly complete and simple picture of the achievement of desirable goals, subject always to the major qualification of the achievement of a desirable income distribution. The price system itself determines the income distribution only in the sense of preserving the status quo. Even if costless lump-sum transfers are possible, there is needed a collective mechanism reallocating income if the status quo is not regarded as satisfactory.

Of course (C) is not a necessary condition for the existence of a competitive equilibrium, only a sufficient one. From proposition 1, it is possible to have an equilibrium and therefore efficient allocation without convexity (when (M) holds). However, in view of the central role of (C) in these theorems, the implications of relaxing this hypothesis have been examined intensively in recent years by Farrell (1959), Rothenberg (1960), Aumann (1966), and Starr (1969). Their conclusions may be summarized as follows: Let (C′) be the weakened convexity assumption that there are no indivisibilities large relative to the economy.

4. *Propositions 2 and 3 remain approximately true if (C) is replaced by (C′).*

Thus, the only nonconvexities that are important for the present purposes are increasing returns over a range large relative to the economy. In those circumstances, a competitive equilibrium cannot exist.

The price system, for all its virtues, is only one conceivable form of arranging trade, even in a system of private property. Bargaining can assume extremely general forms. Under the assumptions (C′) and (M), we are assured that not everyone can be made better off by a bargain not derived from the price system; but the question arises whether some members of the economy will not find it in their interest and within their power to depart from the perfectly competitive price system. For example, both Knight (1921, pp. 190–194) and Samuelson (1967, p. 120) have noted that it would pay all the firms in a given industry to form a monopoly. But in fact it can be argued that unrestricted bargaining can only settle down to a resource allocation which could also be achieved as a perfectly competitive equilibrium, at least if the bargaining itself is costless and each agent is small compared to the entire economy. This line of argument originated with Edgeworth (1881, pp. 20–43) and has been developed recently by Shubik (1959), Debreu and Scarf (1963), and Aumann (1964).

More precisely, it is easy to show:

5. *If (M) holds and a competitive equilibrium prevails, then no set of economic agents will find any resource allocation which they can accomplish by themselves (without trade with the other agents) which they will all prefer to that prevailing under the equilibrium.*

Proposition 5 holds for any number of agents. A deeper proposition is the following converse:

6. *If (C') and (M) hold, and if the resources of any economic agent are small compared with the total of the economy, then, given any allocation not approximately achievable as a competitive equilibrium, there will be some set of agents and some resource allocation they can achieve without any trade with others which each one will prefer to the given allocation.*

These two propositions, taken together, strongly suggest that when all the relevant hypotheses hold, (*a*) a competitive equilibrium, if achieved, will not be upset by bargaining even if permitted, and (*b*) for any bargain not achievable by a competitive equilibrium there is a set of agents who would benefit by change to another bargain which they have the full power to enforce.

The argument that a set of firms can form a monopoly overlooks the possibility that the consumers can also form a coalition, threaten not to buy, and seek mutually advantageous deals with a subset of the firms; such deals are possible since the monopoly allocation violates some marginal equivalences.

In real life, monopolizing cartels are possible for a reason not so far introduced into the analysis: bargaining costs between producers and consumers are high, those among producers low—a point made most emphatically by Adam Smith (1937, p. 128); "People of the same trade seldom meet together, even for merriment or diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." *It is not the presence of bargaining costs per se but their bias that is relevant.* If all bargaining costs are high, but competitive pricing and the markets are cheap, then we expect the perfectly competitive equilibrium to obtain, yielding an allocation identical with that under costless bargaining. But if bargaining costs are biased, then some bargains other than the competitive equilibrium can be arrived at which will not be upset by still other bargains if the latter but not the former are costly.

Finally, in this review of the elements of competitive equilibrium theory, let me repeat the obvious and well-known fact that in a world where time is relevant, the commodities which enter into the equilibrium system include those with future dates. In fact, the bulk of meaningful future transactions cannot be carried out on any existing present market, so that assumption (M), the universality of markets, is not valid.

Imperfectly Competitive Equilibrium

There is no accepted and well-worked out theory corresponding to the title of this section. From the previous section it is clear that such a theory is forcibly needed in the presence of increasing returns on a scale large relative to the economy (hereafter, the phrase "increasing returns" will always be understood to include the prepositional phrase just employed), and is superfluous in its absence.

There are two approaches to a theory of general equilibrium in an imperfectly competitive environment; most writers who touch on public policy questions implicitly accept one or the other of these prototheories without always recognizing that they have made a choice. One assumes all transactions are made according to the price

system, i.e., the same price is charged for all units of the same commodity; this is the *monopolistic competition* approach. The alternative approach assumes unrestricted bargaining; this is the *game theory* approach. The first might be deemed appropriate if the costs of bargaining are high relative to the costs of ordinary pricing, while the second assumes costless bargaining.[4]

It cannot be too strongly emphasized that neither approach is, at the present stage, a fully developed theory, and it is misleading to state any implications about the working of these systems. Chamberlain's (1933), purpose was certainly the incorporation of monopoly into a general equilibrium system, together with a view that the commodity space should be viewed as infinite-dimensional, with the possibility of arbitrarily close substitutes in consumption; Triffin (1941) emphasized this aspect, but the only completely worked-out model of general monopolistic equilibrium is that of Negishi, (1960–61), and he made the problem manageable by regarding the demand functions facing the monopolists as those preceived by them, with only loose relations to reality. Such a theory would have little in the way of deducible implications (unless there were a supplementary psychological theory to explain the perceptions of demand functions) and certainly no clear welfare implications.

Of course, whatever a monopolistic competitive equilibrium means, it must imply inefficiency in the Pareto sense if there are substantial increasing returns. For a firm can always make zero profits by not existing; hence, if it operates, price must at least equal average cost which is greater than marginal cost. Kaldor (1935) and Demsetz (1964), however, have argued that in the "large numbers" case, the welfare loss may be supposed very small. I would conjecture that this conclusion is true, but it is not rigorously established, and indeed the model has never been formulated in adequate detail to discuss it properly.[5]

With unrestricted bargaining it is usual to conclude that the equilibrium, whatever it may be, must be Pareto-efficient for, by definition, it is in the interest of all economic agents to switch from a Pareto-inefficient allocation to a suitably chosen Pareto-efficient one. This argument seems plausible, but is not easy to evaluate in the absence of a generally accepted concept of solution for game theory. Edgeworth (1881) held the outcome of bargaining to be indeterminate within limits, and von Neumann and Morgenstern (1944) have generalized this conclusion. But when there is indeterminacy, there is no natural or compelling point on the Pareto frontier at which to arrive. It is certainly a matter of common observation, perhaps most especially in the field of international relations, that mutually advantageous agreements are not arrived at because each party is seeking to engross as much as possible of the common gain for itself. In economic affairs

[4] Within the framework of each prototheory, attempts have been made to modify it in the direction of the other. Thus, price discrimination is a modification of the price system in the pure theory of monopoly, though I am aware of no attempt to study price discrimination in a competitive or otherwise general equilibrium context. Some game theorists (Luce (1954, 1955 a, b), Aumann and Maschler (1964)) have attempted to introduce bargaining costs in some way by simply limiting the range of possilbe coalitions capable of making bargains.

[5] Suppose that the degree of increasing returns is sufficient to prevent there being more than one producer of a given commodity narrowly defined, but not to prevent production of a close substitute. Is this degree of returns sufficiently substantial to upset the achievement of an approximately perfect competitive equilibrium, as discussed in the last section?

a frequently cited illustration is the assembly of land parcels for large industrial or residential enterprises whose value (net of complementary costs) exceeds the total value of the land in its present uses. Then each owner of a small parcel whose acquisition is essential to the execution of the enterprise can demand the entire net benefit. An agreement may never be reached or may be long delayed; at positive discount rates even the latter outcome is not Pareto-efficient. It is to avoid such losses that the coercive powers of the state are invoked by condemnation proceedings.

There is, however, another tradition within game theory which argues for the determinacy of the outcome of bargaining. Zeuthen (1930, ch. IV) had early propounded one such solution. After von Neumann and Morgenstern, Nash (1950, 1953) offered a solution, which Harsanyi (1956) later showed to be identical with that of Zeuthen. Nash's analysis of bargaining has been extended by Harsanyi (1959, 1963, 1966); variant but related approaches have been studied by Shapley (1953) and Selten (1964). The analysis has proceeded at a very general level, and its specific application to resource allocation has yet to be spelled out. In the simplest situation, bargaining between two individuals who can cooperate but cannot injure each other except by withholding cooperation and who can freely transfer benefits between them, the conclusion of the theories is the achievement of a joint optimum followed by equal splitting of the benefits of cooperation net of the amounts each bargainer could obtain without cooperation. Thus, in a land assembly, if the participation of all parcels is essential, each owner receives the value of his parcel in its present (or best alternative) use plus an equal share of the net benefits of the project. Without further analytic and empirical work it is not easy to judge the acceptability of this conclusion.

An elementary example may bring out the ambiguities of allocation with unrestricted bargaining. Since the perfectly competitive equilibrium theory is satisfactory (in the absence of marketing failures and costs) when increasing returns on a substantial scale are absent, the problem of imperfectly competitive equilibrium arises only when substantial increasing returns are present. In effect, then, there are small numbers of effective participants. Suppose there are only three agents. Production is assumed to take place in coalitions; the output of each coalition depends only on the number of members in it. If the average output of the members of a coalition does not increase with the number of members, then the equilibrium outcome is the perfectly competitive one, where each agent produces by himself and consumes his own product. If the average output of a coalition increases with the number of members, then clearly production will take place in the three-member coalition; but the allocation is not determined by the threats of individuals to leave the coalition and go on their own, nor by threats of pairs to form coalitions (for any one member can claim more than one-third of the total output and still leave the other two more than they could produce without him). But perhaps the most interesting case is that where the average output is higher for two individuals than for either one or three; i.e., increasing returns followed by diminishing returns. For definiteness, suppose that one agent can produce one unit, two agents can produce four units, and all three agents together can produce five units.

Clearly, Pareto efficiency requires the joint productive activity of all three. Since each pair can receive four units by leaving the third agent out, it would appear that each pair must receive at least four units. But this implies that the total allocated to keep the three-man coalition together must be at least six, more than is available for distribution.[6]

(Theories of the Nash-Harsanyi type arrive at solutions in cases like this by assuming that the economic agents foresee these possible instabilities and recognize that any attempt by any pair to break away from the total coalition can itself be overturned. If each is rational and assumes the others are equally rational, then they recognize, in the completely symmetric situation of the example, that only a symmetric allocation is possible.)

The point of this lengthy discussion of possible game theory concepts of equilibrium is to suggest caution in accepting the proposition that bargaining costs alone prevent the achievement of Pareto efficiency in the presence of increasing returns, as Buchanan and Tullock (1962, p. 88) and Demsetz (1968, p. 61) assert.

Risk and Information*

The possible types of equilibria discussed in the previous two sections are not, in principle, altered in nature by the presence of risk. If an economic agent is uncertain as to which of several different states of the world will obtain, he can make contracts contingent on the occurrence of possible states. The real-world counterparts of these theoretical contingent contracts include insurance policies and common stocks. With these markets for contingent contracts, a competitive equilibrium will arise under the same general hypotheses as in the absence of uncertainty. It is not even necessary that the economic agents agree on the probability distribution for the unknown state of the world; each may have his own subjective probabilities. Further, the resulting allocation is Pareto-efficient if the utility of each individual is identified as his expected utility according to his own subjective probability distribution.

But, as Radner (1968) has pointed out, there is more to the story. Whenever we have uncertainty we have the possibility of information and, of course, also the possibility of its absence. No contingent contract can be made if, at the time of execution, either of the contracting parties does not know whether the specified contingency has occurred or not. This principle eliminates a much larger number of opportunities for mutually favorable exchanges than might perhaps be supposed at first glance. A simple case is that known in insurance literature as "adverse selection." Suppose, for example, there are two types of individuals, A and B, with different life expectancies, but the insurance company has no way to distinguish the two; it cannot in fact identify the present state of the world in all its relevant aspects. The optimal allocation of resources under uncertainty would require separate insurance policies for the two types, but these are clearly impossible. Suppose further that each individual knows which

[6] The general principle illustrated by this example has been briefly alluded to by Shapley and Shubik (1967, footnote 5, p. 98).

* Further discussion of this issue is found in the papers by Zeckhauser and Davis & Kamien in this volume.

type he belongs to. The company might charge a rate based on the probability of death in the two types together, but the insurance buyers in the two types will respond differently; those in the type with the more favorable experience, say A, will buy less insurance than those in type B, other things (income and risk aversion) being equal. The insurance company's experience will be less favorable than it intended, and it will have to raise its rates. An equilibrium rate will be reached which is, in general, between those corresponding to types A and B separately but closer to the latter. Such an insurance arrangement is, of course, not Pareto-efficient. It is not *a priori* obvious in general that this free market arrangement is superior to compulsory insurance even though the latter is also not Pareto-efficient because it typically disregards individual differences in risk aversion.

As the above example shows, the critical impact of information on the optimal allocation of risk bearing is not merely its presence or absence but its inequality among economic agents. If neither side knew which type the insured belonged to, then the final allocation would be Pareto-efficient if it were considered that the two types were indistinguishable; but in the above example the market allocation is Pareto-efficient neither with the types regarded as indistinguishable nor with them regarded as distinguishable.

There is one particular case of the effect of differential information on the workings of the market economy (or indeed any complex economy) which is so important as to deserve special comment: one agent can observe the joint effects of the unknown state of the world and of decisions by another economic agent, but not the state or the decision separately. This case is known in the insurance literature as "moral hazard," but because the insurance examples are only a small fraction of all the illustrations of this case and because, as Pauly (1968) has argued, the adjective "moral" is not always appropriate, the case will be referred to here as the "confounding of risks and decisions." An insurance company may easily observe that a fire has occurred but cannot, without special investigation, know whether the fire was due to causes exogenous to the insured or to decisions of his (arson, or at least carelessness). In general, any system which, in effect, insures against adverse final outcomes automatically reduces the incentives to good decisionmaking.

In these circumstances there are two extreme possibilities (with all intermediate possibilities being present): full protection against uncertainty of final outcome (e.g., cost-plus contracts for production or research) or absence of protection against uncertainty of final outcome (the one-person firm; the admiral shot for cowardice "pour encourager les autres"). Both policies produce inefficiency, though for different reasons. In the first, the incentive to good decisionmaking is dulled for obvious reasons; in the second, the functions of control and risk bearing must be united, whereas specialization in these functions may be more efficient for the workings of the system.

The relations between principals and agents (e.g., patients and physicians, owners and managers) further illustrate the confounding of risks and decisions. In the professions in particular they also illustrate the point to be emphasized later: that ethical standards may to a certain extent overcome the possible Pareto inefficiencies.

So far we have taken the information structure as given. But the fact that particular information structures give rise to Pareto inefficiency means that there is an economic value in transmitting information from one agent to another, as well as in the creation of new information. J. Marschak (1968), Hirshleifer (unpublished), and others have begun the study of the economics of information, but the whole subject is in its infancy. Only a few remarks relevant to our present purpose will be made here.

(1) As both communications engineering and psychology suggest, the transmission of information is not costless. Any professor who has tried to transmit some will be painfully aware of the resources he has expended and, perhaps more poignantly, of the difficulties students have in understanding. The physical costs of transmission may be low, though probably not negligible, as any book buyer knows; but the "coding" of the information for transmission and the limited channel capacity of the recipients are major costs.

(2) The costs of transmitting information vary with both the type of information transmitted and the recipient and sender. The first point implies a preference for inexpensive information, a point stressed in oligopolistic contexts by Kaysen (1949, pp. 294–295) and in other bargaining contexts by Schelling (1957). The second point is relevant to the value of education and to difficulties of transmission across cultural boundaries (so that production functions can differ so much across countries).

(3) Because the costs of transmission are nonnegligible, even situations which are basically certain become uncertain for the individual; the typical economic agent simply cannot acquire in a meaningful sense the knowledge of all possible prices, even where they are each somewhere available. Markets are thus costly to use, and therefore the multiplication of markets, as for contingent claims as suggested above, becomes inhibited.

Externalities Illustrated*

After this long excursus into the present state of the theory of equilibrium and optimality it is time to discuss some of the standard concepts of externality, market failure, and public goods generally. The clarification of these concepts is a long historical process, not yet concluded, in which the classic contributions of Knight (1924), Young (1913 pp. 676–684), and Robertson (1924) have in more recent times been enriched by those of Meade (1952), Scitovsky (1954), Coase (1960), Buchanan and Stubblebine (1962), and Demsetz (1966). The concept of externality and the extent to which it causes nonoptimal market behavior will be discussed here in terms of a simple model.

Consider a pure exchange economy. Let x_{ik} be the amount of the k^{th} commodity consumed by the i^{th} individual (i=1, . . ., n; k=1, . . ., (m) and $\bar{x}_k$ be the amount of the k^{th} commodity available. Suppose in general that the utility of the i^{th} individual is a function of the consumption of all individuals (not all types of consumption for all individuals need actually enter into any given individual's utility

*Further discussion of this issue is found in the papers by Davis & Kamien, and Kneese & d'Arge in this volume.

function); the utility of the i^{th} individual is $U_i(x_{i11}, \ldots, x_{imn})$. We have the obvious constraints:

$$\sum_i x_{ik} \leqq \bar{x}_k \tag{1}$$

Introduce the following definitions:

$$x_{jik} = x_{ik}. \tag{2}$$

With this notation a Pareto-efficient allocation is a vector maximum of the utility functions $U_j(x_{j11}, \ldots, x_{jmn})$, subject to the constraints (1) and (2). Because of the notation used, the variables appearing in the utility function relating to the j^{th} individual are proper to him alone and appear in no one else's utility function. If we understand now that there are n^2m commodities, indexed by the triple subscript jik, then the Pareto-efficiency problem has a thoroughly classical form. There are n^2m prices, p_{jik}, attached to the constraints (2), plus m prices, q_k, corresponding to constraints (1). Following the maximization procedure formally, we see, much as in Samuelson [1954], that Pareto efficiency is characterized by the conditions:

$$\lambda_j(\partial U_j/\partial x_{ik}) = p_{jik}, \tag{3}$$

and

$$\sum_j p_{jik} = q_k, \tag{4}$$

where λ_j is the reciprocal of the marginal utility of income for individual j. (These statements ignore corner conditions, which can easily be supplied.)

Condition (4) can be given the following economic interpretation: Imagine each individual i to be a producer with m production processes, indexed by the pair (i,k). Process (i,k). has one input, namely commodity k, and n outputs, indexed by the triple (j,i,k). In other words, what we ordinarily call individual i's consumption is regarded as the production of joint outputs, one for each individual whose utility is affected by individual i's consumption.

The point of this exercise is to show that by suitable and indeed not unnatural reinterpretation of the commodity space, externalities can be regarded as ordinary commodities, and all the formal theory of competitive equilibrium is valid, including its optimality.

It is not the mere fact that one man's consumption enters into another man's utility that causes the failure of the market to achieve efficiency. There are two relevant factors which cannot be discovered by inspection of the utility structures of the individual. One, much explored in the literature, is the appropriability of the commodities which represent the external repercussions; the other, less stressed, is the fact that markets for externalities usually involve small numbers of buyers and sellers.

The first point, Musgrave's "exclusion principle," (1959, p. 86) is so well known as to need little elaboration. Pricing demands the possibility of excluding nonbuyers from the use of the product, and this

exclusion may be technically impossible or may require the use of considerable resources. Pollution is the key example; the supply of clean air or water to each individual would have to be treated as a separate commodity, and it would have to be possible in principle to supply to one and not the other (though the final equilibrium would involve equal supply to all). But this is technically impossible.

The second point comes out clearly in our case. Each commodity (j,i,k) has precisely one buyer and one seller. Even if a competitive equilibrium could be defined, there would be no force driving the system to it; we are in the realm of imperfectly competitive equilibrium.

In my view, the standard lighthouse example is best analyzed as a problem of small numbers rather than of the difficulty of exclusion, though both elements are present. To simplify matters, I will abstract from uncertainty so that the lighthouse keeper knows exactly when each ship will need its services, and also abstract from indivisibility (since the light is either on or off). Assume further that only one ship will be within range of the lighthouse at any moment. Then exclusion is perfectly possible; the lighthouse need only shut off its light when a nonpaying ship is coming into range. But there would be only one buyer and one seller and no competitive forces to drive the two into a competitive equilibrium. If in addition the costs of bargaining are high, then it may be most efficient to offer the service free.

If, as is typical, markets for the externalities do not exist, then the allocation from the point of view of the "buyer" is determined by a rationing process. We can determine a shadow price for the buyer; this will differ from the price, zero, received by the seller. Hence, formally, the failure of markets for externalities to exist can also be described as a difference of prices between buyer and seller.

In the example analyzed, the externalities related to particular named individuals; individual i's utility function depended on what a particular individual, j, possessed. The case where it is only the total amount of some commodity (e.g., handsome houses) in other people's hands that matters is a special case, which yields rather simpler results. In this case, $\partial U_j/\partial x_{ik}$ is independent of i for $i \neq j$, and hence, by (3), p_{jik} is independent of i for $i \neq j$. Let,

$$p_{iik}=p_{ik},\ p_{jik}=\overline{p}_{jk} \text{ for } i \neq j.$$

Then (4) becomes,

$$p_{ik}+\sum_{j \neq i} \overline{p}_{jk}=q_k,$$

or,

$$(p_{ik}-\overline{p}_{ik})+\sum_{j} \overline{p}_{jk}=q_k,$$

from which it follows that the difference, $p_{ik}-\overline{p}_{ik}$, is independent of i. There are two kinds of shadow prices, a price $\overline{p}_{ik}$, the price that individual i is willing to pay for an increase in the stock of commodity k in any other individual's hands, and the premium, $p_{ik}-\overline{p}_{ik}$, he is willing to pay to have the commodity in his possession rather than someone else's. At the optimum, this premium for private possession must be the same for all individuals.

Other types of externalities are associated with several commodities simultaneously and do not involve named individuals, as in the case of neighborhood effects, where an individual's utility depends both on others' behavior (e.g., esthetic, criminal) and on their location.

There is one deep problem in the intepretation of externalities which can only be signaled here. What aspects of others' behavior do we consider as affecting a utility function? If we take a hard-boiled revealed preference attitude, then if an individual expands resources in supporting legislation regulating another's behavior, it must be assumed that that behavior affects his utility. Yet in the cases that students of criminal law call "crimes without victims," such as homosexuality or drug-taking, there is no direct relation between the parties. Do we have to extend the concept of externality to all matters that an individual cares about? Or, in the spirit of John Stuart Mill, is there a second-order value judgement which excludes some of these preferences from the formation of social policy as being illegitimate infringements of individual freedom?

Market Failure

The problem of externalities is thus a special case of a more general phenomenon, the failure of markets to exist. Not all examples of market failure can fruitfully be described as externalities. Two very important examples have already been alluded to; markets for many forms of risk-bearing and for most future transactions do not exist and their absence is surely suggestive of inefficiency.

Previous discussion has suggested two possible causes for market failure: (1) inability to exclude; (2) lack of necessary information to permit market transactions to be concluded.

The failure of futures markets cannot be directly explained in these terms. Exclusion is no more a problem in the future than in the present. Any contract to be executed in the future is necessarily contingent on some events (for example, that the two agents are still both in business), but there must be many cases where no informational difficulty is presented. The absence of futures markets may be ascribed to a third possibility: (3) supply and demand are equated at zero; the highest price at which anyone would buy is below the lowest price at which anyone would sell.

This third case of market failure, unlike the first two, is by itself in no way presumptive of inefficiency. However, it may usually be assumed that its occurrence is the result of failures of the first two types on complementary markets. Specifically, the demand for future steel may be low because of uncertainties of all types; sales and technological uncertainty for the buyer's firm, prices and existence of competing goods, and the quality specification of the steel. If, however, adequate markets for risk-bearing existed, the uncertainties could be removed, and the demand for future steel would rise.

Transaction Costs*

Market failure has been presented as absolute, but in fact the situation is more complex than this. A more general formulation is that

* Further discussion of this issue is found in the paper by Demsetz in this volume.

of transaction costs, which are attached to any market and indeed to any mode of resource allocation. Market failure is the particular case where transaction costs are so high that the existence of the market is no longer worthwhile. The distinction between transaction costs and production costs is that the former can be varied by a change in the mode of resource allocation, while the latter depend only on the technology and tastes, and would be the same in all economic systems.

The discussions in the preceding sections suggest two sources of transaction costs. (1) exclusion costs; (2) costs of communication and information, including both the supplying and the learning of the terms on which transactions can be carried out. An additional source is (3) the costs of disequilibrium; in any complex system, the market or authoritative allocation, even under perfect information, it takes time to compute the optimal allocation, and either transactions take place which are inconsistent with the final equilibrium or they are delayed until the computation are completed (see T. Marchak, 1959).

These costs vary from system to system; thus, one of the advantages of a price system over either bargaining or some form of authoritative allocation is usually stated to be the economy in costs of information and communication. But the costs of transmitting and especially of receiving a large number of price signals may be high; thus, there is a tendency not to differentiate prices as much as would be desirable from the efficiency viewpoint; for example, the same price is charged for peak and offpeak usage of transportation or electricity.

In a price system, transaction costs drive a wedge between buyer's and seller's prices and thereby give rise to welfare losses as in the usual analysis. Removal of these welfare losses by changing to another system (for example, governmental allocation on benefit-cost criteria) must be weighed against any possible increase in transaction costs (for example, the need for elaborate and perhaps impossible studies to determine demand functions without the benefit of observing a market).

The welfare implications of transaction costs would exist even if they were proportional to the size of the transaction, but in fact they typically exhibit increasing returns. The cost of acquiring a piece of information, for example, a price, is independent of the scale of use to which it will be put.

Collective Action: The Political Process

The State may frequently have a special role to play in resource allocation because, by its nature, it has a monopoly of coercive power, and coercive power can be used to economize on transaction costs. The most important use of coercion in the economic context is the collection of taxes; others are regulatory legislation and eminent domain proceedings.

The State is not an entity but rather a system of individual agents, a widely extensive system in the case of a democracy. It is appealing and fruitful to analyze its behavior in resource allocation in a manner analogous to that of the price system. Since the same agents appear in the two systems, it becomes equally natural to assume they have the same motives. Hotelling (1929, pp. 54–55) and Schumpeter (1942, ch. XXII) had sketched such politicoeconomic models, and von Neu-

mann and Morgenstern's monumental work is certainly based on the idea that all social phenomena are governed to essentially the same motives as economics. The elaboration of more or less complete models of the political process along the lines of economic theory is more recent, the most prominent contributors being Black (1958), Downs (1957), Buchanan and Tullock (1962), and Rothenberg (1965).

I confine myself here to a few critical remarks on the possibilities of such theories. These are not intended to be negative but to suggest problems that have to be faced and are raised by some points in the preceding discussion.

1. If we take the allocative process to be governed by majority voting, then, as we will know, there are considerable possibilities of paradox. The possible intransitivity of majority voting was already pointed out by Condorcet (1785). If, instead of assuming that each individual votes according to his preferences it is assumed that they bargain freely before voting (vote-selling), the paradox appears in another form, a variant of the bargaining problems already noted in section 2. If a majority could do what it wanted, then it would be optimal to win with a bare majority and take everything; but any such bargain can always be broken up by another proposed majority.

Tullock (1967) has recently argued convincingly that if the distribution of opinions on social issues is fairly uniform and if the dimensionality of the space of social issues is much less than the number of individuals, then majority voting on a sincere basis will be transitive. The argument is not, however applicable to income distribution, for such a policy has as many dimensions as there are individuals, so that the dimensionality of the issue space is equal to the number of individuals.

This last observation raises an interesting question. Why, in fact, in democratic systems has there been so little demand for income redistribution? The current discussion of a negative income tax is the first serious attempt at a purely redistributive policy. Hagström (1938) presented a mathematical model predicting on the basis of a self-interest model for voters that democracy would inevitably lead to radical egalitarianism.

2. Political policy is not made by voters, not even in the sense that they choose the vector of political actions which best suits them. It is in fact made by representatives in one form or another. Political representation is an outstanding example of the principal-agent relation. This means that the link between individual utility functions and social action is tenuous, though by no means completely absent. Representatives are no more a random sample of their constituents than physicians are of their patients.

Indeed, the question can be raised: to what extent is the voter, when acting in that capacity, a principal or an agent? To some extent, certainly, the voter is cast in a role in which he feels some obligation to consider the social good, not just his own. It is in fact somewhat hard to explain otherwise why an individual votes at all in a large election, since the probability that his vote will be decisive is so negligible.

Collective Action: Social Norms

It is a mistake to limit collective action to State action; many other departures from the anonymous atomism of the price system are

observed regularly. Indeed, firms of any complexity are illustrations of collective action, the internal allocation of their resources being directed by authoritative and hierarchical controls.

I want, however, to conclude by calling attention to a less visible form of social action: norms of social behavior, including ethical and moral codes. I suggest as one possible interpretation that they are reactions of society to compensate for market failures. It is useful for individuals to have some trust in each other's word. In the absence of trust, it would become very costly to arrange for alternative sanctions and guarantees, and many opportunities for mutually beneficial cooperation would have to be foregone. Banfield (1958) has argued that lack of trust is indeed one of the causes of economic underdevelopment.

It is difficult to conceive of buying trust in any direct way (though it can happen indirectly, for example, a trusted employee will be paid more as being more valuable); indeed, there seems to be some inconsistency in the very concept. Nonmarket action might take the form of a mutual agreement. But the arrangement of these agreements and especially their continued extension to new individuals entering the social fabric can be costly. As an alternative, society may proceed by internalization of these norms to the achievement of the desired agreement on an unconscious level.

There is a whole set of customs and norms which might be similarly interpreted as agreements to improve the efficiency of the economic system (in the broad sense of satisfaction of individual values) by providing commodities to which the price system is inapplicable.

These social conventions may be adaptive in their origins, but they can become retrogressive. An agreement is costly to reach and therefore costly to modify; and the costs of modification may be especially large for unconscious agreements. Thus, codes of professional ethics, which arise out of the principal-agent relation and afford protection to the principals, can serve also as a cloak for monopoly by the agents.

REFERENCES

Aumann, R. J. 1964. Markets with a continuum of traders. *Econometrica* 32:39–50.

Aumann, R. J. 1966. The existence of competitive equilibria in markets with a continuum of traders. *Econometrica* 34:1–17.

Aumann, R. J. and Maschler, M. 1964. The bargaining set for cooperative games. In M. Dresher, L. S. Shapley, A. W. Tucker (eds). *Advances in Game Theory. Annals of Mathematics Study*, Princeton, New Jersey: Princeton, University Press, 52:443–476.

Banfield, E. C. 1958. *The Moral Basis of a Backward Society.* The Free Press.

Black, D. 1958. *The Theory of Committees and Elections.* Cambridge, U.K.: Cambridge University Press.

Buchanan, J. and W. C. Stubblebine. 1962. Externality. *Economica* 29:371–384.

Buchanan, J. and G. Tullock. 1962. *The Calculus of Consent*, Ann Arbor, Michigan: University of Michigan Press.

Chamberlain, E. H. 1933. *The Theory of Monopolistic Competition.* Cambridge, Mass. Eighth edition, 1965.

Coase, R. H. 1960. The problem of social cost. *Journal of Law and Economics* 3:1–44.

Condorcet, Marquis de. 1785. *Essai sur l'application de l'analyse à la probabilitié des décisions rendues à la pluralité des voix.* Paris.

Debreu, G. and H. Scarf. 1963. A limit theorem on the core of an economy. *International Economic Review* 4:235–246.

Demsetz, H. 1964. The welfare and empirical implications of monopolistic competition. *Economic Journal* 74:623–641.

Demsetz, H. 1966. Some aspects of property rights. *Journal of Law Economics* 9:61–70.

Demsetz, H. 1968. Why regulate utilities. *Journal of Law and Economics* 11:55–66.

Downs, A. 1957. *An Economic Theory of Democracy.* New York: Harper.

Edgeworth, F. Y. 1881. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences.* London: C. Kegan Paul & Co.

Farrell, M. J. 1959. The convexity assumption in the theory of competitive markets. *Journal of Political Economy* 67:377–391.

Hagstrom, K. G. 1938. A mathematical note on democracy. *Econometrica* 6:381–383.

Harsanyi, J. C. 1956. Approaches to the bargaining problem before and after the theory of games: A critical discussion of Zeuthen's, Hicks', and Nash's Theories. *Econometrica* 24:144–157.

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 IV. Annals of Mathematics Study,* Princeton, New Jersey: Princeton University 40:325–355.

Harsanyi, J. C. 1963. A simplified bargaining model for the n-person cooperative game. *International Economic Review* 4:194–220.

Harsanyi, J. C. 1966. A general theory of rational behavior in game situations. *Econometrica.* 34:613–634.

Hotelling, H. 1929. Stability in competition. *Economic Journal* 39:41–57.

Kaldor, N. 1935. Market imperfection and excess capacity. *Economica, N.S.* 2:35–50.

Kaysen, Carl. 1949. Basing point pricing and public policy. *Quarterly Journal of Economics* 63:289–314.

Knight, F. H. 1921. *Risk, Uncertainty, and Profit.* Boston and New York: Houghton-Mifflin. Reprinted by London School of Economics and Political Science 1948.

Knight, F. H. 1924. Some fallacies in the interpretation of social cost. *Quarterly Journal of Economics* 38:582–606.

Luce, R.D. 1954. A definition of stability for n-person games. *Annals of Mathematics* 59:357–366.

Luce, R.D. 1955a. Ψ-stability: a new equilibrium concept for n-person game theory. In *Mathematical Models of Human Behavior.* Stamford, Conn.: Dunlap and Associates, pp. 32–44.

Luce, R.D. 1955b. k-stability of symmetric and quota games. *Annals of Mathematics* 62:517–555.

Marschak, J. 1968. Economics of inquiring, communicating, deciding. *American Economic Review Papers and Proceedings* 58:1–18.

Marschak, T. 1959. Centralization and decentralization in economic organizations. *Econometrica* 27:399–430.

Meade, J.E. 1952. External economies and diseconomies in a competitive situation. *Economic Journal* 62:54–67.

Musgrave, R.A. 1959. *The Theory of Public Finance: A Study in Public Economy.* New York: McGraw-Hill Book Company.

Nash, J.F.jr. 1950. The bargaining problem. *Econometrica* 18:155–162.

Nash, J.F.jr. 1953. Two person cooperative games. *Econometrica* 21:128–140.

Negishi, T. 1960–1. Monopolistic competition and general equilibrium. *Review of Economic Studies* 28:196–201.

von Neumann, J., and O. Morgenstern. 1944. *Theory of Games and Economic Behavior.* Princeton, N.J.: Princeton University Press. Second edition, 1947.

Pauly, M.V. 1968. The economics of moral hazard: comment *American Economic Review* 58:531–537.

Radner, R. 1968. Competitive equilibrium under uncertainty. *Econometrica* 36:31–58.

Robertson, D.H. 1924. Those empty boxes. *Economic Journal* 34:16–30.

Rothenberg, J. 1960. Non-convexity, aggregation, and Pareto optimality. *Journal of Political Economy* 68:435–468

Rothenberg, J. 1965. A model of economic and political decision-making. In J. Margolis (ed.) *The Public Economy of Urban Communities.* Washington, D.C.: Resources for the Future.

Samuelson, P.A. 1954. The pure theory of public expenditures. *Review of Economic Statistics* 36:387–389.

Samuelson, P.A. 1967. The monopolistic competition revolution. In R.E. Kuenne (ed.) *Monopolistic Competition Theory: Studies in Impact.* New York, London, and Sydney: Wiley, pp. 105–138.

Schelling, T. 1957. Bargaining, communication, and limited war. *Journal of Conflict Resolution* 1:19–36.

Schumpeter, J. 1942. *Capitalism, Socialism, and Democracy.* New York: Harper. Third Edition, 1950.

Scitovsky, T. 1954. Two concepts of external economies. *Journal of Political Economy* 62:143–151

Selten, R. 1964. Valuation of n-person games. In M. Dresher, L. S. Shapley, A. W. Tucker (eds.) *Advances in Game Theory.* Princeton, New Jersey; Princeton University Press, 52:577–626.

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 II. Annals of Mathematics Study.* Princeton, New Jersey: Princeton University Press, 28:307–317.

Shapley, L.S., and M. Shubik. 1967. Ownership and the production function. *Quarterly Journal of Economics* 81:88–111.

Shubik, M. 1959. Edgeworth market games. In A.W. Tucker and R.D. Luce (eds.) *Contributions to the Theory of Games IV. Annals of Mathematics Study.* Princeton, New Jersey: Princeton University Press, 40:267–278.

Smith, A. 1937. *An Enquiry Concerning the Causes of the Wealth of Nations.* New York: Modern Library.

Starr, R. 1969. Quasi-equilibria in markets with nonconvex preferences. *Econometrica* 37:25–38.

Triffin, R. 1941. *Monopolistic Competition and General Equilibrium Theory.* Cambridge, Mass.: Harvard University Press.

Tullock, G. 1967. *Toward a Mathematics of Politics.* Ann Arbor, Michigan: University of Michigan Press.

Young, A.A. 1913. Pigou's Wealth and Welfare. *Quarterly Journal of Economics* 27:672–686.

Zeuthen, F. 1930. *Problems of Monopoly and Economic Warfare.* London, George Routledge & Sons Ltd.

Section B

External Effects (Spillovers) and Inadequate Market Performance

EXTERNALITIES, INFORMATION AND ALTERNATIVE COLLECTIVE ACTION

BY OTTO A. DAVIS and MORTON I. KAMIEN

Otto A. Davis and Morton I. Kamien are both Professors of Economics at Carnegie-Mellon University.

One of the reasons why the operation of an unfettered market system may fail to serve the public interest is the inability of markets to accommodate certain kinds of side effects. Another reason for market failure is that buyers and sellers often lack the quantity and the quality of information necessary for them to choose effectively when engaging in market transactions. The authors of this paper deal with both the problem of side effects or spillovers which are not accommodated in markets, and with the problem of inadequate market information. In dealing with the problem of inadequate information, they cite drug products as an example of a commodity whose distribution requires collective or governmental action as a supplement to the market place. The problems of air and water pollution are employed to illustrate the market failure entailed by side effects or externalities. In the case of spillovers, the authors point out that a number of collective solutions to market failure are available. They discuss solution by prohibition, by directive, by voluntary action, by taxes and subsidies, by regulation, by payment, and by direct public action.

Introduction

Awareness that an action often entails subsidiary as well as direct consequences is commonplace. In choosing an occupation we consider not only the direct monetary remunerations involved but also the security, power and prestige associated with the various endeavors. When purchasing wearing apparel we take into account its attractiveness as well as the protection and comfort which it affords. In the use of drugs we should be acutely conscious of their possibly harmful side effects as well as of their direct curative powers. When purchasing a house we are likely to take into account in addition to the size and age the quality of the neighborhood in which it is located, its proximity to good schools, and public transportation facilities.

In everyday parlance we refer to these secondary attributes of products or actions as "side-effects," "fringe-benefits," or "occupational disease." Our concern with these matters is not wasted on the advertising industry which promotes many products by stressing their desirable side-benefits. Witness the number of advertisements that allude to the masculinity, femininity, youthfulness, and glamour that are to be derived from the use of this or that product. Indeed, some products and occupations have become better known by their side effects than by their direct benefits.

Of course, concern with secondary consequences is not confined to the advertising industry alone. This regard for side-effects finds expression in the selection of products or occupations and the amounts we are willing to pay or sacrifice to avoid or incur them. In other words, many kinds of side effects are accommodated by the market sys-

tem. For example, a desire to live in a "better" neighborhood manifests itself by a willingness to pay more for a house in the preferred neighborhood than for a comparable house elsewhere. Likewise, a strong preference for a relatively secure occupation is satisfied by a willingness to sacrifice potentially higher monetary gains in other occupations. Businessmen find it profitable to be responsive to these desires of consumers. Though the primary function of an automobile is transportation, manufacturers provide a wide variety of models to satisfy the secondary features desired by purchasers. Drug producers attempt to develop new drugs that possess the same beneficial properties as existing ones while reducing undesirable side-effects. Moreover, the responsiveness of producers is spurred by the knowledge that competitors will cater to the preferences of customers. It is for this reason that competition among producers is thought to be desirable. Similarly, competition among buyers assures that goods and services will be allocated in conformity with the relative desires and abilities of the participants to pay.

From the above argument one might be tempted to conclude that a freely competitive economy should provide the goods and services desired by consumers in such a way as to preclude the possibility that another allocative mechanism (such as government) might be judged to be more appropriate for given situations. Given certain conditions and a plausible criterion, one of the major contributions of modern economic theory is the confirmation of this conclusion. Yet, even the most casual observation of the real world discloses that our society often takes recourse to collective governmental action for the provision of certain goods and services. One could allow such a casual observation to bring one to the conclusion that modern economic theory must be either wrong or irrelevant for the real world. While there probably is considerable sympathy for such a conclusion in some circles, it is taken here to be an obviously incorrect deduction. An alternative explanation of this apparent divergence between theory and reality is that collective decisions are necessarily bad and that governments act in a nonsensical manner. This conclusion too is rejected here. Of course, the rejection of this alternative does not imply that governments always make the wisest or best decisions. Instead, the view adopted here is that the governmental decision process can be improved and that we should do what we can to improve it. Another explanation of the apparent divergence between theory and reality—and this one is accepted here—is that the conditions or assumptions upon which the above conclusion about the efficiency of a market system is based are not always satsified in reality. According to this viewpoint, there may be some advantage in the study of particular aspects of economic theory since such a study might produce considerable insight into the detection of situations where market systems cannot be expected to work very well. The mere existence of such situations raise the problem of selecting the proper institutional arrangement under which the activity under consideration may be conducted, and one would be foolish to believe that this problem can be solved at this time in a way which might produce a consensus. Yet, there is considerable advantage in merely knowing where markets might, and where they might not, work tolerably well.

Given the above discussion, the plan of this essay is to present informally the major conclusion of modern welfare economics; a theorem about the allocation of resources by a market mechanism.

Particular attention will be paid to certain of the assumptions upon which this theorem is based. Examples will be used to help make the major points clear. Neither will the problem of selecting proper criteria for institutional choice be overlooked. It is hoped that the outcome of the discussion will be a better understanding of some of the issues and difficulties involved in selecting institutional mechanisms which are capable of attaining an acceptable allocation of limited resources.

The Criterion, the Market, and Optimality *

Obviously, if one is to talk meaningfully about a choice among alternatives, one must have in mind some kind of method for ordering or weighing the various possibilities. Economists have been rather explicit in their choice of an abstract criterion. It is the notion of efficiency or, to use the more technical term, the concept of Pareto optimality. The basic idea is that a situation is inefficient or non-optimal if it is possible to make at least one member of a society better off without making any other member worse off. In other words, a situation is Pareto optimal if it is impossible to make anyone better off without the same time making someone else worse off. It is worth while pointing out that this criterion of efficiency or Pareto optimality need not lead to an unambiguous ordering of alternatives since, theoretically speaking, there exists at least an infinity of positions which are Pareto optimal. On the other hand, there is good reason to insist within the limits of practicality that all solutions be efficient since, by definition, a non-Pareto optimal solution means that someone can be made better off without making anyone else worse off. The qualifying phrase "within the limits of practicality" is used here to denote the fact that although the theoretical possibility of improving at least one person's position without inflicting harm on anyone else must be admitted whenever the situation is not Pareto optimal, the practical means of actually accomplishing such an improvement need not be at all obvious to the frail minds of humans.

The notion of Pareto optimality probably would be neither interesting nor useful were it not for some of the developments of modern welfare economics. The most important of these developments can be viewed as one of the central theorems of economics. It can be stated informally as follows: Given certain assumptions about the technology, the availability of information, the characteristics of goods and services, and the absence of monopoly power, then there exists a set of market prices such that profit maximizing firms and utility maximizing consumers which respond to those prices will automatically cause the economic system to attain a Pareto optimum position. This theorem is, of course, a powerful argument for the organization of our society so that exchange takes place through the mechanism of competitive markets. If the assumptions of the theorem were universally satisfied, then the Government could limit itself largely to programs aimed at the attainment of a desirable distribution of income and be rather certain that the vehicle of competition would cause the system to be efficient.

---

* Further discussion of this issue is found in the papers by Steiner and Arrow in this volume.

There is little, if any, need to review the entire set of assumptions which appear to be required for the above theorem to obtain. Indeed, economists have long been searching for a minimal set of assumptions which will be sufficient for markets to attain Pareto optimality, and it is doubtful that the end of the search is anywhere in sight. Accordingly, it is appropriate to review here only those which seem to cause the greater part of the difficulty in the real world. Fortunately or unfortunately, it will be seen that these difficulties appear to be interrelated in ways that are not always obvious.

Consider first the technology, The assumption here is that all firms have convex production possibility sets. What this supposition means is that there must be an absence of increasing returns to scale.* In other words, it must not be true that ever larger firms can produce the same product at a lower per unit cost than can relatively smaller ones. Of course, it is recognized that there can be increasing returns over a range as long as that range is not significant in terms of industry output.

A second consideration is the availablity of information. Producers are assumed to have knowledge of the available technology. Consumers are supposed to know whether particular goods and services are available as well as their characteristics. More will be said about this below, but it is obvious that this is a heroic assumption. Finally, both producers and consumers are presumed to know the relevant set of prices.

The third condition concerns the characteristics of the goods and services which are to be produced by the economic system. First, not only are there supposed to be no "public" goods—that is, goods such as radio waves or television signals which are noted for the fact that when one listener or viewer "uses" them via reception, the quantity available for use by other persons is not diminished—but the consumption of other goods and services (called "private" goods) is not supposed to directly affect decision units who are not doing the consuming. In other words, although the "side effects" mentioned in the introduction are allowed, there is supposed to be an absence of what in technical language goes under the name of nonpecuniary externalities. Since a large part of this essay is devoted to these external effects, a detailed discussion of the phenomena is postponed. One should note here, however, that the mere presence of externalities is not a sufficient reason for the market to avoid optimality but such a presence is a danger signal which should not go unnoticed.

A final condition worth noting is the absence of monopoly power. It is the competitive market which, under certain circumstances, is supposed to be capable of attaining Pareto optimality. Since relatively little discussion of monopoly will be presented in this essay, it is worthwhile to point out here that monopoly is often related to the other conditions under consideration. For example, it is acknowledged that one of the difficulties associated with increasing returns is that the logical consequence of attempted competition is the emergence of monopoly. Similarly, initial monopoly power can sometimes be maintained due to the fact that technological knowledge is not always available to all and, even when it is, there may be barriers in the form of difficulties in transmission and assimilation.

---

* Further discussion of this issue is found in the paper by Vickrey in this volume.

All of the above assumes that markets either do or can be made to exist. Unfortunately, such does not always seem to be the case.

The remainder of this essay will concentrate upon the dual considerations of information and externality with an emphasis upon the latter. However, the discussion itself will be testimony to the fact that all of these matters are interwoven and cannot always be separated.

Problems of Information

While theoretical discussions traditionally have assumed, as was noted above, that participants in the economic system have full knowledge, there has been a widespread awareness that this assumption is never fully satisfied. Some recent works have taken cognizance of the fact that information is both scarce and costly. Obviously, some kinds of information are more easily acquired than other types and recognition of this continuum of costs is helpful in any attempt to understand the functioning of the market. As an illustration imagine a consumer faced with the problem of selecting a detergent. In the terminology of marketing such a purchase is called a "repeat sale" which emphasizes the fact that the consumer purchases this type of product on a weekly or monthly basis. Detergents and items that are purchased often offer a minimal informational problem for the consumer. There is little cost in "trying out" various brands until one is identified whose characteristics appear most suited for the individual task at hand. The advertising often gives information about the product's characteristics such as its cleaning power or sudsiness. Even here, however, it is clear that laws which promote truth in advertising can help the market perform its proper function since without such regulation advertising might not be a source of information with any reliability.

Items that are not purchased very often, such as consumer durables, offer a more difficult informational problem. Clearly, it is not always practical for the consumer to learn the characteristics of the brands in this class of products by simply trying out various items until one with the desired properties is identified. Even here, for relatively expensive items which are easily transported and which are sufficiently complex to make it desirable for a trained person to instruct the customer in the operating methodology, some companies find it profitable to allow potential customers to keep the item for a trial period during which he can learn at least some of the characteristics of the machine. When this practice is followed, however, salesmen seldom make it easy for the customer to make systematic comparisons among available brands. Hence, other sources of information must be relied upon.

Clearly, one source is often informal and casual conversations in which experiences with various brands are related from one person to another. Another source is advertising, and it would appear that regulations designed to prevent misleading claims from being made are even more important for this class of goods than for the one discussed above since the absence of frequent and periodic purchases make it less easy for the customer to gain comparative information from his own experience. A third source is independent "testing laboratories" such as Consumer's Union. Several points are relevant here. Any serious or casual reader of Consumer's Report knows that for every major product, with the possible exception of automobiles, the information

which can be presented is so scanty and model specific, with many models not being examined, that there is simply not sufficient information for a careful comparison and choice among the models that might be available. Now this statement is not intended as a criticism of the management of Consumer's Union. The point is that knowledge of the characteristics of a product is very analogous to a public good. Once the characteristics are known, then the "consumption" of this knowledge by any one shopper in making his choice does not diminish either the availability or usefulness of that knowledge for any other shopper. All of the well-known difficulties of trying to market a useful public good are relevant here. Although there are obvious costs associated with the production of knowledge of a product's characteristics, the producer of that knowledge cannot hope to recoup anything even approximating its value to the consumer. Neither should one suggest that even if it were possible for him to do so—which is not possible since the information can be transmitted easily from one person to another—should he actually do it since the optimality conditions would require that the transmission take place at marginal cost which is trivially close to zero. This situation means, however, that the producer of this knowledge cannot afford to produce as much information as it would be socially optimal to provide both in terms of the quality of knowledge that could be produced and in terms of the range of coverage of models of the various products.

The above point should not be taken as a criticism of the provision of this class of goods, which is mainly durables, by the private sector. Even the governmental provision of this class of goods would not alter in any basic way the problem under consideration. What is suggested, however, is that a certain kind of governmental regulation can serve to improve the functioning of an otherwise unregulated market. Producers are certainly in a better position than anyone else to know the characteristics of their products. It would appear to be relatively simple to have groups of persons who are knowledgeable about the various products, and who know what characteristics one should consider in making a purchase, draw up a list of these features for each product category. Manufacturers would then be responsible for making this information available to those who deal with the public and these, in turn, could be required to furnish the information to prospective purchasers. Claims about products could be checked in the way that advertising is now regulated. Note that this proposal differs from the existing situation where producers are motivated to furnish the public only that information which is favorable to their product. No one would claim, of course, that information concerning all of the relevant characteristics can be made available. Even the manufacturer may not know, for example, the expected life of a new kind of machine.

There is another class of goods and services where the informational problems can be orders of magnitude greater than those discussed above. The salient features of this class are: (1) Information about the relevant characteristics, even when it is available, is difficult to understand, interpret, and evaluate without the benefit of special training; (2) the consequences of an incorrect choice can be serious to the extent that there is an order of magnitude difference in terms of real costs between this and previous cases. Probably, the best example of a product category belonging to this class is some of the drugs which

are produced. The usual incentives of the marketplace operate here as elsewhere in the sense that consumers should be willing to pay more, ceteris paribus, for a drug which is safer than others which might be available. Consequently, manufacturers should have an incentive to produce safer drugs. The obvious informational difficulty here, however, is the very problem of identifying which drug is safer. It is difficult to consider seriously the possibility of consumers having to use various drugs in order to determine experimentally both their curative powers and their safety when possible side effects may not be immediate and may be irreversible.

While private "testing laboratories" might be relied upon for information, with all the difficulties previously discussed about this kind of institution, it is easy to see why our society has made the collective decision that the Food and Drug Administration should regulate this market under the law. Thus, given the fact that a drug is on the market, one can be certain that manufacturers have at least minimally tried to determine side effects and associated dangers. Of course, given the difficulties associated with the attempt to make such a determination, there obviously can be no certainty that there are no harmful side effects as is illustrated by the example of the Thalidomide case of a few years ago. This case happened under the additional safeguard which society has imposed of allowing particular drugs to be used only when a prescription certifies that one is under the care of a physician.

One rationale for allowing certain drugs to be sold only under prescription is, of course, informational costs. Supposedly, the physican has the necessary information at his fingertips and can exercise his professional judgment in administering any drug which is known to have undesirable side effects for some (unidentified) portion of the population. Presumably, there is little need to mention that the arrangement does not work perfectly, but one of the authors has been pointedly reminded several times of the imperfection when a serious attack of asthma while in a strange town has prompted visits to local physicians who had to be told not only of the available range of drugs but also advised as to which particular one had the properties which made it most desirable to be prescribed for the given condition. In addition, some of the institutional practices are not designed to stimulate price competition between drug manufacturers so that the market is certainly noncompetitive and may be monopolistic. Yet, one of the rationales of this particular institutional arrangement is the costs, which would be associated with the acquisition of information under alternative arrangements so that the monopolistic costs of the present arrangement must be weighed against the costs associated with the conceivable alternatives, which includes modifications, in making an institutional choice.

Problems Associated With Externalities*

All of the above difficulties and the institutional arrangements which have been or might be designed to deal with them are related for the greater part to the impact or effect which given items might

*Further discussion of this issue is found in the paper by Kneese & d'Arge in this volume.

have upon the consumer who purchased them. Therefore, it is only natural to inquire about possible effects of the purchase and consumption of these given items upon other persons or decision units who were not parties to the exchanges. Effects upon those external to or not associated with specified purchases or activities appropriately are called externalities. Alternative terms are called spillovers, external effects, or social effects.

While the literature distinguishes many kinds of externalities, it is necessary for the purposes of this essay to identify only two types. These are technological (or nonpecuniary) and pecuniary externalities.

Let us first deal with the concept of the pecuniary externality. When deciding whether or not to purchase an item an individual will ordinarily take into account his own desire for the item, its price, and his budgetary situation. It will be rare indeed, and generally only in the case of a monopsony, that the individual might even consider that his decision to purchase can contribute to and maybe even increase the demand for that product and thereby cause its price to rise. Of course, in most instances the individual's purchase of a commodity is such a small fraction of the total amount sold that his decision has a negligible impact upon price, although the totality of decision is certainly of importance. Whenever an individual decision does have an effect upon price, it is important to note that not only does he, but also all other purchasers, have to pay the resulting increase or decrease. This change in price, caused by individual decisions, is termed a pecuniary externality. If the individual decision causes the price to rise, which is the usual case associated with an increase in demand, then the phenomenon is a pecuniary external diseconomy to other consumers. Whenever the decision causes the price to fall, which might be illustrated by a decision to join a group travel arrangement which is not yet at capacity, then the phenomenon is termed a pecuniary external economy to other consumers. Of course, by symmetry, a pecuniary external diseconomy to consumers, is a pecuniary external economy to sellers, while a pecuniary external economy to consumers is a diseconomy to sellers.

The important point to note here, however, is that pecuniary externalities, be they economies or diseconomies, pose no problem for the market economy. Indeed, they are the central ingredient of the marketplace. Changing demands cause prices to rise and fall, generally according to whether demand increases or decreases, and the resulting alterations on prices are the essential feature of a marketplace which rations the available goods and services to those whose willingness to pay indicates that they need them most.

Technological externalities are quite another matter. These refer to more or less direct effects, which are not priced, which one decision unit might impose on another. Technological externalities can, and in many instances do, prevent the marketing mechanism from functioning in such a manner as to lead the economic system to a position of Pareto optimality. In such instances, of course, there exists the theoretical possibility that action can be taken to improve the society in the sense that one or more citizens can be made better off without anyone being made worse off. Some examples may serve to illustrate what is at issue here.

Since both of the authors reside in Pittsburgh, it may be appropriate to begin with the example of the manufacture of steel. For the purpose of exposition, imagine that there is no smoke control ordinance. Then, according to the process which is employed, more or less smoke may be discharged into the atmosphere as a byproduct of steel production. Insofar as the manufacturer is interested in profits—and most are—there is motivation to choose that method of production which is most profitable without regard for the associated level of the discharge of smoke. The point is that the manufacturer can be thought of as envisioning the opportunity to dispose of smoke as another resource which contributes to the production of steel. The justification of viewing disposal as another resource is that a reduction in the discharge of smoke could only be achieved by either adopting an alternative and more expensive method of production which emits less smoke or by using the same process but with the addition of smoke control devices. Either alternative involves the use of additional resources such as labor and capital. While these additional resources are not free, there is no charge for the emission of smoke into the atmosphere so that there is little if any motivation to attempt to limit the usage of the resource which might be called smoke disposal.

Although the discharge of smoke into the atmosphere might be viewed as a free resource by the firm, it is certainly not without consequence to those residing within adjoining communities. Not only does smoke contribute to the more rapid deterioration of the exteriors of buildings and certain kinds of equipment—which will certainly mean that compensatory resources will have to be spent in more intensive cleaning, maintenance, and repair—but it certainly contributes to smog which probably has a direct, though not yet fully documented, effect upon the health of at least some of the residents of the community. In other words, to the community at large the discharge of smoke into the atmosphere is not a free resource. Instead, smoke disposal is costly. Pigou would say that such a situation, where the firm does not bear the full costs of its actions, is an instance where private costs diverge from social costs. The essential point to notice about the situation as it has been outlined here, however, is that without some kind of action the steel producer has nothing more than possible humanitarian concerns, which conflict with his interests in profits, to make him take into account the fact that the discharge of smoke imposes costs upon his neighbors. The discharge of smoke is a technological externality. Without some kind of adjustment the system will not be at a Pareto optimum so that there exists the theoretical possibility that at least one person may be made better off without making anyone else worst off.

Of course, smoke is not the only cause of smog. One of the most often mentioned contributors today is the automobile. In order to understand fully the nature of the relevant motivations, imagine the situation prior to the establishment of the regulation which requires that smog control devices be installed upon all new cars. It is obvious that if consumers demand and are willing to pay for smog control devices, the automobile industry would develop and sell these devices in much the same way that it develops and sells special conveniences and optional equipment. The competition among the various manu-

facturers, foreign and domestic, compels the producers to try to give the public what it wants. Would the public demand smog control devices? The answer can be found by examining the consumer's motivation.

Imagine for the sake of argument that the auto industry had developed an effective smog control device which it offered as optional equipment for all new cars. A person who was considering whether or not to order this optional for his new car might reason as follows: Suppose I purchase the smog control devices for my new car. If I purchase and everyone else also purchases, then we will have less smog in the city. On the other hand, my individual car can add only a negligible amount to the smog problem so that if everyone else purchases a device and I do not do so, then the smog will be diminished by almost exactly the same amount and I will have saved the cost of the device. Hence, if everyone else purchases a device, I will be better off if I do not get one installed on my car. Now presume that no one, with the possible exception of myself, purchases a device. Obviously, there will be a smog problem. However, if I purchase a device the problem will not be noticably different since my individual car contributes only negligibly to the situation and I will be out of the money which I paid for the smog control device. Hence, if no one else purchases, I should not purchase either. Obviously, the analysis is the same if some of the other people purchase and some do not. Conclusion: I will be better off, no matter what other people do, if I do not purchase a smog control device.

Since all potential new-car buyers will reason roughly as the representative individual above, the result is that there will be a zero demand for smog control devices. Hence, in the absence of some kind of regulation or collective decision, the automobile manufacturers will have no motivation to develop and market smog control devices. This conclusion holds even if—and it is an if—everyone would be better off if all cars were equipped with smog control devices. The point is that for each prospective purchaser of a device, the benefits from his purchase are widely dispersed while the costs accrue to him. Thus the technological externality associated with the exhaust of a car can prevent the unregulated market from leading the system to a Pareto optimum.

For the final example of this section, consider the problem of the pollution of Lake Erie. Biologists tell us that Lake Erie is dying and that it has "aged" 15,000 years in the past half century. The problem is complex. It was long believed that the major source of the pollution stemmed from the fact that raw sewage and industrial wastes are dumped into the lake. A major source of the raw sewage is antiquated systems, some of which are combined sanitary and storm sewers so that the overflow runs into the lake during periods of rain. For the moment, and for the purpose of discussion, imagine that the entire problem of pollution is caused by the raw sewage so that treatment, which could remove the organic material which otherwise is broken down in the lake by a biological process which consumes its oxygen, could solve the problem. The now familiar dilemma would act to frustrate a pure market solution. Each municipality or sewage district would reap but little of the benefits of its own efforts at treating the sewage, but it would bear the full costs of that treatment. Hence, similar to the above

case of a customer considering the purchase of a smog control device, each would come to the rational decision to continue to allow the raw sewage to flow into the lake even though all might be better off if all installed modern systems with treating devices. Thus the technological externality reflected in reverse in the failure to receive the full benefit of one's expenditure for treatment—the fact that the decisionmaking entity does not bear the full costs of its decision to forego treatment and allow a flow of raw sewage—results in a failure of a pure market solution where no financial incentive to come to the opposite decision is offered from a higher level of government.

In actuality the pollution of Lake Erie is a much more complex phenomenon than is indicated by the above discussion. Even after treatment to remove indigestible solids and to break down organic material so that the sewage is discharged as mostly inorganic products, the residual inorganic matter contains large amounts of nitrate and phosphate which, instead of being swept harmlessly to the sea, tend to remain in the lake long enough to fertilize monstrous growths of algae which use up to an estimated 18 times as much oxygen as the present flow of organic matter from inadequate sewage plants. Thus, the standard treatment of sewage, which is aimed at the organic matter, is not likely to solve the problem even if such treatment were accomplished. One might suggest that one of the "essential" nutrients such as the phosphate should be removed from the waste so that the algae would not grow, and this suggestion brings us to the economics of the situation. Some two-thirds of the phosphorus in municipal waste, which is roughly three-quarters of the total wastes, stems from detergents. Even if the housewife or commercial laundry knew that the detergent used for the wash contributed importantly to the pollution of the lake, which they probably do not know, would there be any incentive to economize on the use of detergents or demand a kind which contained less phosphorus? Again, the familiar dilemma appears. Even if they knew of their contribution to pollution, each could rationalize that their own contribution was negligible, that the benefits to be derived from an individual decision to try to perform the wash in such a manner as to contribute less phosphorus to the sewage was too small to be measured, that the costs of this kind of action was not negligible, so that the rational decision would be to ignore the entire situation.

Thus, the manufacturers of detergents would have no incentive to try to develop products which contain less phosphorus, the municipal sewage systems would have no more incentive than in the previous instance to attempt to remove the material, and the result is that the pure market solution would be to continue the pollution of the lake. Thus, the existence of this technological externality—the fact that those causing the pollution do not bear the full costs of their actions—can cause the market mechanism to lead the system to a situation which is not Pareto optimal.

The Possibility of a Solution: *

The above illustrations are indicative of the fact that the very existence of technological externalities can cause problems for the claims of the efficiency of the unregulated market mechanism. It must be ad-

* Further discussion of this issue is found in the paped by Schultze in this volume.

mitted that the problems caused by these technological externalities are most perplexing. Although only a few years ago economists may have been of the opinion that an adequate solution was available, the consensus now has vanished. Instead, one finds that a variety of solutions have been suggested by various persons, many of whom are not economists, and a goodly number of these have even been tried or implemented in certain situations.

What seems to have been present in many of the discussions and analyses of technological externalities in the literature is a belief that a universal solution might be found. Thus, proposals are often treated as if they are supposed to be "the" solution to the problem of technological externalities. Unfortunately, this belief may turn out to be unfounded, which is the authors' own belief, and it may be that there is no simple and universally acceptable solution to the problem. It may be that, at least for the foreseeable future, our society may have no alternative but to seek pragmatic solutions to the problem.

In keeping with the above, the proper perspective requires that attention be devoted to a consideration of many of the suggested solutions to the problem. Accordingly, the remainder of this essay is devoted to an examination of some of the proposed solutions. It is desirable, of course, that the point of view which is adopted in this discussion be made explicit. The reader is thus warned that the authors view the situation as one which can be conceptually accommodated to cost-benefit analysis. All of the proposals have associated costs and benefits. The problem is to identify which is most suitable for a given situation. It is hoped that the following discussion may be helpful in this regard. The plan is to discuss the more popular of the proposed solutions. Examples will be used to illustrate all of the major points.

Solution by Prohibition

When one is convinced that collective action is necessary to try to improve the situation, or to correct the abuses caused by technological externalities, the first thought that one is likely to have is that action should be taken to prohibit the externalities. After all, if the externalities are prohibited, will not the market system then function so that our economy will be brought to a Pareto optimum position?

Although this course of action may seem appealing at first, it takes little thought to convince one that simple prohibition of activities causing the technological externalities is a poor approach. Obviously, one could not seriously propose that car owners stop driving, that steel manufacturers stop producing, or that municipalities stop disposing of their sewage. Some might think, however, that we should have perfectly clean water or perfectly clean air, so that full treatment is desired. Such a thought, however, misses the fundamental point. Optimality does not require that externalities be eliminated. Instead, optimality requires that externalities be present in the "right amount." Some examples may make this point more clear.

Consider the case of water pollution. Natural biological processes in both lakes and streams give them certain capability of cleaning themselves up. Thus if absolutely no wastes are allowed to flow into these waters, and if all sewage is made "perfectly" clean, this natural capability will not be used. The proper way of viewing this natural capa-

bility is as a resource and, as is true for all scarce and valuable resources, it should be used. In addition, it would be very expensive to make sewage perfectly clean, and benefits must always be balanced against the costs. The problem of pollution arises when so much wastes are dumped into a stream or lake that the capacity of the water to clean itself up without producing objectionable side effects is exceeded.

Air pollution affords a similar example. It would be prohibitively expensive to prevent any contaminants whatsoever from escaping into the air. Further, there is no reason not to use the natural absorbic capability of the atmosphere. The problem of air pollution arises when there are excesses.

The above examples should make clear the fact that strict prohibition of whatever causes a technological externality is insurance that the economic system cannot attain Pareto optimality. What is desired is just the right amount of the externality. So, for example, in the case of water pollution a Pareto optimal solution may in fact call for some deterioration of water quality in certain streams and may even mean complete deterioration of water quality in other streams.

Solution by Directive

Having seen that part of the problem is to get just the right amount of the technological externality, it is tempting to say that the easiest procedure would be to let the Government decide how much of it should be produced. This procedure would involve, for example, governmental determination of just how well the municipalities bordering Lake Erie should treat their sewage in terms of, say, the percentage of organic matter removed and phosphorous content, and an absolute quantity limit above which the sewage would have to be treated until it was pure. Similarly, in the air pollution example the Government would have to specify just how much smoke each factory could emit.

There are several difficulties with this procedure. First, there is the problem of determining just how much of the externality is desirable. This might be called the problem of the overall standard. It is not to be dismissed as a trivial problem. There is, needless to say, a theoretically correct way to determine the standard. It should be set by a careful weighing of costs and benefits. To be specific, consider again the example of the pollution of Lake Erie. For the purpose of illustrating the main point, imagine that it would cost 50 billion dollars per year in operating costs alone to process the sewage in such a manner that the pollution in the lake would diminish from present levels. Now while there may be many benefits to be derived from an unpolluted Lake Erie, it is rather doubtful that they would be valued at anything near 50 billion dollars per year. Hence, in such a case the rational decision would be to tolerate an even higher level of pollution and the standard should be set in such a manner that there would be no increase in the level of processing the sewage. The costs would simply outweigh the benefits. On the other hand, imagine that all of the sewage could be processed if there were an increase in annual expenditures of only $5. Clearly, the yearly benefits from an unpolluted lake would exceed this figure so that the standard should be set so that all sewage would be treated. The benefits would outweigh

the costs. In between these extremes, however, the computations become very difficult. The difficulties of determining the benefits associated with various degrees of pollution are almost insolvable. Hence, there must be a great deal of arbitrariness in setting the overall standard. One of the difficulties here is that the tools for measuring benefits are rather crude. Another problem, which has been overlooked in the above discussion, is that relatively little is known about the ecology of lakes so that there is a great deal of uncertainty concerning what effects treatment might have upon pollution.

Even if the overall standard could be easily determined instead of involving the difficulties discussed above (the problem of measuring benefits is most crucial here), other difficulties would still remain. The overall standard must be translated into directives for each of the entities which emit pollutants. In principle the directives should be adjusted so that the marginal effectiveness of the last dollar spent upon the processing of wastes should be equated for all of the pollutors. In practice the marginal effectiveness of dollars spent for treating wastes probably cannot be determined since for any given pollutor, the effectiveness depends upon the policies which the other pollutor follow. In other words, there would be arbitrariness at this level, too.

If one takes another example, such as air pollution, the problems are even more difficult although the principles involved are the same. The overall standard should still be determined by weighing and comparing the benefits and the costs of the various possibilities. However, the uncertainties are even greater since one does not know, for example, the exact relationship between the level of pollution and the health of the residents. Also, there is the fact that in most urban areas the amount of pollutants which can be released into the atmosphere for any given standard depends upon the weather and especially the prevailing winds.

None of the above comments should be interpreted as meaning that the policy of controlling externalities by directives is to be dismissed as being obviously inappropriate. What is intended here is to point out that there are difficulties associated with the procedure. Further, one should observe that in addition to the above, this particular approach also involves an administrative cost of policing the directives, which cannot be ignored.

Solution by Voluntary Action

Some argue that collective action is not needed to correct the market solution when there are technological externalities. It has often been pointed out that there is motivation for private parties to act to correct the situation by a variety of methods. Two which are often discussed are the methods of bribes and merger. These will be discussed in order.

Consider once more the example of a steel producer discharging smoke into the atmosphere unchecked by a smoke control ordinance. The previous discussion indicated that this situation potentially gives rise to a divergence between private and social cost or between the private and social benefit of steel production. To avoid the adverse effects

of smoke discharge the community might resort to bribing the steel producer to decrease or discontinue altogether the discharge of smoke. The rationale for this behavior is that as long as the amount of the bribe needed to induce the steel manufacturer to reduce smoke discharge is less than the damage inflicted on the community, then the community will on net be better off by paying the bribe. Of course, the community acting in its self-interest should never offer a bribe that exceeded the value of the damage inflicted via smoke discharge. The steel producer should in turn accept or reject the bribe in accordance with his best interests. Accordingly, if the bribe exceeded the amount he would have to spend on means to reduce smoke discharge, he should accept the bribe and effect the desired reduction; and if the costs were too great, he should not. In any case a quantitative measure of the damage suffered by the community from smoke would have been presented to the manufacturer and in such a way as to make him cognizant of this figure when deciding how much smoke to discharge.

Moreover, whatever the final level of smoke discharge is it could be Pareto optimal if there were such a thing as perfect bargaining. One can reason as follows: Acceptance of the bribe by the manufacture indicates that he is at least as well off as before, while payment of the bribe by the community indicates that it is at least as well off as before. Consequently, the situation is improved and if the bargaining were perfect any departure from the agreed upon position would only improve the position of one of the parties at the expense of the other. It is also true that rejection of the bribe by the manufacturer under perfect bargaining leads to a Pareto optimal solution. By rejecting the bribe the manufacturer would disclose that the value of this resource (release of smoke into the air) is of greater value to him than to the community.

The method for avoiding a divergence between private and social cost just described is purely voluntary and leads to a Pareto optimal allocation of resources when bargaining is perfect. It would, therefore, appear to the ideal way in which to resolve such problems. Unfortunately, bargaining is not perfect and there are several impediments to its widespread use. The first difficulty is associated with the valuation of smoke damage suffered by the community. The most direct way of estimating the damage is to ask each member of the community how much he would be willing to contribute to the bribe to be offered to the manufacturer. In principle, each individual should be willing to contribute the amount he would have to expend to avoid the damage from smoke by other means. Unfortunately, however, the individual may realize that if he contributes nothing toward the bribe while others contribute positive amounts and smoke abatement is effected, he will reap the benefits of smoke abatement at no cost to himself. If all members of the community adopt this attitude, no bribe will be offered and the scheme will fail. In other words, the public good nature of the benefit from smoke abatement impedes the realization of the necessary collective action by the community. The second difficulty with the bribe procedure is that it requires that the community know all the available methods for manufacturing steel, as these are related to smoke control, and the associated costs so that they might prevent the manufacturer from cheating. For suppose after the bribe has been accepted by the manufacturer, the demand for steel rises

and output increases. The producer can now legitimately argue that a larger bribe is required for him to maintain the previously agreed upon level of smoke discharge. Unless the community is completely knowledgeable about steelmaking technology it cannot be sure that the manufacturer is not expanding his output more than would be optimal for him in the absence of the bribe. Thus, a seemingly ideal scheme for avoidance of a divergence between private and social cost is marred by difficulties in implementation.

Another voluntary scheme for interalizing nonpecuniary externalities free of some of the implementation difficulties mentioned above is the merger of the entities involved when this is possible. To illustrate how this procedure might work, consider the following situation: Suppose a firm discharges wastes into a stream which are harmful to fish life. Suppose, further, that a fishery operates downstream from the firm. In the absence of any governmental regulation the upstream firm will discharge waste into the stream without regard to the damage inflicted on the fishery in the way of smaller catches or tainted fish. Were the firm and the fishery to merge under a single ownership, then it would be in the new consolidated firm's best interest to take account of the losses incurred by its downstream subsidiary as a consequence of the actions of its upstream plant. The consolidated firm should balance the cost of disposing the waste at the upstream plant by means other than discharge into the stream against the costs incurred by the downstream fishery as a result of waste discharge into the stream in such a way as to maximize the combined profit from the two operations. Since in this example Pareto optimality corresponds to joint profit maximization by the two entities, merger will assure a Pareto optimal allocation of the resource in question;—viz, the stream. It might also be noted that the profit of the consolidated firm will always be at least as great as the combined profits of the two firms operating in isolation. The reason for this is that the merged firm always has the option of adopting the operating policies of the two firms working independently and consequently can achieve a profit level at least as large as the combined profit of the previously independent firms. The difference between the profits of the consolidated firm and the combined profits of the individual firms reflects the loss to society from the presence of a nonpecuniary externality.

Two difficulties with the merger solution can be pointed out. The first is the entities have to be firms. The second in that merger is feasible only when the number of entities involved is small. As the number of decision making units to enter into the consolidation increases the chances of effecting the merger decline. This is primarily due to the fact that it becomes increasingly more difficult to persuade potential participants that it is in their best interest to join the coalition. Individual units may find it profitable to postpone entry into the coalition so as to extract a larger portion of the joint profits from the merged entity. The third difficulty is that a merged entity might become so big as to cause a distortion in the allocation of resources via monopoly or monopsony power. In this case the losses from the presence of nonpecuniary externalities have to be weighed against the losses to society from the resource allocation distortions engendered by imperfectly competitive markets.

Solution by Taxes and Subsidies*

If voluntary arrangements among the entities effected by nonpecuniary externalities are impractical or not forthcoming, collective governmental action might be justified. In the economics literature the classic form of government intervention in this situation is the payment of a subsidy to units that by their actions confer external economies upon other units and the levying of taxes upon those entities that by their actions confer external diseconomies upon other units. In essence, the idea is to encourage those activities that contribute to the "common good" and discourage those that detract from the "common good".

To illustrate the working of this scheme, consider again the example of an upstream firm and a downstream fishery. Suppose now, however, that the waste discharged into the stream by the firm provides food for fish in the stream and is therefore beneficial to the fishery. Since by hypthesis voluntary negotiation by the parties involved is ruled out here, the fishery has no way of communicating the magnitude of the benefits it derives from the firm's discharge of waste into the stream. Consequently, the amount of food provided to the fish may not be ideal. In this case a government subsidy to the firm for the discharge of waste can in principle be devised to achieve the desired result. Likewise, if as in our earlier description of the situation the waste discharged is harmful to fish life, then in principle a tax can be imposed on the firm that reflects the damage imposed on the fishery.

We have stressed the "in principle" nature of these conclusions because of the immense informational requirements necessary for the implementation of this scheme. A little reflection will make it apparent that the governmental agency imposing a tax or offering a subsidy will need to know the production technologies of all the entities involved. In effect the governmental agency will have to solve the same problem that the directors of the merged firm discussed above solve. Instead of issuing orders regarding the quantities of each product that each of the subsidiaries should produce so as maximize joint profits, a practice that the executives of the merged firm might follow, the agency would attempt to achieve the same results via the payment of subsidies and/or the issuance of taxes. Suffice it to say the information that would be available to the directors of the merged firm is rarely available to an outside governmental agency. Bits of this information may of course be available and it may be possible at a cost to obtain additional information.

The amount of information required also depends on the nature of the productive technologies involved. Less information is required by the agency for the successful implementation of a tax-subsidy scheme if the underlying productive technologies are separable or additive than if they are not. For example, if the cost of producing the upstream firm's product and the cost of waste treatment are additive then the tax on waste discharge, if that is what is called for, simply depends on the amount of waste discharge. On the other hand if these costs are not additive then the tax must vary not only in accordance with the level of waste discharge but also with the amount of the

*Further discussion of this issue is found in the paper by Knetsch in vol. 3 of this collection.

firm's primary product. Of course the information requirements mount enormously as the number of economic entities involved increases. Despite all these difficulties an attempt to achieve optimal resource allocation via taxes and subsidies might be justified if the losses to society from the presence of nonpecuniary externalities is large enough. In essence what has to be balanced in this situation is the cost of acquiring the needed information against the losses to society if nothing is done or another of the imperfect policies is followed.

Solution by Regulation

Another collective action which is often suggested is governmental regulation. For example, the official governmental response to the fact that cars contribute to air pollution in our cities has been to reduce the range of consumer choice by simply requiring that all new cars be equipped with devices which are supposed to reduce the level of pollutants that escape from the engine. Such a regulation obviously permits an escape from the dilemma described earlier where a rational calculation would cause the consumer to refrain from purchasing a control device.

Regulation also has problems associated with it. There is often uncertainty associated with the imposition of a regulation. In the case of automobiles, for example, there is uncertainty as to whether the devices will be effective in reducing the level of pollutants which are discharged, especially as the cars grow old. There are also problems of enforcement. For instance, there is speculation that if the devices are effective and the pollutants are kept in the engine rather than being spewed into the atmosphere, the life of the engine is likely to be shortened and repairs will have to be made more frequently. These are costs of the regulation which, along with the costs of the devices, have to be weighed against possible benefits which, as was discussed earlier, are rather difficult to compute. Note, however, that if the devices do have the anticipated effect upon engines, then each owner has an incentive to take action to render the devices ineffective and to allow the pollutants to escape into the atmosphere since by doing so he can increase the life of his engine and reduce his repair bills. Certainly, the owner cannot be expected to keep the device in good working order, or to repair it when it breaks, since such an action would not be in his own self-interest. Hence, the regulation cannot be expected to be successful in reducing air pollution, even if the devices work, unless it is accompanied by the practice of periodically inspecting all cars and requiring that the devices be maintained in good working order.

The above discussion should make clear the fact that solution by regulation is not as simple as might first seem to be the case. The administrative costs of enforcing the regulations are relevant and should not be overlooked. Regulation differs from solution by directive in that under the latter scheme government prescribes the possibly different activity level of each economic entity involved while under the regulation alternative uniform requirements are imposed upon all entities or only uniform permissible bounds on certain activities are set.

Neither should one ignore the fact that a regulatory solution is of necessity inflexible. This point too can be better understood by recourse

to the above example. Many of our Nation's motor vehicles are operated for a considerable portion of their life in nonurban areas where air pollution is not thought to be a problem. Ideally, vehicles operating in these areas should not be required to have smog control devices so that the natural ability of the atmosphere to accommodate a certain level of pollutants could be utilized. Obviously, it is impossible to design regulations which would accomplish this ideal due to the very mobility of motor vehicles and the population. Hence, for many externalities solution by regulation is inherently incapable of bringing the system to a Pareto optimal solution because regulations are by nature inflexible.

Solution by Payment

One of the obvious ways of trying to accommodate the system to the presence of technological externalities is to attempt to provide a financial incentive for the desired actions to be taken. In the pollution of Lake Erie, for example, one of the problems is supposed to be the fact that Cleveland has an archaic sewage system which combines storm and sanitary sewers. The capacity of the facilities for treatment is sufficiently limited that anything but modest rainfall is supposed to create such a flow that the capacity of treating facilities are exceeded, allowing raw sewage to flow directly into the lake. As was explained earlier, the citizens of Cleveland do not bear the full cost of their archaic sewage system since, although it certainly contributes to the pollution of Lake Erie, persons residing outside of Cleveland desire to use the lake for various purposes. These other persons thus bear part of the costs of pollution including that portion of the pollution which stems from Cleveland's sewage system. Of course, what is true for Cleveland is also true for many of the other cities and towns in the lake area. Since each does not bear the full costs of the results from its own system, none have full incentive to remedy the situation.

One possible policy step is for a higher level of government—for example, the Federal Government—to provide an incentive for a remedy. Such a provision can be accomplished, at least in part, by Federal subsidy for the capital costs of improving the sewage facilities. Such a subsidy is an incentive to the local community.

One difficulty with this measure of policy is its crudeness. It does not easily provide proper coordination for all of the relevant units in the system. Another limitation of this policy is that it is suited only for those kinds of externalities, such as the one discussed here, where the capital costs are the only really significant feature which prevents the situation from being improved.

Solution by Direct Public Action

It sometimes happens that there are simple direct actions which can be taken to ameliorate the effects of an externality. Probably the clearest example involves fishing. Consider a lake or stream where many people come to fish. Now at least after some level of activity, the catch begins to affect the future fish population. Thus, when a fisherman makes a catch, he can affect the future population of fish

and thus lower the pleasure and profits of other fishermen. No individual fisherman, of course, has incentive to take into account the effect which his own activity has upon others. In extreme cases, of course, the population of fish could be exhausted.

An obvious remedy for this situation is for the Government to continually stock the lake or river so that the fish population is never diminished to the danger point. The externality is then more or less eliminated by this direct action.

Needless to say, there are obvious problems associated with this policy of direct action. Not the least of these is the fact that it is suitable only for very special situations.

Concluding Remarks

It should be obvious by now that there is a whole menu of policies which can be fashioned to deal with problems caused by technological externalities. None of these policies, at least at our present level of knowledge, appears to be perfect. Neither does it seem to the authors that any one clearly dominates in the sense that it would be the best of the imperfect lot for each and every technological externality. Thus it is argued here that policies must be designed with particular situations in mind and what is best for one externality may be inappropriate for another. Accordingly, it is appropriate to conclude this essay with a few remarks about what seems to be the appropriate procedure for the selection of policy.

The tools of cost-benefit analysis appear to provide the proper perspective. In a given situation, the policymaker should consider the problem and imagine the application of each of the alternative approaches to it. The principle of selection is simple. Each measure of policy (including that of doing nothing) will have costs and benefits associate with it. The policymaker should select that measure for implementation which produces the greatest net benefits.

PERVASIVE EXTERNAL COSTS AND THE RESPONSE OF SOCIETY

BY ALLEN V. KNEESE and RALPH C. d'ARGE

Allen V. Kneese is the Director of the Quality of the Environment Program at Resources for the Future, Inc. He is currently on leave of absence at the University of New Mexico. Ralph C. d'Arge is Assistant Professor of Economics at the University of New Mexico.

Drs. Kneese and d'Arge argue that "widespread concern with air and water pollution, urban congestion, landscape deterioriation, and other environmental impacts of economic growth" is evidence that externalities and spillovers are *pervasive* phenomena in our society. Because these effects are present in most of the production and consumption process of our economy, they cannot be viewed as "somewhat freakish anomalies" which disturb an otherwise smoothly functioning market system. For this reason, they argue that most of the present approaches advocated by social scientists to deal with this problem are of limited usefulness.

To effectively analyze this pervasive kind of external effect, the authors outline a more comprehensive model. This model provides a framework to assist public decisionmakers in developing an appropriate collective response to these pollution-type problems. They argue that coherent environmental controls must be devised such that decentralized voluntary exchange can again dominate resource allocation and conclude by commenting on the implications of their analysis for both public policy and research. The rigorous statement of the models on which their paper is based are presented in two technical appendices.

I. *Introduction*

"MANY PERSON" EXTERNALITIES

In the past few years externalities have received unprecedented attention in the economics literature. Much has been done to systematize definitions and clarify policy issues.[1] But almost the entire literature is devoted to "two party" externalities and references to more pervasive effects are rare. In preparing his recent book, Buchanan found almost no literature of theoretical interest regarding "many person" externalities.[2]

This situation exists despite widespread concern with air and water pollution, urban congestion, landscape deterioration, and other environmental impacts of economic growth. These external costs affect groups of various sizes up to and including the population of the entire earth. While highly appreciative of the substantial contributions which have recently been made, we cannot help but feel that they do not provide a fully suitable framework for dealing effectively and efficiently with today's externality problems.

[1] Two excellent survey articles have recently appeared. See E.J. Mishan, "Reflections on Recent Developments in the Concept of External Effects," *The Canadian Journal of Economics and Political Science*, February 1965, and R. Turvey, "On Divergencies Between Social Cost and Private Cost, *Economica*, November 1962.

[2] See James M. Buchanan, *The Demand and Supply of Public Goods*, Rand McNally & Co., Chicago 1968, p. 188.

Our perspective in the present paper is drawn on the theoretical side from the "second best" literature and on the more applied side from a consideration of the particular problem of environmental pollution. We feel that the conventional partial equilibrium approaches may amount to efforts at fine tuning of a system which is grossly out of focus. Also we believe that the concept of consumption which is at least implicit in most recent economic literature may have contributed to a failure to develop economic theory in the form best suited to deal with problems of widespread environmental pollution. In short, it is our view that external costs cannot realistically be treated in the traditional fashion, as somewhat freakish anomalies which may affect isolated parts of an otherwise smoothly working economic system. Instead, our view is that they are inherent in the production and consumption processes of highly developed economies. In this context it becomes a function of government to adjust the framework for voluntary economic exchanges so that they may more realistically be thought to lead to efficient resources allocation.

In the next section we discuss environmental pollution and explain why we feel that the present economic approaches to externalities are of limited usefulness with respect to it. In the following section we outline a model combining aspects of economic general equilibrium analysis and a material balance approach drawn from the concept of mass conservation. We argue that a model of this type could be useful to government in devising coherent sets of environmental quality constraints or standards. Following this we discuss the role of decentralized decision making, as well as governmental initiatives when environmental considerations (pervasive externalities) are introduced into an otherwise competitive economy. In the last part we interpret the previous sections in terms of policy and research considerations. The next section and appendix I draw on parts of an earlier paper.[3] All mathematical development of theoretical models used is consigned to two appendices.

II. Services and Material Flow

Standard economic theory is concerned with services which yield flows of utilities, not with physical substances. Material objects are merely the vehicles which transmit some of these services. They are exchanged because of consumer preferences for the services directly associated with their use or because they can help to add value in the manufacturing process.[4] Yet we persist in referring to the "final consumption" of goods as though material objects such as fuels, materials, and finished goods somehow disappeared in the void—a practice which was comparatively harmless so long as air and water and landscape, the residuals-receiving media, were almost literally "free goods." Of course, residuals from both the production and consumption processes remain; and they usually return to the environment to render disservices (like killing fish, increasing the difficulty of water treatment, reducing public health, soiling and deteriorating buildings, despoiling

[3] See Robert U. Ayres and Allen V. Kneese, "Production, Consumption, and Externalities," *American Economic Review*, June 1969.

[4] Both Fisher and Knight cautioned economists that the basic magnitude to which economics reasoning applies is, in Knight's words, "service not good." See I. Fisher, *Nature of Capital and Income*, New York, 1906, and F. H. Knight, *Risk, Uncertainty and Profit*, Boston and New York, 1906.

landscapes, etc.). Control efforts (often involving large public expenditures) are aimed at eliminating or reducing those disservices which flow to consumers and producers whether they want them or not, and which, except in unusual cases, they cannot control by engaging in individual exchanges. It is important to understand, however, that "treatment" of waste residuals as such does not reduce the amount of residuals, but merely changes their form. Only reuse of waste materials or changes in the durability of products can reduce the total weight of residuals associated with a given level of production of services.

To elaborate on these points, we find it useful initially to view environmental pollution and its control as a materials balance problem for the entire economy. The physical inputs to the productive system are fuels, foods, and raw materials which are partly converted into "final goods" and partly become waste residuals. Except for increases in inventory, "final goods" also ultimately enter the residuals stream. Thus articles which are "consumed" really only render certain services. Their material substance remains and must either be reused or discharged to the ambient environment.

In an economy which is closed (no imports or exports) and where there is no net accumulation of stocks (plant, equipment, inventories, consumer durables, or residential buildings), the amount of residuals which is inserted into the natural environment must be approximately equal to the weight of basic fuels, food, and raw materials taken from the environment and entering the processing and production system (plus oxygen taken from the atmosphere). This result, while obvious upon reflection, leads to the at first rather surprising corollary that residuals disposal involves a greater tonnage of materials than basic materials processing. In chart I materials flow through the economy is presented schematically. In other papers order of magnitude estimates of residuals tonnages, based on materials balance calculations, have been provided for the U.S. economy as a whole.[5] In open economies, such as those of regions, the import and export of materials would, of course, affect the amount of residuals discharged to the immediate environment of the area where production and consumption occur.

Generally speaking, the residuals which are discharged to the environment result in external costs on a regional scale in the form of air, water, and land pollution. However, regions affected by discharges to the various environmental media are by no means always coterminous. Sometimes (junk in a backyard is an example) only a few nearby parties are affected. At the other extreme some experts are concerned about the effect on worldwide weather patterns of rising levels of carbon dioxide in the atmosphere and the adverse results of oxygen depletion from the whole atmosphere.[6]

---

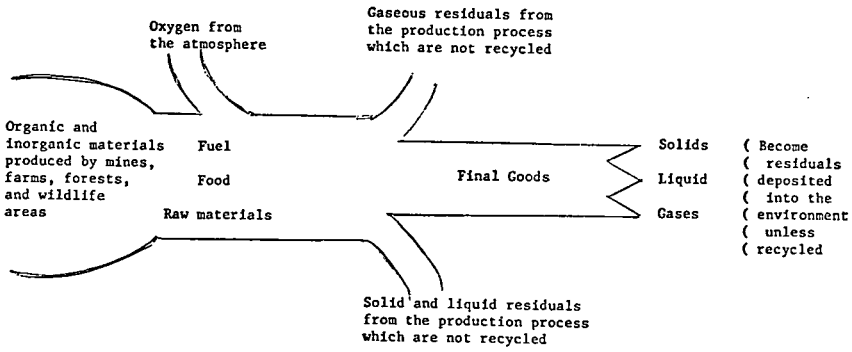
[5] See particularly R. U. Ayres and A. V. Kneese, "Environmental Pollution," in U.S. Congress, Joint Economic Committee, *Federal Programs for the Development of Human Resources*, vol. 2, Washington, 1968.

[6] See, for example, L. Cole, "Can the World Be Saved?" paper presented at the 134th meeting of the American Association for the Advancement of Science, Dec. 27, 1967. It is disquieting to hear meteorological experts disputing whether coastal cities will be drowned because the icecaps melt due to the "green house" effect of rising CO_2, or whether Siberia and Canada will become Arctic wastes due to the "opacity" effect of particulate discharges to the atmosphere.

Defining a proper region for environmental analysis and management is a deep matter with which we shall not concern ourselves in this paper. For present purposes we will simply assume that appropriate regions can be defined.

Chart I

Schematic of the Goods-Residuals Production Process



Given the population, industrial production, and transport services in an economy, different combinations of social policy would lead to quite different relative burdens placed on the various residuals-receiving environmental media; or, given the possibilities for recycle and less residual-generating production processes, the overall burden to be placed upon the environment as a whole.[7]

Moreover, it is frequently possible to improve an environmental medium through collective investment in large-scale facilities. One example is reservoir storage to augment low riverflows which ordinarily are associated with critical pollution (high external cost situations). Many public investments, as those in transportation facilities, have incidental effects on environmental quality. Investments involving public good aspects must enter into any effort to manage environmental quality in an optimal fashion.

To review our points briefly: (1) Technological external diseconomies are an inherent and normal part of production and consumption; (2) they become progressively more important as the population rises and the level of output increases; (3) they cannot be properly dealt

[7] To take one extreme example, a region which went in heavily for electric space heating and wet scrubbing of stack gases (from powerplants and industries), which ground up its garbage and delivered it to the sewers and then discharged the raw sewage to watercourses, would protect its air resources to an exceptional degree. But this would come at the sacrifice of placing a heavy residuals load upon water resources. On the other hand, a region which treated municipal and industrial waste water streams to high level but relied heavily on the incineration of sludges and solid wastes would protect its water and land resources but at the expense of discharging waste residuals predominantly to the air. Finally, a region which practiced high-level recovery and recycle of waste materials and fostered low residual production processes to a far-reaching extent in each of the economic sectors might discharge very little residual waste to any of the environmental media

with by considering environmental media such as air and water in isolation from each other; (4) uncoordinated and ad hoc restrictions are not sufficient for their optimum management; and (5) public investment affects the amounts and effects of residuals and must be planned in light of them.

In view of these factors it is important to develop not only improved measures of the external costs resulting from differing concentrations and duration of residuals in the environment, but more systematic methods for projecting emissions of external cost-producing residuals, technical and economic tradeoffs among them, and the effects of recycle on environmental quality.

To help further this end, we have developed in appendix I a formal materials balance model and related it to some conventional economic models of production and consumption. The main objective of this model is to make some progress toward defining a system in which flows of services and materials are simultaneously accounted for and related to policies for environmental management. In the following section we present a brief verbal description of the formal model.

III. General Economic Equilibrium and Materials Balance

The takeoff point for our model of materials balance in general economic equilibrium is a mathematical depiction of general interdependency among consumers and producers in the various sectors of the economy, first presented by Leon Walras many years ago.[8] This basic theoretical model has been extended to include intermediate exchanges among industries and has been made operational by many later scholars, most prominent of whom is Wassily Leontief. The model we use consists of a large system of simultaneous equations which relates the services of resources (including labor) to final demands by consumers and intermediate demands by industries. This is done by combining a system of equations which relate resources services to product output and a system which relates commodity production to final demand. Another set of equations relates the prices of intermediate goods (services) to prices of basic resources.

To connect this economic exchange system to the physical flow of materials, it is convenient to add two other sectors to the ones normally included in such models—an "environmental" sector and a "final consumption" sector. The environmental sector provides physical inputs to the production process, and the final consumption process emits physical outputs to the environmental sector. Since mass is conserved, the flows from and to the environment must be equal. Thus the system balances in physical terms.

Equations are written which represent materials flows *from* the environment to all other sectors and which balance materials flows into and out of the environment sector, and materials flows to and from the final demand sector. The final stage of the model building is to derive equations which connect the flow of waste materials back to the environment, to the final demands for goods and services in the economy.

The treatment could be simplified slightly if we assumed that there is no recycling *per se*. Thus, in the context of the model, we could sup-

[8] See Ayres and Kneese, *American Economic Review, op. cit.* for citations to the literature.

pose that all residuals return to the environmental sector, where some of them (e.g., wastepaper) become "raw materials." They would then be indistinguishable from new raw materials, however; and price differentials between the two would be washed out. In principle, this is an important distinction to retain and the model does retain it.

IV. Finding Environmental Quality Constraints

In the previous section we have described verbally at least the bare outline of a procedure for accounting for materials flow in conjunction with a general economic interdependency model. The physical flow of materials between various intermediate (production) sectors and the final (consumption) sector tends to be accompanied by, and correlated with, a (reverse) flow of dollars. But, the physical flow of materials from and back to the environment is only partly reflected by actual dollar flows; namely, land rents and payments for raw materials. There are three classes of physical exchange for which current private property arrangements provide no basis for counterpart economic transactions; these are—

(1) Private use for production inputs of "common property" resources, notably air, streams, lakes, and the ocean (e.g., water from streams in industrial production; oxygen from the atmosphere in combustion, etc.);

(2) Private use of the assimilative capacity of the environment to "dispose of" or dilute wastes and residuals; and

(3) Inadvertent or unwanted material inputs to productive processes—diluents and pollutants.

All these services or disservices are physically transferred at zero price, because there exist no social institutions which would permit the resources in question to be "owned," and exchanged in the market.

Of the three categories mentioned, (2) and (3) seem to be the most important at the moment.[9] One method of dealing with these externalities would be for a public agency to devise a coherent set of environmental standards and implement them by direct control of discharges to the common-property environmental media.* Alternatively the agency could devise a set of shadow prices to provide appropriate incentives for attaining the environmental standards. Which would be the more efficient system is a deep and interesting question which we will not address here.[10] For present purposes we will assume that, if a set of environmental standards can be devised, it will be efficiently implemented.

The usual approach of public officials has been to attempt to achieve the highest standard possible in a particular environmental medium without regard to any sort of economic considerations. Economists, on the other hand, have tended to favor cost-benefit analysis of particular standards in a partial equilibrium context. There seem to be important reasons, growing out of the above analysis, why neither of these approaches will produce satisfactory results even if they are successful in achieving their specific aims.

[9] But see the paper by Cole, *op. cit.*

[10] For discussion of this issue in connection with water quality, see Allen V. Kneese and Blair T. Bower, *Managing Water Quality, Economics, Technology Institutions,* the John Hopkins Press, 1968, pt. II.

*Further discussion of this issue is found in the paper by Davis & Kamien in this volume.

(1) The "highest" standard in one medium may well imply the "lowest" in another. A coherent set of standards can only result from an analysis which gives due regard to the conservation of mass.

(2) Even if "benefits" can be imputed in a partial equilibrium context, say by the use of property values,[11] the solution which maximizes net benefits in a partial equilibrium sense appears unlikely to have much normative significance in the presence of pervasive and interdependent externalities resulting from residuals discharge.

Thus partial and *ad hoc* approaches appear as likely to lead away from optimality as toward it. An effort must be made to find coherent and consistent environmental controls in the context of which decentralized voluntary exchange may perhaps again assume its role with respect to resources allocation. In the following section, we examine some theoretical aspects of this question.

V. Environmental Standards and Decentralized Decisionmaking

A. Some General Notes on Second-Best Problems*

In previous sections, a model was outlined which demonstrates the "pervasiveness" of externalities associated with interrelationships between production, consumption, and environmental sectors. It is important to note that even if a particular type of productive activity does not directly utilize inputs from the environmental sector, it may do so indirectly through its demand for intermediate products from sectors that do. Thus, while selling insurance may not produce waste residuals directly, the use by this activity of paper means that insurance adds indirectly to demand for environmental waste disposal services. In fact, it is difficult to imagine any economic activity which does not directly and (or) indirectly contribute to demands on the unpriced environmental sector. And if this is true, a nearly universal divergence (of greater or lesser degree) between prices and social costs is implied.[12]

The question addressed in this part of our paper is whether decentralized decisionmaking coupled with environmental planning on the part of a governmental unit can reestablish or approach an optimum in social product.[13] This is not to say that we imagine such an optimum as a pragmatic possibility. At most, the analysis could yield reference points against which to compare alternative planning approaches. Specifically, we wish to know how desirable it might be to rely on decentralized decisions regarding the environmental sector where these de-

[11] See Ronald G. Ridker, *Economic Costs of Air Pollution,* Frederick A. Prager, New York. 1967. chs. 6 and 7.

[12] This can be seen most readily if we observe that equations (5a) or (5b) in appendix I relate intermediate product prices (and final demand prices) to imputed prices of raw materials and services. Given the absence of a viable market for environmental services, the right side of (5a) or (5b) will not contain prices for these services, thus leading to an inequality between prices and social costs; that is, provided environmental services have a positive social cost.

[13] It must be emphasized that the Pareto conditions derived from decentralized decisions without consideration of environmental pollution are *not* optimal in the sense that *some* adjustment in allocation of resource inputs, and in production levels of final products, could make everyone better off without simultaneously making anyone worse off.

* Further discussion of this issue is found in the papers by Arrow and Steiner in this volume.

cisions are tempered (or constrained) by collectively imposed controls.[14]

Before embarking on this discussion, we will undertake a brief review of several issues in the context of what economists call "second best" problems.[15] Second best versus first best is defined and related to the concept of Pareto optimality quite well by Davis and Whinston: "A Pareto optimum problem is one which, given the market clearing conditions and the *technology* of the economy, all normative behavioral rules can be determined so that solutions of the system achieve a * * *

[14] By standards, in this context, we mean broadly any type of controlling mechanism on the utilization of environmental services; e.g., taxes, subsidies, or direct control on emissions or quality deterioration. Generally, a particular level of control can be implemented in a variety of ways, provided the behavior of participants such as firms, toward each type of controlling mechanism is predictable.

[15] The "second best" situation is usually cast in the following mathematical terms. Given some function $H(Y_1 \ . \ . \ . \ Y_n)$ that is differentiable and where partial derivatives are defined as H_i, second order partial derivative as H_{ij} or H_{ii}, and where H is subject to a constrain and J_{ii} as previously defined, the maximum of H given J using a LaGrange multiplier method is:

$$\max \psi = H - \lambda J$$

which yields the first order Paretian conditions:

$$H_i - \lambda J_i = 0 \qquad (i = 1, 2, \ . \ . \ . \ n).$$

If the additional restriction is added,

$$H_1/H_n = \rho(J_1/J_n)$$

where $\rho \neq 1$ then the first order conditions become

$$H_i - \lambda J_i - \delta[(H_n H_{ji} - H_j H_{ni}/H_n^2) - \rho(J_n J_{1i} - J_j J_{ni}/J_n^2)] = 0 \qquad (i = 1, 2, \ . \ . \ . \ n)$$

where δ is a second LaGrange multiplier. Note, if there are m such additional restrictions, the first order conditions become

$$H_i - \lambda J_i - \sum_{j=1}^{m} \delta_m[(H_n H_{ji} - H_j H_{ni}/H_n^2 - \rho_m(J_n J_{ji} - J_j J_{ni}/J_n^2)] = 0$$

$$(i = 1, 2, \ . \ . \ . \ n.$$

If the H and J relationships are separable in that

$$H = H_1(Y_1) + H_2(Y_2) + \ . \ . \ . \ H_n(Y_n)$$

then in the case of only one additional restriction, the first order conditions become:

$$H_i - \lambda J_i = 0, \qquad (i = 2 \ . \ . \ . \ n-1)$$

and

$$H_i - \lambda J_i - \delta[(H_n H_{1i} - H_1 H_{ni}/H_n^2) - \rho(J_n J_{1i} - J_1 J_{ni}/J_n^2)] = 0 \qquad (i = 1 \text{ and } n),$$

since

$$H_{ij} = 0, \text{ and } J_{ij} = 0, \text{ for } i \neq j.$$

However, the materials balance model presented in section III and appendix I illustrates that separability cannot be presumed to exist between *any* productive sector and the environmental sector such that first order Paretian conditions could be optimal with no adjustment or governmental intervention.

maximum, [while] a second best optimum problem is one in which, given the market clearing conditions and the technology of the economy, at least one [but possibly more] of the behavioral rules is nontrivially specified, *cannot be changed*, and neither can the behavior of the deviant(s) be altered by *any* policy, and the remainder of the normative behavioral rules are to be chosen so as to achieve a * * * maximum." [16] Quite clearly, whether our problem falls within the realm of the "second best" by Davis and Whinston definition depends upon the extent to which we must accept as immutable the noninclusion of imputed costs of environmental services in decentralized decisions. Conceptually, it may be possible to impute costs of utilizing the environment such that a return to full Paretian conditions is possible (see app. II). If so, the imposition of zero costs for environmental services as a "first best" or "second best" type of problem will affect the kind of behavioral rules that are superimposed to reduce or counteract these social-private cost discrepancies. There appear to be three methodological options:

(1) To presume that complete correction for all deviations is plausible such that the "first best" Paretian conditions can be attained through appropriate dosages of environmental standards, taxes, subsidies, or other policy instruments;[17]

(2) [or that] Pervasity of externalities is so encompassing, and (or) detailed information on the general equilibrium system so costly, that deviations between private and social costs from this source must be viewed as totally immutable;

(3) [or that] The deviations between social and private costs are only partially correctable such that a "second best" in the Davis-Whinston sense must be imposed following (or in conjunction with) the *partial removal* of deviations between social and private costs for environmental services.

In a welfare sense (disregarding information costs), the three options can be ranked. Case 1, of complete correction, allows the system to achieve the "first best" optimum. Case 3, while not achieving the "first best" optimum, is less constrained than case 2, such that the "second best" optimum achieved through appropriate governmental applicaton of environmental standards would be at least as good in a welfare sense as could be derived under case 2.[18]

---

[16] See O. Davis and A. Whinston, "Welfare Economics and the Theory of Second Best," *Rev. Econ. Stud.*, vol. XXXII, January 1965. A divergence appears to have developed in the meaning of "second best" in the literature. While Davis and Whinston presume the behavior of the deviant(s) from Paretian conditions is immutable, and given this, ask how might other participants (e.g., firms, individuals) alter their behavior to achieve a "second best"; McManus and Bohm, among others, view the "second best" problem as trying to correct for the deviant's behavior by "doing something about the deviant," be it a monopolist or groupings of technological and institutional externalities, as is analyzed here. Finally, Morrison has broadened the definition of "second best" theory to include replacement of old constraints, as well as adding new constraints, to existing ones. M. McManus, "Private and Social Costs in the Theory of Second Best," *Rev. of Econ. Stud.*, vol. XXXIV, July 1967; C. C. Morrison, "Generalizations on the Methodology of Second Best," *West. Econ. Jour.*, vol. 6, No. 2, March 1968.

[17] Davis and Whinston have suggested the informational content for such global corrections is "difficult to obtain" and in any case "in any real economy there will always be imperfections somewhere." See O. Davis and A. Whinston, "Piecemeal Policy in the Theory of Second Best," *Rev. Econ. Stud.*, vol. XXXIV, July 1967.

[18] By "less constrained" we refer to the constraints imposed by assuming social and private cost gaps cannot be altered and thus are more binding than if these constraints could be relaxed, though not completely. Since the value of a constrained maximum cannot be less if one or more of the constraints is relaxed (though it might be constant), our conjecture on rankings must follow. Thus, whether one views pervasive externalities as natural, or only transient and alterable facets of the economic system, influences the type of, and level of achievement of, alternative governmental actions. However, of more importance to the discussion here is the realization that case 1 may offer less social

B. SYNOPSIS OF FINDINGS ON ENVIRONMENTAL STANDARDS

In appendix II, a simplified model of an economy is developed in order to study public policies directed toward ameliorating the social inefficiencies created by nonpricing of environmental services. As we note again in the final section of this paper, the model rests on drastically simplifying assumptions, and the conclusions we are able to draw from it can be considered no more than suggestive. The model takes as its starting point the materials balance model briefly discussed in section III and developed explicity in appendix I. Inherent in the model developed here is the assumption of competition with consequent fulfillment of Paretian optimal conditions (prices equal marginal social costs) everywhere in the economy except for environmental services. The model also contains the assumption that all products are produced under conditions of constant costs; that is, total production costs divided by physical production is constant regardless of the level of production, and under conditions of no substitution between factors of production.[19] Given these assumptions, we are able to derive explicit environmental standards via taxes on final products such that the economy in principle would operate efficiently so that the Paretian optimal conditions are met by every sector. This conclusion leads us to believe that if a government agency could procure the necessary information on materials balances and economic interdependencies for the entire economy, a coherent and consistent set of environmental standards could be determined.

From our findings, it is our view that once such a set of optimal environmental standards were imposed on firms or industries, these standards would become a part of the natural impediments to business enterprise comparable to other resource scarcities. Voluntary exchanges in markets and individual decisionmaking for the most part could be preserved. Of course, these conclusions are strongly predicated on the underlying assumptions of the model we *selected* to study, including the assumption of no noncompetitive firms or industries and on the assumption that environmental standards could be *efficiently* implemented.

If waste flows influence the level of well-being of consumers, then the Government agency, in addition to knowledge of technical interdependencies and materials balances, must have knowledge of preferences of consumers in order to establish optimal environmental standards. The extent of public reaction to aesthetic degradations, smog, pesticides, fish kills, et cetera, in the recent past is indicative that consumer preferences must be taken into account in the formulation of controls on the use of environmental services.

---

welfare if no corrections are made than either options 2 or 3. No correction for excess use of waste assimilation or removal would be expected to result in lower levels of social welfare than if a small degree of correction is imposed. This is not to argue that any amount of correction will bring about an increase in welfare, only that *certain* levels of correction brought about by imposition of environmental standards will. For a heuristic argument of the point regarding "small degrees of correction," see E. J. Mishan, *The Costs of Economic Growth*, London: Staples Press, 1967.

[19] The assumption of nonsubstitutability between factors including environmental services is a major impediment to the interpretation of our conclusions as regards realistic environmental policies. There is substantial evidence that substitution possibilities are widespread. See for example, G. O. G. Löf and A. V. Kneese, *The Economics of Water Utilization in the Beet Sugar Industry*, Resources for the Future, Inc., Baltimore: Johns Hopkins Press 1967, p. 79.

We find that a set of coherent and consistent standards are identifiable in the case where industry behavior resulting from the assumption that environmental services are virtually free, is *immutable*. This nonoptimal but unchangeable behavior is partially (or, in certain cases, completely) compensated for through government regulation of consumer purchases through taxes. The taxes charged in this case were found to be functionally related to the amount of environmental services embodied in each final product. Once such taxes are imposed, markets can be allowed to function without controls, with individual choice and decisionmaking preserved. However, in this case of total immutability of industry's noncosting of environmental services, the Government agency would need information on market demand relationships and individual preferences in order to establish an optimal set of taxes in addition to the informational requirements regarding materials balances.

The final case studied was where environmental services can *partially* be charged to industries, but not at levels necessary to achieve Paretian conditions throughout the economy. In this case, the Government agency would not only need to regulate waste flows by industries through environmental standards, but also regulate consumer purchases through taxes and (or) subsidies.

It is important to note that, once such environmental standards were implemented, in all three cases studied, markets would remain free in the sense that buyers and sellers could establish prices. Also, once taxes were imposed on environmental services utilized by industries, these industries could still operate as private decision bodies not manipulated by a Government agency. In addition, once taxes or subsidies were imposed on consumers as regards purchases of goods, individual choice would still prevail as to their combinations of purchases.

Undoubtedly the greatest impediment to developing optimal or suboptimal (as in cases 2 and 3) control measures is the almost limitless amount of information required, which may well be the *most* binding constraint in that the complexity of pricing or tax rules, the cost of information, and the propensity to induce greater costs from information errors, may overwhelm consideration of efficiencies analyzed here. Also, the losses in welfare encountered by omitting many of the tax measures may be slight compared to informational costs of obtaining these measures, yet one cannot establish whether this is in fact true without substantial expenditure on information.

C. PRICE AND WELFARE IMPACTS OF ENVIRONMENTAL STANDARDS

Our next consideration is the impact of environmental standards imposition on prices of intermediate and final products and prices of resource inputs other than environmental services; i.e., land and labor. Conceptually, the imposition of such standards will have two distinguishable effects on prices: a general rise (reduction) in product (resource input) prices, and a change in relative prices of both products and resource inputs. The increase in particular final product prices will be related to the total amount and type of environmental

services required per unit of product, including not only the direct utilization of environmental services in the product's manufacture, but also indirect utilization through the use of other products which require environmental services. Thus, the production of certain chemicals requires the input of other previously processed products, and the total input of environmental services to chemicals not only includes services directly used by chemicals, but, in addition, environmental services required by the previously processed products.

Given explicit knowledge of the relationship between waste flows to the environmental sector and particular levels of final demands (equation (13), app. I) by displacing final demand by one unit (or dollar) for one final product, holding all other final demands constant, and repeating this procedure for each final product, total environmental service requirements per unit of each type of final product is calculable. Those final (or) intermediate products with relatively high environmental service requirements, if standards are imposed, should exhibit higher prices relative to those final products with low or nearly zero waste flows per unit of final demand. This discussion, however, implicitly assumes completely inelastic changes in technology (as regards reduction in residuals flow) in response to changes in costs of residuals flow.

Environmental standards that are in any sense restrictive will raise costs of environmental services relative to costs of other resource inputs. The change in relative resource prices may induce substitution of these resource inputs for environmental services which would tend to alter costs of inputs in relation to each other. The particular technologies imposed for illustrative purposes in section III and appendix I do not allow for such substitutions and thus imply, given a particular set of other resource prices, the total impact of environmental standards will be transmitted to product prices.

In terms of existing economic models, not much can be said regarding changes in welfare when more than one deviation (e.g., externality plus monopoly) pervades the economy. Even in the case of a single deviancy, questions of income distribution have not been satisfactorily included.[20]

Foster and Sonnenschein have recently been able to prove that a "radial decrease in distortion will unambiguously improve welfare," provided there is only one consumer, no products are inferior, and all products are produced at constant costs per unit.[21] By "radial decrease in distortion" Foster and Sonnenschein mean a uniform reduction in the absolute or percentage differences between a consumer's marginal rates of substitution and producers' marginal rates of transformation. Since our model contains a constant cost assumption, and if all consumers are similar in their preferences for products and environmental services, we can infer from the above results that a radial decrease in

[20] See E. J. Mishan, "Second Thoughts on Second Best," *Oxford Economic Papers,* N. S. 14, October 1962.

[21] See E. Foster and H. Sonnenschein, "Price Distortion and Economic Welfare," forthcoming *Econometrica.*

distortion between prices and marginal costs will generally increase welfare.[22]

The above result regarding improvements in welfare, as Foster and Sonnenschein prove, is dependent to a large extent on the assumption of constant costs. If costs are not constant, or if preferences (and incomes) of individuals are dissimilar, the above conclusion regarding the change in welfare from imposing environmental standards may not be correct. However, from our heuristic argument regarding the probable change in welfare brought about by imposition of a particular set of environmental standards, we believe once a materials balance is developed for a region, that a quasi-coherent (not optimal—but probable welfare improving) and consistent set of environmental standards could be developed.

VI. Comments on Policy and Research

A. An Approach to Regional Environmental Management

We propose that something like the approach outlined in the following paragraphs is more likely to lead to desirable results than the present partial and ad hoc efforts, particularly in light of the findings in section V. We visualize that it would be applied to a "region" where heavy demands are made on the environment to receive waste residuals.[23] As a first step, a more or less refined materials balance would be estimated for the area. This would involve some major complications since regional economies are very "open." In other words, all material (including goods at various stages through the manufacturing process) imports and exports would have to be accounted for. Getting this information, as well as the remainder of that needed to estimate materials balances for various industries, would involve considerable primary data collection. But once completed, the materials balance would provide a picture of the residuals flows in the area. Economic base-input-output models could then be used to project levels of activity and

[22] The degree of distortion associated with nonpricing of environmental services is equal, utilizing the equations set forth in appendices I and II, to:

$$\mu_\theta \sum_{j=1}^{N} (C_{if} - C_{fi}A_{ik}) \qquad k=1, 2 \ldots N$$

thus,

$$\psi\left[\mu_\theta \sum_{j=1}^{N} (C_{if} - C_{fi}A_{ik})\right] \qquad k=1, 2 \ldots N,$$

where

$0<\psi<1$, imposed as a per unit tax on products 1 through k would generally lead to an improvement in welfare. Note that the tax on each product is proportional to each product's demands on environmental services.

[23] A few examples of the sort of regions we have in mind are the east coast megalopolis, the San Francisco Bay area, Greater London and the Thames Estuary area, the Ruhr, the Saar, Upper Silesia, the Sao Paulo region, and so forth. Highly concentrated connurbations of this type do not constitute a large part of the earth's land area, but do contain a disproportionate part of its population, economic activities, and externalities.

industry mixes into the future, as well as transportation requirements, population, and other parameters of interest in constructing a new materials balance. In the first instance, present technology and low levels of materials recycle and byproduct recovery could be assumed. This procedure would differ from conventional approaches in that it would relate levels of residuals generation logically to the industrial and population base and account for all residuals in an internally consistent model. Conventionally, emissions of air borne, liquid borne, and solid wastes are extrapolated separately.

As a further step, this procedure would permit analysis of the overall impact on all residuals of control measures instituted for one or more of them. As we have discussed in section I, for a given level of economic activity, a given efficiency of energy conversion, and a given degree of recycle and byproduct recovery, reduction of one type of residual must come at the expense of creating another. Furthermore, residuals should be classified and quantified in terms of their potential recoverability as a basis for economic analysis of this alternative.[24] Also, and quite important, the overall residuals implications of projected changes in waste treatment and recovery technologies could be tested.

Another major effort would have to go into mathematically simulating the relevant natural system, especially the meteorological and hydrological subsystems, in such a way that residuals concentrations and their probability and duration could be estimated, as well as secondary effects of residuals discharges. The latter might include photochemical reactions in the atmosphere and the reduction of dissolved oxygen in water courses resulting from the discharge of organic wastes to them.

Tradeoffs between discharges to different environmental media, the costs of controlling discharges, the costs of recycle over and above those induced by internal profit considerations, as well as pertinent information of various types on the probable damaging effects of alternative concentrations and durations of residuals in the different environmental media would have to be considered. On this basis a reasonably consistent and coherent set of standards might be established.[25]

Once established, the environmental standards would become part of overall framework within which voluntary market exchanges take place. We have previously argued that, if these environmental standards (although themselves perhaps not fully optimal) are considered as fixed constraints on the system, the conditions of Pareto optimality through partially adjusted choice can be met within them and be deemed to have the same normative significance as when achieved under the "natural" constraints of conventional resource scarcities, private ownership, and technological production functions. This result is very significant because it, in large measure, preserves the role of

[24] A distinction should be drawn between recycle and byproduct recovery processes which market conditions will make internally profitable to the firm and those which might be stimulated by public policy—residual taxes, for example.

[25] Among the many other difficult questions we do not address in this paper is by what authority these standards would be set. What we have in mind is a decision by some representative political body charged with protection and enhancement of the environment. Bodies of the type we visualize already exist on a much more limited scale. The Delaware River Basin Commission is an example. An augmented Thames Conservancy in England and Ruhrverband in Germany are further examples of the type of regional environmental agencies which might suit the purpose.

decentralized decisionmaking with its inherent efficiencies.[26] Once the economic system has adjusted to these "approximate" standards, the resulting relative prices might provide a firmer basis for "partial" damage estimation and refinement of controls.

B. IMMEDIATE NEXT STEPS

Our analysis suggests to us a need to move forward rapidly toward a fuller understanding of the economics, politics, and technology of environmental pollution management. The best way to do this, in our opinion, would be to mount a well-financed multidisciplinary research project to analyze materials flow as related to the present, and projected future, economic structure of a particular region. As explained above, this analysis should be combined with a study of the concentration and duration of the residuals generated in the various environmental media. The phenomena of degradation and dilution in the environment are influenced by stochastic processes in nature. Atmospheric wind patterns, temperatures, and temperature inversions are examples, as are variations in streamflow. Moreover, there are complex interrelationships between the biological phenomena, important in the degradation of residuals, and these physical phenomena. For example, the basic equations of the "oxygen sag" in residuals-receiving water include temperature as a major variable. Accordingly, close integration of studies in the various natural sciences as well as with the economic studies would be needed.

(1) *Obstacles*

There are many institutional and professional obstacles to taking the sort of approach we have outlined and tried to defend—even for a study. Air, water, and solid wastes pollution are the provinces of different agencies at the Federal and usually at the State and local government levels. Moreover, only limited efforts have been made to wed economic analysis to the available meteorological and hydrological models. Even within the natural sciences, institutional and professional barriers to a fully interrelated approach exist and are important obstacles. There is a relatively new agency, the Environmental Science Services Administration (ESSA), whose name suggests the appropriate breadth. But this implication is deceptive. ESSA deals solely with geophysical phenomena; whereas, as we have argued above, one of the most important natural science interfaces is between physical and biological phenomena.[27]

---

[26] Another area of great importance which we pass over in this paper is the role of government in implementing large-scale environmental control systems in the region. Research on water quality management has show clearly that such measures as reservoir regulation of river flows, mechanical aeration of residuals-receiving water courses, and treatment of municipal and industrial wastes together in collective regional treatment plants can enter efficiently into regional water quality management systems. See Robert K. Davis, *The Range of Choice in Water Management*, the Johns Hopkins Press, 1968. Similar points could be made with respect to other environmental media. In the context of our present model, collective management systems would have to be analyzed primarily in cost minimization terms.

[27] Other examples could be given. Some experts believe that one of the most important sinks for atmospheric CO_2 is in the vast quantities of phytoplankton found in the bays, estuaries, and shallow coastal waters. These waters are very much subject to pollution by toxic industrial waters which kill these organisms.

(2) *Institutional changes*

We are not in a favorable institutional posture with respect to the type of approach we have discussed in this paper—neither in regard to analysis nor implementation. We suggest that reorganization and consolidation at the Federal level should seriously be considered. This would involve the creation of a new environmental agency to take over control functions now scattered in the Department of Health, Education, and Welfare, the Department of the Interior, and other agencies. With respect to scientific and research services, the responsibilities of ESSA could be broadened to include biological and social aspects needed to provide comprehensive models and analysis of environmental systems. For implementation of programs, regional environmental management agencies of some type will, we believe, be needed.

The regional case study we have proposed earlier should be suggestive of the type of the regional management agency needed. In fact, study of institutional and organizational questions should be built into the project. Obviously, present governmental arrangements are not conducive to the execution of the regional study. A special group would have to be authorized and funded by Congress or by a private foundation.[28]

(3) *The challenge to economic theory*

Finally, we see an urgent need to develop more relevant and operational economic models for dealing with pervasive externality phenomena. A few economists have observed that external diseconomies increase rapidly (nonlinearly) and pervasively with economic and population growth, but comparatively little has been done to formulate analytical or normative models based on this insight.[29] Assuming independence of production and utility functions in our basic economic models tends to become less defensible. The simplifying assumptions necessary in the model development of sections III and V, of separability in environmental services and utility functions, of nonjoint production, and simplified aggregation over interfirm complementary-competitive relationships are examples of the lack of realism in most current economic analyses. In our opinion, economic theorists face no more urgent task than to devise improved models for the analysis of environmental pollution, urban congestion, landscape deterioration, and the host of other externality phenomena which accompany economic growth.

---

[28] The Delaware Estuary area commends itself as a study region. It contains one of the great industrial areas of the world and much relevant work has already been done there by the Regional Science Center at the University of Pennsylvania and the Federal Water Pollution Control Administration, among others.

[29] See for example, Tibor Scitovsky, "External Diseconomies in the Modern Economy," *The West. Econ. Jour.*, vol. IV, No. 3, summer 1966.

APPENDIX I. *The General Equilibrium—Materials Balance Model*

The following analysis is a formal statement of the general equilibrium—materials balance model described verbally in section III. Its elaboration involves the following quantities:

resources and services: $\overbrace{r_1, r_2, \ldots, r_M}$

products or commodities: $\overbrace{X_1, X_2, \ldots, X_N}$

final demands: $\overbrace{Y_1, Y_2, \ldots, Y_N}$

product or commodity prices: $\overbrace{P_1, P_2, \ldots, P_N}$

resource prices: $\overbrace{V_1, V_2, \ldots, V_M}$

The M basic resources are allocated among the N sectors as follows:

$$r_1 = a_{11}X_1 + a_{12}X_2 + \ldots a_{1N}X_N$$

$$r_2 = a_{21}X_1 + a_{22}X_2 + \ldots a_{2N}X_N$$

$$\vdots$$

$$r_M = a_{M1}X_1 \quad a_{M2}X_2 + \ldots a_{mn}X_N$$

or

$$r_j = \sum_{k=1}^{N} a_{jk}X_k \qquad j = 1, 2, \ldots, M \tag{1a}$$

In (1a) we have implicitly assumed that there is no possibility of factor or process substitution and no joint production. In matrix notation we can write:

$$\begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_M \end{bmatrix} = \begin{bmatrix} a_{11}, & \ldots & a_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{M1}, & \ldots & a_{MN} \end{bmatrix} \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_N \end{bmatrix} \tag{1b}$$

where $[a]$ is an M x N matrix.

A similar set of equations describes the relations between commodity production and final demand:

$$(2a) \qquad X_j=\sum_{k=1}^{N} A_{jk}Y_k \qquad j=1, 2, \ldots, N$$

$$(2b) \qquad \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_N \end{bmatrix} = \begin{bmatrix} A_{11}, & \ldots & A_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{NL}, & \ldots & A_{NN} \end{bmatrix} \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix}$$

and the matrix $[A]$ is given by

$$(3) \qquad [a]=[I-C]^{-1}$$

where $[I]$ is the unit diagonal matrix and the elements C_{ij} of the matrix $[C]$ are essentially the well-known Leontief input coefficients. In principle, these are functions of the existing technology and, therefore, are fixed for any given situation.

By combining (1) and (2), we obtain a set of equations relating resource inputs directly to final demand, viz,

$$(4a) \qquad r_j=\sum_{k=1}^{N} a_{jk}\sum_{l=1}^{N} A_{kl}Y_l=\sum_{l=1}^{N} a_{jk}A_{kl}Y_l=\sum_{l=1}^{N} b_{jl}Y_l \qquad j=1, 2, \ldots, M$$

or, of course,

(4b)

$$\begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_M \end{bmatrix} = \begin{bmatrix} a_{11}, & \ldots & a_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{M1}, & \ldots & a_{MN} \end{bmatrix} \begin{bmatrix} A_{11}, & \ldots & A_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{N1}, & \ldots & A_{NN} \end{bmatrix} \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} b_{11}, & \ldots & b_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ b_{M1}, & \ldots & b_{MN} \end{bmatrix} \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix}$$

We can also impute the prices of N intermediate goods and commodities to the prices of the M basic resources, as follows:

$$(5a) \qquad P_k=\sum_{j=1}^{M} b_{jk}V_j \qquad k=1, 2, \ldots N$$

or

$$(5b) \qquad [P_1, P_2, \ldots P_N]=[V_1, V_2, \ldots V_M] \begin{bmatrix} b_{11}, & \ldots & b_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ b_{M1}, & \ldots & b_{MN} \end{bmatrix}$$

In order to interpret the X's as physical production, it is necessary for the sake of consistency to arrange that outputs and inputs always balance, which implies that the C_{ij} must comprise all materials exchanges, including residuals. To complete the system so that there is no net gain or loss of physical substances, it is also convenient to introduce two additional sectors; viz., an "environmental" sector whose (physical) output is X_0 and a "final consumption" sector whose output is denoted X_f. The system is then easily balanced by explicitly including flows both to and from these sectors.

To implement this further modification of the Walras-Cassel model, it is convenient to subdivide and relabel the resource category into tangible raw materials $[r^m]$ and services $[r^s]$:

$$\begin{array}{cc} r_1, r_2, \ldots -r_L & r_{L+1} \ldots -r_M \\ \vdots \quad\quad \vdots & \vdots \quad\quad \vdots \\ \underbrace{r_1^m, r_2^m, \ldots -r_L^m} & \underbrace{r_1^s \ldots r_P^s} \\ \text{raw material (units)} & \text{service (units)} \end{array}$$

where, of course,

$$L+P=M \tag{6}$$

It is understood that services, while not counted in tons, can be measured in meaningful units, such as man-days, with well defined prices.

The coefficients $[a_{ij}]$, $[b_{ij}]$ are similarly partitioned into two groups; for example,

$$\begin{array}{cc} b_{1j}, \quad b_{2j}, \quad \ldots b_L & b_{L+1,j}, \ldots b_{Mj} \\ \vdots \quad\quad \vdots & \vdots \quad\quad \vdots \\ b_{1j}^m, b_{2j}^m, \ldots b_{Lj}^m & b_{1j}^s, \quad \ldots b_{Pj}^s \end{array}$$

These notational changes have no effect whatever on the substance of the model, although the equations become somewhat more cumbersome. The partitioned matrix notation simplifies the restatement of the basic equations. Thus (1b) becomes:

$$M\left\{\begin{bmatrix}\vdots\\ r\\ \vdots\end{bmatrix}=\begin{bmatrix}r^m\\ \ldots\\ r^s\end{bmatrix}\begin{matrix}L\}\\ P\}\end{matrix}=M\left\{\begin{matrix}\{L\\ \{P\end{matrix}=\begin{bmatrix}b_m\\ \ldots\end{bmatrix}\begin{bmatrix}\vdots\\ Y\\ \vdots\end{bmatrix}\right\}N \tag{7}$$

We now wish to focus attention explicitly on the flow of materials through the economy. By definition of the Leontief input coefficients (now related to materials flow), we have:

$C_{kj}X_j$ (physical) quantity transferred from k to j

$C_{jk}X_k$ quantity transferred from j to k

Hence, material flows from the environment to all other sectors are given by:

$$(8) \qquad \sum_{k=1}^{N} C_{\theta k}X_k = \sum_{j=1}^{L} r_j^m = \sum_{j=1}^{L}\sum_{k=1}^{N} a_{jk}^m X_k = \sum_{j=1}^{L}\sum_{k=1}^{N} b_{jk}^m Y_k$$

using equation (1), as modified. Obviously, comparing the first and third terms,

$$(9) \qquad \underbrace{C_{\theta k}}_{\substack{\text{total material}\\ \text{flow } (\theta \text{ to } k)}} = \underbrace{\sum_{j=1}^{L} a_{jk}^m}_{\substack{\text{all raw materials}\\ (\theta \text{ to } k)}}$$

Flows into and out of the environmental sector must be in balance:

$$(10) \qquad \underbrace{\sum_{k=1}^{N} C_{\theta k}X_k}_{\substack{\text{sum of all raw}\\ \text{material flows}}} = \underbrace{\sum_{k=1}^{N} C_{k\theta}X_\theta + C_{f\theta}X_\theta}_{\substack{\text{sum of all return}\\ \text{(waste) flows}}}$$

Material flows to and from the final sector must also balance:

$$(11) \qquad \underbrace{\sum_{k=1}^{N} C_{kf}X_f}_{\substack{\text{sum of all}\\ \text{final goods}}} = \underbrace{\sum_{k=1}^{N} C_{fk}X_k}_{\substack{\text{sum of all}\\ \text{materials}\\ \text{recycled}}} + \underbrace{C_{f\theta}X_\theta}_{\substack{\text{waste residuals}\\ \text{(plus accumulation)}}}$$

$$(12) \qquad X_f = \sum_{j=1}^{N} Y_j$$

Substituting (12) into the left side of (11) and (2a) into the right side of (11), we obtain an expression for the waste flow in terms of final demands:

$$(13) \qquad C_{f\theta}X_\theta = \sum_{j=1}^{N}\sum_{k=1}^{N} (C_{jf} - C_{fj}A_{jk})Y_k$$

APPENDIX II. *Establishing Environmental Standards*

The consideration of policies to establish an economic approach to environmental quality is complicated by both the complexity of technologic and economic relationships involved in a comprehensive materials balance approach, and the institutional-legal impediments to implementation of appropriate standards. In this appendix, we attempt to analyze three possible cases of public response to correct for external costs.

Part A. *Optimal Environmental Standards*

In order to explore some of the effects of adding constraints (or modifying existing ones) on the utilization of environmental services, we adopt the following simplified model of the economy. Given the following definitions:

ω_i = the utility function of the ith person, assumed to have arguments only of consumption directly by the person, exhibit global concavity, and where $i=1, 2, \ldots z$.

$\overline{Y}_i$ = the total quantity of goods consumed by the ith person; i.e., a specific bundle of commodities.

y_{ik} = the quantity of product (final demand) of product k consumed by the ith person, where $k=1, 2, \ldots N$ as defined in Appendix I.

Y_k = production of the kth product available for final demands where $k=1, 2, \ldots N$.

r_j = the total quantity of resource service j available for producing all final products.

$$\max [\omega_1(\overline{Y}), \omega_2(\overline{Y}_2), \ldots \omega_z(\overline{Y}_z)] \tag{14.1}$$

subject to:

$$\sum_{i=1}^{z} y_{ik} \leq Y_k \qquad k=1, 2, \ldots N \tag{14.2}$$

$$C_{j0}X_0{}^1 \geq \sum_{j=1}^{N}\sum_{k=1}^{N}(C_{jj}-C_{fj}A_{jk})Y_k \qquad j=M \tag{14.3}$$

$$r_j \geq \sum_{k=1}^{N} b_{jk}Y_k \qquad j=1, 2, \ldots M-1 \tag{14.4}$$

$$\overline{Y}_i,\ Y_k,\ y_{ik} \geq 0 \tag{14.5}$$

for all i and k.

These relationships simply state that we wish to maximize welfare of the economy as measured by the utility levels of persons 1 through z, subject to the conditions that all markets for final goods are cleared, and that no more resources, including environmental services, are used

[1] Note $C_{f0}X_0$ is not identical to $C_{f\theta}X_\theta$ in equation (13). $C_{f0}X_0$ is a measure of environmental assimilative capacity while $C_{f\theta}X_\theta$ is a direct measure of waste flows originating from productive activities. In the analysis that follows, we will use interchangeably $C_{f\theta}X_\theta$ and

$$\sum_{j=1}^{N}\sum_{k=1}^{N}(C_{jf}-C_{fj}A_{ik})Y_k$$

to denote the magnitude of waste flows.

than are available. Following Davis and Whinston,[2] we reinterpret the utility bundles of goods $\overline{Y}_i$ as a scalar function (positively weighted vector sum) by attaching constant weights to each person's utility function and rewrite (14.1) as:

$$\max W=\sum_{i=1}^{z} \alpha_i \omega_i(\overline{Y}_i) \tag{15}$$

For each constraint or grouping of constraints (14.2) through (14.5), there are non-negative LaGrangean multipliers that reflect the cost of these constraints. Let us denote these as μ_k for (14.2), μ_θ for (14.3), and μ_j for (14.4). These multipliers can be interpreted as prices in that they measure the change in welfare (measured in cardinal units of α_i) per unit change in the constraint.[3] In a competitive market situation, excluding considerations of adjustment processes (either dynamic or comparative static), the multipliers would equal prices of final goods (η_k), or costs of direct and indirect resource inputs, including environmental services.

The first order consumer optimum conditions for the above model of the economy are:

$$\alpha_i \frac{\partial \omega_i}{\partial y_{ik}} - \eta_k \begin{cases} =0 \\ \\ <0 \end{cases} \quad \text{provided } y_{ik} \begin{cases} >0 \\ =0 \end{cases} \quad \begin{matrix} k=1, 2, \ldots N \\ i=1, 2, \ldots z \end{matrix} \tag{16}$$

---

[2] O. Davis and A. Whinston, "Welfare Economics and the Theory of Second Best," *op. cit.*

[3] Let $W^*=\sum_{i=1}^{z} \alpha_i \omega_i^*(\overline{Y}_i^*)$ denote the optimal value of (15) given constraints (14.2) through (14.5), then for small changes in one of the market clearing constraints (14.2), there will be small shifts in the optimal value of all variables. Let us further assume the new equilibrium also satisfies the optimal conditions (16) through (20) given in the text. Also, let b_k denote relaxation of one of the market clearing constraints (14.2). Then:

$$\partial W^*/\partial b_k = \sum_{i=1}^{z} \alpha_i (\partial \omega_i^*/\partial \overline{Y}_i^*)(\partial \overline{Y}_i^*/\partial y_{ik})(\partial y_{ik}/\partial b_k).$$

From constraint (14.2) the following shift is obtained.

$$\sum_{i=1}^{z} (\partial \overline{Y}_i^*/\partial y_{ik})(\partial y_{ik}/\partial b_k)=1 \qquad \text{for } j=k \text{ and zero if } j \neq k.$$

Multiplying this equation by η_k^* and subtracting it from $\partial \omega^*/\partial b_k$ above, one obtains:

$$\partial W^*/\partial b_k=\eta_k^*+\sum_{i=1}^{z} [\alpha_i(\partial \omega_i^*/\partial \overline{Y}_i^*)(\partial \overline{Y}_i^*/\partial y_{ik})-\eta_k^* \partial \overline{Y}_i^*/\partial y_{ik}](\partial y_{ik}/\partial y_k)$$

But the square bracketed term equals zero by constraint (16), such that $\partial W^*/\partial b_k=\eta_k^*$. Note that η_k^* is dimensionally similar to α_i. If αi is dollars per "util," i.e., the reciprocal of the marginal utility of money income, then η_k is dollars per unit quantity of Y_k.

which in verbal terms indicates that the ith person's marginal utility with respect to the kth final product weighted by the interpersonal utility weight α_i, equals the price of product k if it is consumed, and must be less if the product is not consumed.

The market clearing equations require that all markets for final demand products be cleared in order for equilibrium to be maintained, or if supply of final products exceeds demand, prices become zero. Note that demand for final products cannot exceed supply because of the direction of initial inequalities in (14.2).

$$(17) \qquad \sum_{i=1}^{z} y_{ik} - Y_k \begin{cases} =0 \\ \text{provided } \eta_k \\ <0 \end{cases} \begin{cases} >0 \\ k=1, 2, \ldots N \\ =0 \end{cases}$$

The first order Pareto conditions for environmental services yields:

$$(18) \qquad \sum_{j=1}^{N} \sum_{k=1}^{N} (C_{jf} - C_{fj}A_{jk}) Y_k - C_{j0}X_0 \begin{cases} =0 \\ \text{provided } \mu_\theta \\ <0 \end{cases} \begin{cases} >0 \\ M=\theta \\ =0 \end{cases}$$

which states that if environmental services are being fully utilized at the level $C_{f0}X_0$, then the price for these services must be positive and equal to μ_θ. Likewise, the Pareto conditions for other resource services are:

$$(19) \qquad \sum_{j=1}^{N} b_{jk} Y_k - r_j \begin{cases} =0 \\ \text{provided } \mu_j \\ <0 \end{cases} \begin{cases} >0 \\ j=1, 2, \ldots M-1 \\ =0 \end{cases}$$

The environmental services constraint represented by equation (14.3) and Pareto optimum relation (18) could be partitioned into types of environmental services provided, of course, the components summed to $C_{f\theta}X_\theta$. The simplification is made here that environmental services are of sufficient degree of homogeneity for a meaningful single price μ_θ to be obtained.

For purposes of simplification, we will also assume the basic production unit in the region is a particular sector k that produces one product Y_k. We note in passing, however, that this simplifying assumption, along with the above assumption on homogeneity of environmental services, would collapse in attempting to derive specific optimal policy prescriptions. For the model to provide exacting optimal environmental standards, each set of constraints would require disaggregation for individual firms with differing requirements on environmental media, and to take into account locational advantages and disadvantages.

Finally, the Pareto conditions for utilization of resource inputs in the production processes require:

$$(20) \qquad \eta_k - \mu_\theta \sum_{j=1}^{N} (C_{jf} - C_{fj}A_{jk}) - \sum_{j=1}^{M-1} \mu_j \; b_{jk} \begin{cases} =0 \\ <0 \end{cases}$$

$$\text{provided } Y_k \begin{cases} >0 \\ =0 \end{cases} \quad k=1, 2, \ldots N$$

This last Pareto condition indicates the price of the kth final good be just equal to the total costs of producing it, where total cost includes direct resource input costs to the kth product plus indirect costs of resource inputs used to produce other products necessary for production of the kth product. Thus, μ_θ and the μ_j must be interpreted as *gross* resource input costs per unit of final demand.

Now if the competitive economy were operating without cognizance of constraint (14.3), the first order conditions in (20) would change to:

$$(21) \qquad \eta_k - \sum_{j=1}^{M-1} \mu_j \; b_{jk} \begin{cases} =0 \\ <0 \end{cases} \quad \text{provided } Y_k \begin{cases} >0 \\ =0 \end{cases} \quad k=1, 2, \ldots N$$

and constraint (20) would be excluded from consideration. So long as (18) was not binding, however; i.e., $\mu_\theta=0$, environmental services were of such magnitude as to be virtually free, no important difference would arise between the Pareto optimum outputs of Y_k specified by (20) or (21). That is, assuming for the moment that $C_{f\theta}\bar{X}_\theta$ does not enter into the utility functions of individuals 1 through z. However, once the assimilative capacity or other aspects of environmental media became relatively scarce such that $\mu_\theta>0$, then the specification of optimal levels of final demands would change. Also, a divergence would arise between the first order production conditions (20) and (21). This point can be easily seen by comparing solutions for η_1 from the first order conditions for products k and 1 in the two cases, assuming there is only one additional resource with $\mu_j>0$:

$$(22.1) \qquad \eta_k - \mu_j \, [b_{jk} - b_{j1}] = \eta_1^*$$

$$(22.2) \qquad \eta_k - \mu_j \, [b_{jk} - b_{j1}] - \mu_\theta[(C_{jf} - C_{fj}A_{jk}) - (C_{jf} - C_{fj}A_{j1})] = \eta_1^{**}$$

thus,

$$(22.3) \qquad \eta_1^* \neq \eta_1^{**}$$

provided: (*a*) environmental media are not limitless such that μ_θ is zero, or (*b*) each product does not require an identical quantity of total environmental services per unit of product.

If the governmental agency responsible for establishing standards has complete information regarding constraint (14.3) previously given, it would be able to impose environmental standards such that condition (20) could be met by each industry k. Thus, the Pareto optimum conditions adjusted for pervasive externalities could be fulfilled throughout the economy, both for consumers and producers.

Since in our exposition $(C_{jf}-C_{fj}A_{jk})$ is assumed to be a constant such that demands for environmental services are proportional to final product, a tax per unit of final demand for each product k would induce each industry to fulfill condition (20). That is, provided the unit tax equaled

$$\mu_\theta(C_{jf}-C_{fj}A_{jk})$$

In doing so, Pareto optimality, including consideration of scarce environmental media, is achieved.

Thus far, we have only analyzed the case where pervasive externalities, induced by non-consideration of environmental services, entered into the sphere of production activities. We now turn to the case where waste flows simultaneously affect levels of welfare of each individual. This can be accomplished by including $C_{f\theta}X_\theta$ in the utility function of each individual. Note that this implies environmental pollution or waste flows is a "public good" in the Samuelson-Musgrave sense.[4] A first order condition regarding the consumer sector is added, in addition to the conditions specified under (16).

$$\mu_\theta+\sum_{i=1}^{z}\alpha_i\frac{\partial\omega_i}{\partial C_{f\theta}X_\theta}=0 \tag{23}$$

which states the sum of the marginal disutilities of all individuals induced by an additional unit of waste flow to the environmental sector must equal the resouree price μ_θ for environmental services. Note that $\partial\omega_i/\partial C_{f\theta}X_\theta$ is usually expected to be negative, in that the greater the waste flow, the lower the utility level each individual is expected to have. The inclusion of (23) into the model greatly complicates the derivation of environmental standards, for now not only must the Government agency have complete knowledge of the production aspects and material balances of the economy, but also the utility functions for each individual. In any case, we have shown that even if externalities pervade the entire economy, if a Government agency had the necessary information, was unconstrained as regards action, and no additional constraints on behavioral rules in the Davis and Whinston sense arose, a full Paretian optimum is possible. The Government agency could, in principle, formulate a *coherent* and consistent set of environmental standards.

In the previous section we implied that if such a set of standards could be found, once implemented, they would be viewed as "internal" within the system, and individual decisionmaking could still be preserved. Here, we have demonstrated that such a set of standards con-

[4] P. A. Samuelson, "The Pure Theory of Public Expenditure," *Rev. of Econ. Stat.*, Vol. 36, November 1954; P. A. Samuelson, "Diagrammatic Exposition of a Theory of Public Expenditure," *Rev. of Econ. Stat.*, Vol. 37, November 1955; R. A. Musgrave, "The Voluntary Exchange Theory of Public Economy," *Quar. Jour. Econ.*, Vol. 53, February 1939.

ceptually *does exist* and could be implemented via taxes or subsidies, and still retain individual decisionmaking as regards markets for final products and utilization of resources by industry *other* than environmental services. Difficulties arise, however, in empirically determining the production relationships, materials balances, and utility functions needed to establish the optimal set of environmental standards. It is very likely that costs of obtaining the requisite information would far exceed the gains from achieving the Pareto conditions (18), (20), and (23), at least in the near future. Thus, we endeavor to analyze the two other cases posed in part A of this section: To presume total pervasity of externalities such that the deviations between private and social costs, and (or) benefits, are immutable, and the alternative that externalities are only partially immutable; i.e., a partial correction for waste flows is possible. We now turn to look at these less theoretically satisfying, but possibly more relevant and interesting, cases.

Part B. *Complete or Partial Immutability and Second Best Policy*

In this section we raise the question: given *complete* or *partial* immutability in discrepancies between private and social costs due to pervasive externalities from residuals discharge, can a "second best" solution be found. By *complete immutability* we are thinking of the case where it is impossible for the Government to exercise *direct* controls over the use of environmental services by private industry. For example, the Government could not impose unit taxes, or emission controls on waste flows of firms. By *partial* immutability, we are identifying those cases where the Government can impose taxes or controls on firms utilizing environmental services, but where these taxes are less than the optimal level devised earlier. We also wish to know if such a "second best" solution impedes the function of voluntary exchanges in markets through individual actions.

Morrison [5] has demonstrated that, without complete knowledge of utility functions, production capabilities (and materials balances in terms of our analysis) plus *complete* information on the form of private-social cost discrepancies, the "second best" solution is generally indeterminate. Thus, the same or a greater amount of information is required to establish coherent and consistent environmental standards in the "second best" case, as is required in the "optimal" case examined previously. This is an important finding for analyzing externalities, since it suggests that a completely coherent set of environmental standards may never be formulated. The costs associated with measuring all requisite data inputs is undoubtedly extremely high, and may exceed the gains from increased efficiency, at least for some level of accuracy.

Let us assume that the discrepancies between private and social costs in production resulting from residuals discharge are immutable, in the sense explained above. Then, given the assumption of competition and constant cost technologies, each industry will fulfill the first order condition (21) of equating prices to marginal costs of production other than environmental services. If these conditions are viewed as un-

[5] C. C. Morrison, "Generalizations on the Methodology of Second Best," *West. Econ. Jour.*, vol. 6, No. 2, 1968.

changeable, then they become additional constraints on the system (14.1) through (14.5). An alternative in the "first best" sense would be to impose direct supply control constraints by Government agencies on all waste residual producing industries. In doing so, however, we are *adding* an additional constraint to the system (14.2) through (14.5), (15), and (21). The additional constraint very likely will conflict with constraints (21), and thus make irrelevant the "second best" case studied here, and it would remove entirely the possibility of voluntary market exchanges following its imposition. However, supply controls could be implemented such that welfare would reach the "first best" optimum, provided the assumption of constant costs was valid. In addition, we are interested only in solutions where in equilibrium $\eta_k = P_k$, $k=1, 2, \ldots\ldots N$, and we assume there is an inverse demand relationship relating P_k to demand for all products: [6]

$$(24) \quad P_k = f_k\left(\sum_{i=1}^{z} y_{i1}, \sum_{i=1}^{z} y_{i2}, \ldots \sum_{i=1}^{z} y_{iN}\right) \qquad k=1, 2, \ldots N$$

Since the Government agency cannot adjust the first order conditions for any industry, it cannot influence directly the activities of industries in our example. In the more general case of factor substitution not explicitly analyzed here, each firm or industry would be utilizing nonoptimal combinations of productive factors, including environmental services. With Leontief type technologies containing no factor substitutions, relative prices are distorted directly from not including costs of environmental services. The Government agency would have the option to develop public companies competitive with private industries, though we will not pursue this case, as it has been studied elsewhere.[7] Of course, since many industries contribute importantly to waste flows, to force these industries to adjust for their waste residuals implies nationalization of these industries, otherwise the competitive public companies would require continuous subsidization. The agency's other option would be to adjust consumer demand to compensate for the nonoptimal behavior patterns of industry.

Taking into account equation (24), and the additional constraints given by conditions (21), an alternative set of first order conditions for consumers is developed, which replaces (16):

$$(25) \quad \alpha_i \frac{\partial \omega_i}{\partial y_{ik}} - \eta_k - \rho_k f'_{ik} \begin{cases} =0 \\ <0 \end{cases} \text{ provided } y_{ik} \begin{cases} >0 \\ =0 \end{cases} \quad \begin{matrix} k=1, 2. \ldots N \\ i=1, 2. \ldots z \end{matrix}$$

[6] Negishi has shown that even though such a "perceived demand" relationship is presumed to be equivalent to the actual demand curve of an equilibrium, there exists no mechanism in the model which gives rise to equivalence at the point of the second best solution. Unless the second best solution is known to the * * * [industries] * * * beforehand, it is impossible to perceive a demand curve like * * * [our (24)] * * * which is equivalent to the actual demand curve of the point of the second best solution." Negishi goes on to show one can reformulate the perceived demand relationship such that a consistent equilibrium for a single firm is obtained, but which leaves output in at least one other sector of the economy indeterminate. T. Negishi, "The Perceived Demand Curve in the Theory of Second Best," *Rev. Econ. Stud.*, vol. XXXIV, No. 99, July 1967.

[7] See for example, O. Davis and A. Whinston, "Welfare Economics and the Theory of Second Best," *op. cit.*

The first order conditions of production in *equilibrium* become:

$$(26) \qquad \eta_k - \mu_\theta \sum_{j=1}^{N} (C_{jj} - C_{fj}A_{jk}) - \sum_{j=1}^{M-1} \mu_j b_{jk} - \rho_k f'_k \begin{cases} =0 \\ <0 \end{cases}$$

$$\text{provided } Y_k \begin{cases} >0 \\ =0 \end{cases} \qquad k=1, 2 \ldots N$$

where in both (25) and (26), ρ_k is a LaGrangean multiplier associate), with first order condition (21).[8] Note, by the added constraint (21d

$$(27) \qquad \mu_\theta \sum_{j=1}^{N} (C_{jj} - C_{fj}A_{jk}) = -\rho_k f'_k$$

The adjustment to each consumer's demand for product k is functionally related to the demand for environmental services by producers of k, since:

$$(28) \qquad f'_k = \sum_{i=1}^{z} f'_{ik}$$

Therefore, we can rewrite (25) *in equilibrium* as:

$$(29) \qquad \alpha_i \frac{\partial \omega_i}{\partial y_{ik}} = \eta_k = \frac{f'_{ik}}{f'_k} [\mu_\theta \sum_{j=1}^{N} (C_{jj} - C_{fj}A_{jk})] \begin{cases} =0 \\ <0 \end{cases}$$

$$\text{provided } y_{ik} \begin{cases} >0 & k=1, 2 \ldots N \\ =0 & i=1, 2 \ldots z \end{cases}$$

Here each consumer is taxed (or receives a subsidy) according to his purchases of products requiring environmental services. The greater a consumer's demand for products requiring relatively more environmental services, the larger is his tax (or smaller the subsidy). Note, that f'_{ik}/f'_k can be rewritten in terms of price flexibilities as

$$(\epsilon_{ik}/\epsilon_k)\ (y_{ik}/\sum_{i=1}^{z} y_{ik}).$$

Thus, taxes in this case would be functionally related to: demand relationships for individual consumers as well as for markets, and to the proportion of total demand accounted for by each consumer. In the special case where there is only one consumer, the per unit consumption tax would equal the per unit production tax derived in part A.

---

[8] In deriving these first order conditions, we assume as in the model of materials balances in appendix I that b_{jk} is constant for all j and k, as is $(C_{ij} - C_{fi}A_{jk})$.

The last case we consider is where the constraint set (21) is not entirely immutable but can be adjusted toward the optimal first order condition (20), but not fully achieve (20). An exposition in symbolic form of this hybrid case will not be developed here, but one can conceptually view the problem as follows: Constraints (21) contain a parameter that can be changed, but this parameter is bounded by zero and some positive proportion less than one of $\mu_o(C_{jj}-C_{jj}A_{jk})$. The Government agency then would select these parameters in conjunction with fulfilling the "second best" solutions by appropriate adjustment in the revised conditions (21) and for consumers via adjustments in conditions (16). Thus, a partial removal of discrepancies between private and social costs is possible, but not complete removal such that price of product k is equated with costs including costs of environmental services. We have already argued [9] that *generally* this hybrid case will lead to an improvement in welfare compared to the case of total immutability of constraints (21).

In this case, the Government agency would need to intercede in final product markets (as regards demand curves "perceived" by firms) and in the indirect use of environmental services by consumers. Of course, decentralized decision making could be allowed to prevail once environmental standards and consumer taxes or subsidies were established. The standards and controls would then function as "natural" impediments to the private enterprise system much as antitrust, crop-acreage control, business-practice, or "luxury" tax laws are viewed currently. However, the "second best" types of solutions appear to complicate rather than simplify the sets of rules for constraining "piecemeal" decisions on the part of Government agencies, and private decision making. Applying the "second best" rules in the hybrid case here discussed, an industry would be required to observe demand relations established by Government edict, and consumers along with producers taxed according to demands on environmental services. While decentralized decision making could function within these constraints, it is doubtful if anyone would claim that decentralized decisions had not, for the most part, been emasculated.

[9] See footnote 16, p. 95, text.

Section C

Conditions of Technology, Uncertainty, and Communication and the Role of Government

DECREASING COSTS, PUBLICLY ADMINISTERED PRICES, AND ECONOMIC EFFICIENCY

BY WILLIAM S. VICKREY

William S. Vickrey is Professor of Economics at Columbia University.

When markets fail to be competitive, the price system of the private sector tends to product inefficient results. In modern industrial societies, the prevailing technology often leads to decreasing costs as output increases. If this decreasing cost situation is sufficiently severe, markets will tend to be dominated by but a single firm. In such conditions, public action is necessary to insure that the undesirable results of monopoly are avoided.

In this paper, Professor Vickrey deals with the appropriate social responses to situations in which decreasing costs prevail. Among the responses designed to encourage economic efficiency are public regulation of the prices of firms with decreasing cost technology, operation by a public entity, and also provision of subsidies to permit prices to be established at levels that are more conducive to efficient utilization. Professor Vickrey analyzes and discusses a wide range of issues related to these responses. These pertain to the definition of marginal costs, the financing of the required subsidy if marginal cost pricing is enforced, and considerations of equity in implementing a marginal cost pricing policy. Urban transportation and privately operated utilities are used as examples of major decreasing cost sectors. "A thoroughgoing and imaginative attack on the problem of pricing the output of decreasing cost . . . activities . . . is capable of yielding very substantial returns in increased output, better service, and a higher overall level of real national income."

I. THE SIGNIFICANCE OF DECREASING COSTS

One important class of situation in which the free competitive enterprise system tends to fall significantly short of generating the best possible results is that in which there are increasing returns to scale, or alternatively stated, decreasing costs as the volume of output increases.

In a world of handicrafts and farming, attempts to operate on a larger scale usually tended to encounter difficulties in management, communications, or incentives. No marked improvement in efficiency could be obtained through operating on a larger scale, productive units remained small, and no individual producer had much if any control over the price at which he would be able to sell his output. The results produced by the balancing of supply and demand under these conditions, however much they might leave to be desired from the standpoint of the distribution of income, at least were free from serious waste, given the resources and techniques available at the time. In modern technological societies, however, a large and increasing part of productive activity takes place under conditions where an increase in the scale either permits certain fixed costs to be spread over a larger output, or permits the use of larger and more efficient equipment, or in other ways results in a reduction in cost per unit.

Such cost reduction can be, on the one hand, a relatively sharp reduction on a comparatively small operation, as in the case of the publication of a relatively small edition of a book, where a large part of the cost is in the writing, editing, typesetting, and makeready, with the cost of printing and binding additional copies once the run has been started making up a relatively small part of the total. In other cases the relative reduction may be small, but concerns an important segment of production, as when increased production of a given model of refrigerator or automobile permits the initial costs of design and tooling up to be spread over a larger volume of production or brings within the range of feasibility more elaborate and efficient methods of production and so brings the average cost down.

The most important cases of decreasing cost, however, tend to occur in transportation and public utility operations, where the operations are necessarily spread over a wide area and cannot in their very nature be concentrated in a few locations. This spreading out tends to require relatively large initial commitments of resources in order to extend service to the customers. As the amount of service within a given area increases, the spreading of this initial commitment over more units of service results in a substantial reduction in the cost per unit of service. Thus if for example the traffic on a railway line doubles, the cost of building and maintaining the right-of-way and trackage appropriate to the level of traffic will much less than double: a double track line is capable of handling much more than twice as much traffic as a single track line. Similarly the cost per circuit mile of installing a bundle of telephone circuits goes down fairly rapidly as the number of circuits to be installed along a given route increases. The cost of providing a given number of mail deliveries per day does not rise nearly in proportion to the number of pieces delivered, especially in the less densely populated areas.

The most extreme cases of economies of scale arise with a class of service sometimes called a "public good." Here the use of the service by additional individuals has no direct effect on costs at all. Radio and television broadcasting is perhaps the clearest example, in that the act of tuning in normally has no noticeable direct effect on either the broadcaster or on the other listeners. Other examples are highway bridges on lightly traveled roads where no traffic likely to develop would generate appreciable congestion or call for the construction of other than a minimal bridge, or a theater operator who shows matinee films in an auditorium larger than would ever be filled by those wanting to see the film at that hour. Other examples sometimes cited are the provision of a firework display, or a mosquito abatement program. In such instances the average cost per user or per beneficiary falls in direct proportion to the number of these.

It is important to note that "decreasing cost" in this context refers to alternative levels of average cost for different volumes of service at a given time, and not to any tendency of cost to decline over time through technological innovation, except to the extent that this innovation would not have taken place without the growth in volume. Thus a decrease in the cost of raising corn through the introduction of hybrid seed is not the kind of "decreasing cost" that is relevant here, in that this innovation is almost equally applicable whether the total acreage planted to corn is large or small.

Wherever these economies of scale are of significant magnitude, a serious dilemma is created between the objective of securing an efficient use of the Nation's economic resources and the objectives of adhering to concepts of equity, accounting conventions, price stability, and the decentralization of decisionmaking on the other. Under such conditions the free competitive enterprise system cannot be relied upon to secure the full benefits that it is technologically possible to derive from the situation. Some form of collective intervention is necessary if inefficiency is to be avoided.

II. Decreasing Costs and the Failure of the Free Competitive Market*

A major virtue of a competitive free enterprise system is that it has a tendency to generate prices that reflect the relative costs of providing additional quantities of various goods and services, so that individuals and businesses are ordinarily led, in choosing among alternative ways of accomplishing their objectives, to select the alternatives that accomplish these ends so as to minimize the effort or sacrifice required of the remainder of the community to provide the necessary inputs. Thus the given level of output or satisfaction tends to be achieved at the lowest possible sacrifice in terms of effort or alternative uses to which the resources required might have been put. Where production takes place at constant or increasing costs, a price that will properly reflect the cost of producing the final units of output can be charged and at the same time make it possible for the producing firm to make at least a normal profit, and in many cases leave something over as a special profit or rental for the use of scarce specialized resources. And where the economies of scale are only moderate, a price that will cover the total costs of production plus a normal profit may have to be too high to represent closely the additional costs of additional units of output, but the difference may be small enough to be tolerable and not lead to any gross inefficiency in the allocation of resources.

Where the economies of scale are substantial, however, there is a marked divergence between the level of price necessary to cover the entire costs of the operation (including the element variously referred to as "fixed overhead" or "startup costs") and the price which would properly reflect the costs of additional outputs to prospective buyers and thus lead them to make appropriate choices in terms of the impact of their decision on the economy as a whole. This cost of additional output is usually referred to as the "marginal cost." Marginal cost is sometimes loosely identified with the "variable costs" of the accountant, or in some contexts to "out-of-pocket" costs, but as will be seen marginal cost often contains elements of what is usually termed "fixed cost," insofar as in the long run additional output may require the installation of additional fixed plant or capacity. The costs of providing added capacity will often not increase in proportion to the increase in total capacity, as when a highway is doubled from two lanes to four lanes and the traffic that can be carried without undue congestion and delay is much more than doubled. In such cases

*Further discussion of this issue is found in the papers by Steiner and Arrow in this volume.

marginal cost will not fully cover a proportionate share in these fixed costs.

For these reasons it is usually best, in discussions of this kind, to avoid altogether the use of the terms "fixed cost" and "variable cost" and refer instead to marginal cost, on the one hand, and to the "intramarginal residue" of costs on the other; the latter term designates that part of the total cost that would not be covered by the aggregate revenues generated by prices set equal to marginal cost for each class of output.

Where there are such economies of scale, then, setting price at marginal cost so as to induce purchasers to make correct decisions regarding the extent of their use of the product will entail operating at a loss, and any price which will yield even the bare normal profit will necessarily be such as to cause purchasers to use less of the product then would fully exploit the potential of the situation. Some degree of departure from the free competitive system is thus essential in such cases if substantial inefficiency is to be avoided.

III. Subsidization of Decreasing Cost Industries

A. Marginal cost as an absolute pricing standard*

For a long time it was held by many economists that the proper solution for such situations would be to set prices at the marginal costs in such industries, and to cover the deficits that would result, which would be equal to the "intramarginal residue," out of general tax revenues or from other charges collected in some way unrelated to the consumption of the particular service. Only in this way, it was felt, could users of the service (whether as ultimate consumers or as productive units using them as inputs) be led to make proper choices. For example, suppose a given shipment can be handled by rail at a marginal cost of $100, whereas by truck the cost would be $120. Then unless the shipper would value the added convenience of the truck service at more than $20, the most economical way to ship would be by rail, from the standpoint of overall efficiency. Yet in order to cover its intramarginal residue the railroad might have to charge $150, on the basis of a fully allocated cost including a share in the intramarginal residue, whereas the trucker, whose economies of scale are considerably smaller, might have to charge at most $130 to cover his costs, including a full contribution to his residue, if any. Under these circumstances the shipper would be induced to ship by truck, in spite of the fact that this would cause more resources to be used up in the transportation process, and thus to become unavailable for other purposes, without yielding any comparable advantage to the shipper.

If, on the other hand, the intramarginal residue of the railroad could be financed out of general tax revenues, it was held, so that the railroad rate could be lowered to $100 so as to represent marginal cost, and similarly for the trucking service, then the shipper would be faced with prices that would truly reflect the costs that would be incurred as a result of his choice of one mode or the other, and he would be induced

*Further discussion of this issue is found in the papers by Krutilla and Milliman in this volume.

to make a decision between shipping by rail, using a trucking firm, or operating a truck on his own, that would be consistent with overall economic efficiency, properly balancing the advantage to him of one mode or the other against the respective costs to be borne by the rest of the community.

B. MODIFICATION OF THE MARGINAL COST PRINCIPLE FOR A CONTEXT OF GENERALLY IMPERFECT COMPETITION

There are, however, two major difficulties with applying this solution of pricing at marginal cost in its pure and simple form. One is that mild degrees of decreasing costs with their attendant tendencies to imperfect competition and monopoly, with prices significantly above marginal cost, constitute a large proportion of modern industrialized societies. To deal with all such cases by attempting to bring prices down to marginal cost everywhere and to subsidize the intramarginal residues of each such industry would involve a vast expansion of the role of government and a severe reduction in the role of free competitive enterprise. On the other hand, to apply simple marginal cost pricing only to a selected group of industries where the economies of scale are particularly striking and the inefficiencies resulting from unsubsidized or unregulated prices are serious would go too far in the opposite direction. Pricing such service exactly at marginal cost produce a situation where some services priced at marginal cost are competing in the market more or less directly with services priced significantly above marginal cost. This would result in a substantial, though relatively smaller, distortion, of the same character as, but in the opposite direction to, that produced by full-cost pricing of the decreasing cost services.

Thus, for example, if in the above case the superior service of the truck were worth $25 to the shipper, but the railroad service were priced at a marginal cost of $100, while the truck charge was $130, including a $10 margin above the marginal costs, the shipper would be inclined to choose the rail shipment, whereas the overall efficiency, making due allowance for the added convenience to the shipper, would be better served by his using the truck, the $25 worth of convenience to the shipper being greater than the added $20 in the marginal cost of the shipment. A proper choice between these two modes of shipment would be induced if the railroad rate were subsidized to the extent of $40 rather than $50, so as to bring the rate down to $10 and make the difference between the truck rate and the rail rate the same as the difference between the truck marginal cost and the rail marginal cost. While this would be better than leaving the rail rate at $150, at least from the standpoint of inducing correct choices of mode by shippers, it would still leave some distortion, in the case of the shipper whose alternative was to use his own trucks, where the cost to the shipper would not contain any excess above the marginal cost comparable to that contained in the rate quoted him by the common carrier trucker.

This situation illustrates what is sometimes referred to as the general principle of second best adaptation, which in effect states that whenever there is a general rule, such as setting price equal to marginal cost, that cannot be applied uniformly throughout all parts of an interrelated system, such as an industrial economy, then the best adaptation

that can be made in controlling a selected part of the economy is likely to be one that will also depart somewhat from the general theoretical rule. In effect, if the rule cannot be followed exactly everywhere, it is generally best not to try to follow it exactly at any given restricted area, but to allow for some appropriate deviation. In the case of marginal cost pricing, the application of this second best principle would call roughly for prices in the controlled sectors to be set above marginal cost by an amount corresponding to the differentials prevailing on the average in the uncontrolled sectors, with the greater weights being attached to the differentials encountered in the goods and services most closely competitive with those of the controlled sector. While it is sometimes claimed that allowing such margins would result in a policy too complex to be carried out in practice, actually rough rules of thumb for carrying out such a policy can be developed that would produce results reasonably close to the best that would be theoretically possible, given the constraints against tampering with the uncontrolled sector, and in any case vastly superior to so-called full-cost pricing.

C. THE PROBLEM OF FINANCING THE INTRAMARGINAL RESIDUE

i. *Costs of Subsidizing From General Tax Revenues*

A more serious obstacle to the carrying out of a policy of basing prices on marginal costs, with or without the addition of such a representative margin, is the fact that the funds necessary to provide the corresponding subsidies are not freely available without cost, but must in practice come from taxes that themselves have adverse effects upon economic efficiency. These adverse impacts arise not only from the added administrative costs of collection and added costs of taxpayer compliance, but in terms of the impact of the taxes themselves in interfering with the making of economic decisions in an efficient manner.

The importance of this impact of taxes can be summarized in a coefficient or ratio termed the "marginal cost of public funds," which is, conceptually, the ratio of (1) the dollar equivalent of the burden imposed on taxpayers by the collection of a given increment of taxation to (2) the net revenue, after all additional expenses of collection have been met, derived from the tax increase. It would of course be futile to attempt to improve the efficiency of the subsidized industries by means that would involve even greater inefficiencies induced by the collection of the taxes required to finance the subsidy. The circle would indeed be complete if, based on the plea that the subsidy should be paid for by taxes on those who benefit, the tax increase took the form of an added excise tax on precisely the service being subsidized: this would, of course, defeat the entire purpose of the subsidy.

There are, however, marked differences in the marginal cost of public funds obtained from various sources, and indeed, if we look at the increasing of the rates charged for a given service and the resulting reduction in the required subsidy as a kind of tax, it will be found that in many cases such revenues have a very high marginal cost indeed. Suppose, for example, that one considers an increase in the fare on a municipally owned transit system from 15 to 20 cents, and that as a consequence of this 33-percent increase in the fare, ridership falls off by 10 percent. Then for every 100 riders who previously paid 15 cents, yieding a gross revenue of $15, there will under the new fare be only

90 riders paying 20 cents, yielding a gross revenue of $18, or a net increase of only $3. The 90 riders who continue to ride will, however, be burdened with an added 5 cents each, or a total of $4.50, and in addition the 10 former riders who have been driven to some other alternative that they consider inferior to the transit ride at 15 cents, yet preferable to the transit ride at 20 cents, will be worse off by amounts ranging between 0 and 5 cents, average about 2.5 cents, or a total of 25 cents, which when added to the burden on the continuing riders brings the total burden to $4.75, or about $1.58 for each $1 of gross revenue, and if operating costs are unaffected, as they well may be, this is also the marginal cost of net revenues.

To be sure, the municipality can respond to the reduction in patronage by reducing the service, but this saving in cost will be to a large extent offset by the consequent deterioration of the value of the service to the remaining riders because of the reduced frequency. Reducing service, indeed, could induce a further reduction in patronage and loss of revenue. It seems clear, therefore, that under such circumstances the marginal cost of public funds obtained from increasing transit fares is of the order of 1.5 or more. While most large city governments are fairly hard pressed financially, it seems almost certain that substantial funds could be obtained from other sources at a much lower marginal cost than this, even under the most adverse circumstances. Moreover, in this case an increase in transit fares, considered as a tax, is one of the most regressive taxes of any significance, in that approximately the same dollar burden is imposed on individuals drawn from a wide range of income levels and ability to pay.

The prescription that one gleans from these two considerations is that in the pricing of a service subject to sharply decreasing costs or showing substantial economies of scale, one should first consider a price exceeding marginal cost by a margin corresponding to the margin existing in competing services, and then increase the price further to the point where the marginal cost of obtaining additional net revenue from the increase in the price is roughly equivalent to the marginal cost of obtaining net revenues from such other sources as would be the likely candidates for adjustment in increase in the overall budget of the government concerned.

ii. *Subsidy Through Tax Abatement: Preservation of Independence*

Once one has gone this far, one may find that in cases where the economies of scale are not too extreme, one is not too far from a set of prices that would enable the operation to be carried on on a self-liquidating basis. Or if this is not the case initially, one might find that by suitable adjustment of the tax structure, with exemptions from taxes where this can be done without unduly impairing the integrity of the tax, a reasonable approximation to this situation can be brought about. One can certainly, under this heading, begin by abolishing any special taxes on the utility operation as such, such as any special franchise taxes, gross receipts taxes, or special excise taxes; property taxes on improvements and personal property actually used in providing the service can also be abated in most cases without causing serious difficulties of definition or wasteful incentives. Taxes on land and on corporation income may be another matter.

If by tax abatement or otherwise the operation can be brought to a point where the prices otherwise desirable on efficiency grounds are not too far from covering total costs, there may be a distinct advantage, from the standpoint of creating greater independence of administration and providing incentive for operating efficiency and economy, to increasing the rates sufficiently further so that the element of explicit subsidy is largely eliminated. Even if this cannot be done without pushing prices too far from their efficient levels, it may be possible to come sufficiently close to the efficiency level of prices by arranging for a subsidy to be provided to an extent to be determined independently of the manner in which the service is supplied. The objective in this case would be to leave management with full freedom to operate on a reasonably independent basis, without having to be concerned on the one hand about whether a contemplated action might result in an increase or decrease in the subsidy, or on the other about having to justify its outlays to the agency that would be called upon to foot the bill. Moreover if operations-dependent subsidization is eliminated, it may be somewhat easier to allow the operation to be carried on by private enterprise rather than by a Government agency, and this may under some circumstances be considered to be an advantage.

It is essential to the operating efficiency of the enterprise, however, that the subsidy be actually independent of the manner in which the service is supplied, and not take the form of payment for some particular category of cost. Making the subsidy cover some specific element of cost can be disastrous to the economical use of resources, as has occurred with the New York Transit System. There the fact that expenditures classifiable as capital outlays could be subsidized out of general city revenues through assumption of the capital charges, while operating expenses had to be met from revenues, has resulted in current service being cut to the bone while outlays for equipment, appurtenances, and new facilities have been laid out and programed on a relatively lavish scale.

iii. *The Use of Discriminatory Pricing*

Possibilities for developing a self-liquidating rate structure that would bring the allocational efficiency of the operation reasonably close to the absolute maximum that might be achieved with the aid of a subsidy are increased when certain types of discrimination in the rate structure are introduced. Perhaps the best example of this is the electric power industry, where by judicious use of the classification of customers and of block rates, sometimes refered to as a "multipart tariff," a substantial amount of the intramarginal residue is covered in ways that do not impinge seriously on the marginal consumption of the larger consumers. Even here, however, many small customers remain at points on the schedule where they are paying a rate at the margin far above the marginal cost of supplying them with more electricity, while others fail to take full advantage of their opportunities because of their lack of understanding of the price they would have to pay for additional use, given the complexity of the rate schedule, the intangible nature of that which they are buying, and the difficulty of correlating the size of the bill at the end of the month (or in some cases, 2 months) with particular uses of electricity.

Even here, the degree to which the existing pricing system falls short of producing optimum results is suggested by the wide variations in rates for comparable service that are charged in various parts of the country. These variaitons are far greater than can be accounted for by differences in the cost of power at the bus-bar, or differences in the characteristics of the neighborhoods serviced, or environmental difficulties in installing the distribution network. In large measure they reflect attitudes toward the setting of rates ranging from highly consumer-oriented attitudes in the case of the distributors of TVA power, to highly profit- and security-oriented attitudes in the case of many of the privately operated utilities.

iv. *Pricing of Utility Services According to the Size of Premises*

Actually, much better results could be achieved in electricity distribution by introducing into the domestic rate structure a variable block size that would vary with the size of the premises served. Thus for a large house the amount of current that would be paid for at the 6-cent rate, before the lower rates would begin to apply, would be greater than the amount in this block for a smaller house. With judicious determination of such a rate schedule, even the relatively poor would, if living in correspondingly small quarters, have available to them a low marginal rate for additional consumption. Such a rate schedule would also have a more favorable impact on individuals of low income than present rate schedules.

Something of this sort is indeed done in some cases where the blocks are in terms of the total connected load, but in this case the concept behind the rate is more nearly one of adjusting for the extent to which the consumer contributes to peak loads rather than an explicitly discriminatory one. Moreover this form of charge may act as an undesirable deterrent to the installation of additional equipment, often without relation to any likelihood that the installation of the equipment would in fact result in heavier peak loads. Actually, even the variation of the size of the rate blocks in accordance with the size of the premises would to some extent constitute in effect an extra charge levied for living in larger quarters, and to this extent would have a tendency to induce undue stinting on housing space. The amount of this charge, however, would be small relative to the other costs of expanding one's living space, and it does not seem likely that the effects of this sort would be sufficient to outweigh the overall advantages of such a rate scheme in relation to better utilization of electric power distribution facilities. These side effects would, of course, on general second best principles, indicate that this method of pricing should probably not be pushed all the way to allowing the final follow-on rate to be reduced quite to the level of the marginal cost of additional electric power.

v. *Financing Utility Services by Charges on Land Ownership*

Plans such as the above are possible for electricity by reason of the use of at least some electricity being a virtual necessity: high initial block rates are unlikely to induce any significant number of customers to decide to do without electricity altogether. Such plans would probably be possible for water supply as well. And while such methods can also be applied to gas or telephone service, there is here a much greater danger of driving some customers to give up the service entirely, so that the degree to which such discrimination can be applied is much more limited.

There is, however, a good deal to be said for an even more radical method of covering the intramarginal residues of utility services of this type, at least in part. This is to levy a charge, possibly on a front-foot basis, on all owners of property past which the distribution mains run, whether or not they actually take the service. The justification for making such a charge to persons who do not actually use the service would be somewhat similar to the practice of requiring the purchaser of a car to pay a price covering the cost of the headlights that necessarily come with the car, even though he intends never to drive at night, or the cost of a body and engine capable of carrying six persons at 90 miles per hour, even though the purchaser never intends to carry more than two persons or drive more than 60 miles per hour, simply because those are part of the specifications of that particular type of automobile. In a similar manner, the availability of utility services is a part of the specifications that come with a particular parcel of land, and should be paid for whether or not used.

Charging on such a basis could appropriately be extended not only to the conventional utilities, but to mail service, garbage collection, parcel delivery services, cable TV, and even, though possibly on some other basis than frontage, fire protection and transit service. Assessment of charges of this sort for occupancy of property having these services available whether or not they are used would go far to discourage inappropriate land use, whereby potential occupiers that do need these services are forced to move elsewhere and require costly extensions of the services into new territory. To some extent leap-frogging and sprawl would be discouraged and the development of urban areas induced to follow a more rational pattern.

In many cases, however, the possibilities for abating the inefficiencies resulting from prices too far above marginal cost through such multi-part pricing, discriminatory pricing, and other devices are limited, so that even with the best possible price policy, substantial inefficiencies will result under any scheme that is in any significant sense self-liquidating. Urban transit is also certainly an example of this, and intercity transportation may be another. Other possible candidates are telecommunications services, cable TV, and newspaper delivery.

D. THE PROBLEM OF PRESERVING MANAGERIAL INCENTIVES

In such cases, the problem must be faced squarely as to whether the added possibilities for inefficiency and lax management introduced by the payment of a subsidy are so serious as to warrant the toleration of the allocational inefficiency inherent in operating such services on a self-liquidating basis. A great deal will depend not on pure economics, but on the political realities of the situation. One danger is that public operation on a subsidized basis may lead to a tendency to accede to demands of customers that the service be expanded beyond the point where the benefits conferred would fully justify the added costs, given that the customers would not themselves be bearing the full costs, or even, in some cases, the full additional costs of the increment. Another possibility is that the operating agency may not be as keen to hold costs down as it would if it had to finance the operation directly out of the charges that it was levying itself. Still a third possibility is that the operating agency may spend an excessive proportion of its administrative effort in attempts to justify an increase in the subsidy and hence in the possible scope of its operations.

These problems will exist to some extent whether the operation is a private one or is carried on by a public agency, but may be to a degree more severe where there is a distinct interest represented by a private management responsible in part to private shareholders. For this reason a decision for subsidized operation may entail public rather than private operation.

If the decision is for subsidized public operation, we can then be content to have the rates set by the application, without further modification, of the principle of equalizing the marginal cost of public funds secured by various rate increments with each other and with the general marginal cost of public funds secured from other sources. If, however, the operation is to be placed on a self-liquidating basis, or is to be subsidized only to a limited extent defined independently, such as by abatement of or exemption from taxes, or by the dedication of certain revenues, it will no longer be possible to secure equality of the marginal cost of net revenues to the operating agency with the marginal cost of public funds generally for the responsible government.

IV. Ratemaking in Decreasing Cost Industries Required To Be Self-Supporting

A. Equalizing the Marginal Cost of Net Revenues

If the independent operation has only a single price under its control, as is substantially the case, for example, with many transit operations having a single flat fare, the level of the fare is determined by the financial requirement, and there is no further room for discretion. In most cases, however, there will be a number of interrelated rates capable of being adjusted, and some principle is needed for doing this. One way of doing this would again be to use the principle of equalizing the marginal cost of net revenues for each of the rates under the control of the agency, but in this case not requiring that this internal marginal cost of net revenues be equated with the marginal cost of public funds. In effect, as long as one can get additional net revenue of $100 from customers in group A at a cost to them of only $120, while it would require a burden on group B of $130 to get the same amount of net revenue for the operating agency, it will be possible to decrease the aggregate burden imposed on the customers in order to met a given net revenue requirement of the operating agency by increasing rates to get additional net revenues from group A and lowering the rates on group B so as to give up a corresponding amount of net revenue. If this procedure is carried out until equalization of the marginal cost of net revenues has been achieved, the aggregate of benefits to customers will have been maximized.

B. Equity, Efficiency, and the Overlapping of Oppositely Affected Groups *

Justification for such a rule does of course require that one be able to state that a dollar of benefit to group B is to weigh as much as a

* Further discussion of this issue is found in the papers by Weisbrod, Bonnen, and Freeman, in this volume.

dollar of burden to group A in the process of decisionmaking. This would obviously not be valid if group A were known to be significantly underprivileged relative to group B, for example. In practice, however, it can often be stated with some confidence that there is no significant difference between the income levels, on the average, of the two groups. Indeed, the two groups may actually consist in large measure of the same individuals in different capacities, and in extreme cases the two groups may be almost indistinguishable, especially if one considers that equity should properly not be concerned with business firms as such but only with the ultimate individuals who are the customers, employees, or stockholders in the various firms.*

For example, if the question is one of increasing rush hour transit fares and correspondingly lowering offpeak fares, it must be considered that many rush-hour riders also ride at other times as well, or if not, members of their families do. Or if the question is one of adjusting relative railroad rates on steel sheets on the one hand and coal on the other, by the time the freight rates have been absorbed in costs and passed on to ultimate consumers of electric power, canned goods, automobiles, and the like, it is likely that few if any ultimate consumers would be able to tell, in the long run, that his own share of say a $100,000 additional burden imposed on shippers of steel sheets would be greater than his own share of say a $120,000 saving thereby made possible to shippers of coal.

Moreover, even though one might in any particular stuation be able to say that the two groups of ultimate consumers are distinguishable, and that in some sense it would be unjust in the particular instance to impose a burden on one group even for the sake of a substantially greater benefit to be conferred on the other, yet when one considers not merely the particular decision in isolation, but the net result of a large number of similar decisions, each to be made on the basis of maximizing the net total balance of gains minus losses, one can have a quite different feeling about the matter. Even though an individual might come out on the short end of one of these decisions, there will be others in which he gains and probably gains relatively more. If in effect one thinks not of making a particular decision in isolation, but rather of deciding on a policy to be carried out in making a whole class of decisions ranging into the indefinite future, it is likely that few, if any, individuals would be found, after it is all over, feeling that they had come out worse in the end as a result of consistently applying a rule of balancing dollar gains and losses in this way, as compared with any other rule likely to be applied consistently. And even if there were a few who could look back after the event and say they did worse, if instead we look at the situation prospectively, as indeed all decisionmaking must do, there will be practically no individuals, even provided with all of the relevant information available at that time, who would be able to correctly anticipate with any confidence that they would be worse off under this policy than under another one.

Thus here more than in most situations, there is good reason to look to economic efficiency rather than to be held back by rules of justice or concepts of equity applicable to the individual case. The individual

*Further discussion of this issue is found in the paper by Krutilla in this volume.

who correctly understands that it is proposed in the name of justice to take $11 from his left-hand pocket in his capacity of coal consumer, directly and indirectly, in order that $10 may be put in his right-hand pocket in his capacity of indirect consumer of steel sheets is likely to have a fairly low opinion of such "justice."

This is not to say that there may not be cases where the impacts are so immediate, glaring, and substantial, and the sense of injustice generated so acute, that efficiency may be set aside for the sake of justice. If, for example, one builds a new rapid transit line in a given direction, and proposes that since this new line will be, for a time at least, very lightly loaded, so that marginal cost of carrying additional passengers is low, and that the fare should be correspondingly low, while marginal cost on crowded lines in other directions calls for a correspondingly higher fare, this is likely to strike many as piling insult on injury. It may be regarded as a sufficient, though perhaps necessary discrimination if, simply because one must do one thing at a time, one particular area is singled out for the new line; to give the beneficiaries of the new line a lower fare in addition while maintaining the higher rate on the older crowded line many not only provoke hard feelings but make the original decision as to the location of the new line or the sequence in which the new lines are to be built a politically disruptive one.

In such cases the felt injustice involved in adjusting rates for maximum efficiency and the political tensions generated may have to be taken into account. The redistributive effects in such a case may be so substantial and so concentrated on identifiable individuals as to make it unlikely that adequate compensation would come from the random occurrence of countervailing decisions resulting from applying the efficiency rule in other cases. One possibility, of course, would be to cover some of the cost by levies on the property affected, as suggested above. But if this cannot be managed, it may be necessary for the sake of avoiding substantial dissatisfaction with the equity of the situation to retain some uniformity of fare structure. In doing this, however, it must be realized that one is incurring a significant degree of economic inefficiency through encouraging overutilization of the old facility and underutilization of the new, including a tendency to cause development along the new line to be less well adapted to the facilities being provided than would otherwise be the case. Moreover, it is important to note that deviations from the efficiency rule are justified in this manner mainly in cases where there is a very substantial direct impact on identifiable individuals. No such deviation is justifiable on grounds of impacts on industries or businesses, generally speaking.

V. Equity and Efficiency in Urban Transportation *

A. The Subsidization of Commuters

A similar issue can arise in considering, on the above principles, the subsidizing out of general revenues of a service used predominantly by wealthier groups. An example of this would be the subsidi-

*Further discussion of this issue is found in the paper by Nelson in vol. 3 of this collection.

zation of rail commuter service to the wealthier suburbs, as distinct from local transit service. In this case it might well be held unconscionable to provide additional subsidies out of funds that at the margin would be derived largely from added taxation impinging on lower incomes, given the difficulties in obtaining any substantial additional amount of revenues from upper income groups without a major overhaul in the entire philosophy of the income tax structure, particularly in relation to the various loopholes with which it is riddled. The situation here is particularly acute in that such subsidization would be added to other inequities that run in the same direction, such as the tax advantages accorded homeowners relative to tenants and the avoidance by many suburbanites of a large part of their share in the burden of supporting metropolitan amenities and obligations.

In practice, however, the question of such subsidy to suburban mass transportation must be considered in the light of the even heavier subsidization of the alternative mode of commutation by private car. Failure to subsidize mass commutation facilities may merely mean that more traffic is diverted to highways, generating added congestion impinging not only on the commuters themselves but on others caught in the general congestion. This in turn is likely to lead to a demand for additional facilities constructed in considerable measure with Federal funds or with funds derived from motorists using the facilities at noncongested times and places and who will derive relatively little benefit from expansion and duplication of facilities for commuters. Thus the end result may be that avoidance of a more obvious subsidy to suburbanites through subsidized mass transportation may merely lead to a far greater, though less patent, subsidy to these same commuters as motorists.

Moreover the existing and prospective subsidies to private car commuters are so great that in many cases not even the offer of free mass transportation service would suffice to produce an efficient choice of mode of travel as between private automobile and mass transit. The subsidy, computed in marginal terms, to the private car commuter often far exceeds the entire marginal cost of the mass transit service. If the private car trip involves a subsidy of $1 per trip on a marginal basis and the marginal cost of the corresponding transit trip is $0.60, it would require the payment to the passenger of a bonus of $0.40 for each rush-hour ride on the mass transit service to cause the costs borne by the commuter to correctly reflect to him the cost differential involved in his choice between the two modes, so that he would be induced to make a proper choice in terms of weighing his own preferences against the resources used in making the two alternatives available to him. And even if such a bonus could be financed, it would on the other hand bias the choice between living in the city or close to the job and living in the suburbs, or between taking a job in the city and commuting versus finding a job in a decentralized office in the suburbs.

B. THE PRICING OF FACILITIES THREATENED WITH CONGESTION

There is, indeed, no adequate solution to the problem of inducing efficient patterns of commuter travel and providing the facilities for satisfactory service at minimum cost that does not involve in one way or another imposing a specific charge on the private car commuter that

will bring home to him the full social costs of his trip. These costs should in principle include not only the full cost of duplicating or enlarging facilities to take care of peak hour volumes of traffic, as well as a major share of the cost of routes and facilities which also provide substantially improved service for offpeak users, but additional elements to reflect the contribution of automobile traffic to air pollution, and the cost of automobile accidents, to the extent that these are not fully covered by insurance or are covered in a way that does not suitably impinge on the decision to make the particular trip. These costs must of course be brought home to all users of the congested facilities at congested times, and not just to the regular daily users.

Such a suggestion may to some conjure up visions of toll booths sprouting up everywhere, with the queues at the booths and the salaries of the toll collectors eating up most of whatever benefits might be derived from improved management of the traffic. There are, however, a number of alternative devices for the collection of such tolls without excessive expense for toll collectors or delay or inconvenience to motorists, with a degree of flexibility and adaptability capable of varying the charges so as to provide a reasonable approximation to the desired optimization. These devices range from simply requiring all vehicles entering designated congested areas to display a license marked, torn, or chemically activated by breaking a capsule, so as to indicate the time and place for which it is being made valid, with or without provision for returning used licenses for credit of the unused value; through meters on cars actuated by pulses emitted from roadway cables, to sophisticated scanning or locator systems that would permit a bill to be computed for each vehicle and mailed to the registered owner monthly. Institution of such congestion tolls would not only facilitate a rational solution to the problem of the commuter, but would vastly increase the efficiency and utilization of the downtown street facilities.

Indeed, it is not usually realized what a drastic improvement can be achieved with relatively slight reductions in the amount of travel actually performed. If, for example, we have a reduction of only 10 percent in the vehicle-miles of travel attempted over a given street network, this can easily result in a 25 to 50 percent reduction in the number of cars attempting to move at any one time, corresponding to an increase in speed of 20 to 80 percent, depending on the severity of the initial congestion conditions. When to this is added the provision of adequate incentives to move as much of the freight as possible at night and other uncongested periods, and deterrents to the preemption of street space by excessively protracted parking and double parking and by the use of excessively large vehicles, it should be clear that an almost revolutionary improvement in downtown traffic conditions is within reach in relatively short order. Thus the use of congestion charges will not only reduce the need for subsidy to the commuters from the wealthier suburbs, but will have a crucial revitalizing impact on the core cities and the central business district. Nor is the contribution of the resulting revenues to hard-pressed city treasuries a negligible element. Here, indeed, is one source of revenue with a marginal cost of public revenue of much less than 1.0. In many instances, indeed, the motorists will pay the added charges and find themselves better off than they were originally, so that the marginal cost of public revenues may be zero or even negative.

VI. Regulation of Privately Operated Utilities

A. The Need for Limitation of the Exploitation of Natural Monopolies

There remain those cases where for one reason or another it seems desirable to leave the supply of the service in question in private hands. In such cases the existence of substantial economies of scale would, in the absence of regulation, lead to the development of an intolerable degree of monopoly power. In cases such as this the ability of the uncontrolled monopoly to raise prices and exploit the consumer is far greater than that existing in cases where monopoly exists without strong economies of scale, since in the latter case the ever-present threat of entry by new firms places a limit on the degree to which monopoly power to exploit can be exercised over any long period. Here the economies of scale mean that costs of potential small scale competitors would be higher than those of the monopoly, making it much more difficult for them to compete and leaving the monopoly free to raise prices considerably above its own costs. Accordingly, some form of price regulation is essential if the greater efficiency of the private operation is not to be completely outweighed by the allocational inefficiency induced by monopoly pricing, to say nothing of the inequity involved in such exploitation.

B. Methods of Securing Equity for Shareholders

Regulation of the price or a utility, however, introduces a new element into the situation: the problem of securing equity as between the shareholders and the consumers. Basically the problem is one of making it clear, at the time the stockholder makes his investment, whether indirectly in the form of reinvestment of earnings or directly in the form of purchase of shares, just how much and under what conditions he is entitled to receive as a return on his investment. As long as the terms of this implicit contract are reasonably clear and cover substantially all of the more likely contingencies, and the terms of the contract are adhered to, the investor can have no cause to complain. The difficulty is that there is often a conflict between the selection of a type of contract that would be most conducive to such certainty and one that would permit the closest approach to economic efficiency.

In the United States, the problem has in the past been complicated by the intervention of the courts in making a 14th amendment issue out of the matter. Given the fact that most active regulation was initiated after considerable investment had already been committed in the absence of any clear contract, implicit or otherwise, and that even where franchise contracts existed their validity could often be questioned on grounds of undue influence, bribery, or fraud, the question of determining what investors were entitled to tended to be regarded as a matter to be determined *de novo* as of the current instant without too much regard to the past. The doctrine arose that the investors were entitled to receive the equivalent of a "fair return" on the "fair value" of the property "used and useful" in the public service. How this "fair value" should be determined, however, has been a matter for great uncertainty and controversy. An appraisal in terms of what a

willing buyer would pay for the property, as is usual in property tax assessment, will not do for this purpose, since this value would depend on the rates to be established, leading to a circular self-justification of almost any value. Nor, for the same reason, could the market value of the shares be used for this purpose, though they might be used more directly as a test of the reasonableness of the rates. While there seems to have been a tendency on the part of many regulatory authorities to be almost deliberately vague concerning the basis on which they have determined this "fair value", possibly from a not altogether groundless fear of seeing their determinations overthrown, and practice varies widely from one jurisdiction to another, discussion of fair value determinationfor ratemaking purposes has generally focused on two distinct concepts, that of "current value" and that of "historical cost", plus some minor variants, one of the more important of which is that of historical cost adjusted to allow for price level changes.

i. *Current Value as a Rate Base*

The current value rate base was originally thought of as reflecting what it would cost to reproduce the physical plant under current conditions, with due allowance made for the difference between the value of a new plant and that of the partially worn and used plant. But when it was applied to specific cases it became clear that in many instances the same type of plant would not in fact be constructed were an attempt to be made to provide the service anew from scratch. This problem extended not merely to the replacement of brick walls by reinforced concrete, or to the laying of new gas mains under a shoulder or park strip rather than tearing up pavement to reproduce the old mains that had been laid before the street had been paved, but of installing entirely new types of system with different operating costs and characteristics, such as microwave instead of carrier current systems, or higher pressure and temperature turbines and boilers rather than the older less efficient ones, and the like. To make any sense at all, the current value concept had to be redefined in terms of the value the old plant would have in competition with or as an alternative to fully modern plant capable of rendering the same service, with necessary adjustments for differences in operating cost. And in some cases, even the service rendered by a modern plant constructed *de novo* would have different characteristics and a different value and market.

When thus pursued to its logical conclusion, the concept of current value becomes one that is extremely difficult and costly to evaluate in practical terms. Where cases have been taken through the courts, the delays have often been such as to render the data collected obsolete by the time the decision is reached. And at best the manner in which the concept would be evaluated could hardly be said to give the kinvestor any real feeling of security, which in turn would tend to result in an increase in the cost of securing additional capital funds, because of the risk engendered by this uncertainty. It is hardly too much to say that to the extent that reliance has been placed on current value as a basis for rate regulation, regulation has been a rather dismal failure.

ii. *The Historical Cost or Prudent Investment Approach*

In principle, the use of the historical cost rate base places the investor in somewhat the position of a preferred stockholder, entitled to a

fair return on the book value of the property used in providing the service, subject only to the condition that on the one hand the investment be made with reasonable care and prudence in the light of the information available at the time, and on the other to the condition that the demand for the service be sufficient to permit this return to be earned at rates that would be reasonable as to level and not unduly discriminatory among the various classes of customers. If the demand for service should fall off severely, as for example happened with streetcar lines, so that maintaining the return to the investor would be either entirely impossible or possible only at rates that would be considered entirely exorbitant, then the investors would be required to shoulder a loss, the possibility of which would have to be considered in determining what constitutes a fair rate of return. While some uncertainty may still remain as to what is exorbitant and what, if any, lower return the investor is to be entitled to under such circumstances, this uncertainty concerns only what is usually a fairly remote contingency, and is in effect the minimum amount of uncertainty consistent with any reasonable degree of private autonomy for the enterprise, unless, indeed, there is to be some form of government guarantee to bail the investors out in the event of such a severe decline in demand.

In one respect, however, the position of the utility investor may differ from that of the preferred shareholder, in that while the preferred share normally specifies a fixed dividend rate, the rate of return to which the utility shareholder is entitled is ordinarily left to be adjudicated in the light of current market conditions. Presumably the rate of return allowable would be higher when interest rates are high, and conversely. The multiplicity of market interest rates prevailing at any one time, the uncertainty of the degree of risk to be assigned to the utility investment and the corresponding selection of comparable market rates, and especially the existence of fixed interest indebtedness, with the attendant leverage and especially the pervasive baneful influence of the corporation income tax introduce an element of uncertainty at this stage that is almost as great as that involved in the determination of a replacement cost rate base, though with less of a tendency to generate prolonged procedural delays. The problem could readily be solved, at least in principle, by agreement in advance at the time the investment is made that the rate of return should be either a specified percentage or should be determined by reference to some specified index of interest rates. But it would still be necessary to leave some margin above the minimum rate necessary to attract capital in order to be sure that the utility would be able to finance needed expansion without impairing the equity of antecedent investors. Such explicit advance agreement has not been a frequent practice, however.

Actually, while in principle the use of a historical cost rate base should result in a constant money dividend to shareholders, or at least one varying only with general changes in interest rates and the leverage in the capital structure of the company, actual application of the principle seems to fall short of this for a number of reasons, among them the time lags and delays in the regulatory process, the intermingling of regulated and nonregulated business in the accounts of the utility, and inconsistencies between the acts of different regulatory bodies. For example, in spite of the fact that A.T. & T. had

paid a $9 dividend steadily from 1922 to 1958, and the FCC had declared its adherence to a historical cost rate base, the effective dividend rate was thereafter increased, subsequent to a stock split, reaching the equivalent of $13.20 in 1966, with nothing in the overall trend of interest rates or the capital structure of the company to warrant such an increase.

In many jurisdictions outside the United States, the entire rigmarole of rate base and rate of return determination is short-circuited by the simple device of requiring that dividends be limited to a fixed amount, with new capital requirements obtained by sales of shares on a competitive basis. Any accumulation of a capital surplus in excess of a normal reserve level is then the signal for a rate reduction, and any depletion of the surplus below a suitable level is the signal for a rate increase. But to work properly, such a method requires that all of the activity of the operating corporation be brought under a single regulatory rule.

iii. *Adjustment for Changes in the General Price Level and in Interest Rates*

Whether applied in terms of rate of return and rate base, or in terms of direct control over disbursements to shareholders, one major difficulty with such a historical cost method of determining utility rates is that in the event of substantial price inflation (or deflation) the capital charges included in the cost of the service will fall severely out of line with other prices, and in extreme cases could lead to rates significantly below the optimum level from the standpoint of resources in the case of inflation, or, in the case of severe deflation, to rates that are excessively high. There is also the claim on behalf of the investor that in the case of inflation he is being deprived of the expected real return on his investment, though here it can be pointed out, at least in the case of a well-defined policy, that he is in no worse a position than the investor in a bond or other nonconvertible obligation with a return fixed in money terms. Thus it is primarily the impact on resource allocation rather than the impact on the investor that is of crucial significance here.

The problem of changes in the general price level can be adequately taken care of by adjustment of the historical cost or of the allowable dividend by the application of an appropriate index number of general prices. The only requirement is that the adjustment should be one clearly agreed on in advance, and that it should be applied evenhandedly in both directions and not become a case of heads I win, tails you lose. It is, to be sure, not at all clear whether, in terms of current capital markets, a security offering a return fixed in money terms or one fixed in proportion to a general price index would result in a lower cost of capital on the average, but whatever difference there is is probably of minor importance.

In addition, even where regulation is based on control over dividend disbursements, it may be appropriate to make adjustments in the dividend rate based on variations in market rates of interest. Provided that an appropriate index can be found that is sufficiently robust to stand the pressure possibly involved in making it the criterion for a substantial variation in price and rate levels, and that the manner of computing the adjustment is clearly specified at the time investment

commitments are made, there should be no difficulty with such an adjustment either. Such an adjustment would have a considerable beneficial effect in bringing utility rates more into conformity with what efficiency requires, causing them to move more in conformity with the rates of other substitute commodities and services.

iv. *Adjustment for Obsolescence*

A somewhat similar deviation of the regulated price level from what would be optimal in the light of current circumstances occurs when technological developments drastically alter the economics of the situation. Ideally, when a new technological development supervenes that drastically lowers the cost of providing a service, economic efficiency would call for lowering the price immediately to the full extent of the cost reduction. If the technological development was correctly foreseen a sufficient time in advance, it would be appropriate to have increased depreciation charges during the period of anticipation, so as to write off the obsolescent investment, by the time the new development is ready for application, to a value in terms of which this old investment yields services at a total cost comparable to that obtained for services produced with the new equipment. Raising rates to cover such increased depreciation charges would also appropriately tend to curtail demand during the period when providing for additional service would require installing additional equipment of a type shortly to become obsolete.

For example, suppose a unit of coaxial cable costing $100,000 can provide 10,000 units of service a year, and will have a physical life of 20 years, so that a charge of $1 per unit of service over the 120 years would provide an annuity of $10,000, sufficient to amortize the cost at 7.5 percent interest. If operating cost and maintenance per unit of service amounts to 50 cents, the total cost per unit of service will be $1.50. Suppose, further, that at some point of time it becomes apparent that beginning 2 years thereafter microwave systems will become available, in terms of which the cost per unit of service would be cut in half to 75 cents. From then on, the value of the coaxial cable system must be evaluated in terms of the cost of providing equivalent service by the cheapest competing alternative. Accordingly, subtracting the 50 cents of current operation and maintenance costs from the 75 cents value of the service leaves a contribution to capital charges of 25 cents per unit, or $2,500 per year for the remaining 18 years. At the end of the second year, therefore, the value of the cable unit must be set at the discounted value of this annuity, or roughly $24,000. To make the book value agree with this appraisal at that time requires a depreciation charge of roughly $76,000 over the first 2 years, or about $3.80 for each of the 20,000 units of service rendered over this period. Adding the operating and maintenance cost, the total cost of a unit of service during this period is therefore $4.30 rather than the $1.50 that conventional accounting procedures would indicate. If under these circumstances the service is nevertheless offered at a rate of $1.50 per unit, many individuals to whom the service is worth, say, only $2 per unit will demand and get the service, even though they are thereby requiring the commitment of resources involving a net cost of $4.30 per unit. Efficiency at this point requires that the rate be raised either to $4.30, or to the point where demand is kept within the capacity of the plant already installed, whichever rate is lower.

But when the new development is not foreseen long enough ahead of time to allow such writeoffs to be made (e.g., when the price increases needed to cover the indicated writeoffs during the short time remaining before the new technique is effectively available would be so large as to curtail demand below the capacity of the equipment already installed), or when an improper or inadequate depreciation policy has been followed for other reasons, the question arises as to who should pay for the dead horses. While there may be something to be said for making this one of the risks that the investors should be called upon to bear, it is here almost impossibly difficult to define with any precision the circumstances under which and the degree to which a burden of this type would in practice be imposed on investors. Moreover the threat that such a burden would be imposed would create a strong incentive for management, on behalf of investors, to overestimate the amount of obsolescence about to occur, while at the same time underestimating the amount of obsolescence or depreciation deemed to have occurred up to the current period. To attempt to resolve through the regulatory process such an issue, which is even more a matter of opinion than the concept of current value itself, is to give up most if not all of the advantage of the historical cost approach as a relatively well-defined and easily applied method.

In practice, a fairly frequent situation would be one where the inadequate past depreciation was the result of insistence by the regulatory authority on low depreciation rates and correspondingly low current service rates, over the objections of the utility management. Utility stockholders would then have reason to be particularly aggrieved if they were to be made to suffer because of a judgment by the commission that turned out to be demonstrably erroneous in the light of hindsight.

In terms of equity, indeed, it would seem that if, in the period prior to the time the imminence of the innovation became apparent, consumers were benefiting from low rates made possible by low depreciation rates that in the light of hindsight turned out to be inadequate, it is probably fairer that subsequent consumers in the same categories, many of whom will in fact be the same individuals, should pay for the dead horses, rather than that the burden should fall on the shoulders of the investors who derived no benefit from the error. And much the same reasoning can be applied whether the error was produced by developments that could not reasonably have been anticipated with any confidence, or was due to erroneous policy or poor foresight.

To be sure, the saddling of consumers with the burden of paying for benefits accorded their predecessors because of what in the light of hindsight proved to have been mistaken evaluations of technological and other trends and prospects may, in extreme cases, seriously inhibit the full exploitation of new developments that permit a service to be produced at drastically reduced costs, or lead to underutilization of facilities rendered redundant by shifts in demand. But to threaten investors with loss in the event of technological developments beyond those allowed for in past depreciation charges, or to promise them a windfall in case expected development fails to materialize would be to introduce a significant incentive against energetic research and innovation.

In some cases, of course, the impact of the change may come through the competition of alternative services that are beyond the competence of the regulatory authority to control, and in such cases there may be no possibility of maintaining the promised return on investment out of charges for the service. It is primarily with respect to such risks that the investor may be entitled to demand a higher rate of return than that obtained on gild-edged investments. At a slightly less extreme level, moreover, the rate structure that would be necessary to obtain the promised rate of return may be so close to the unconstrained monopoly level as to imply a marginal cost of net revenues that is intolerably high. For example, one might in extreme cases have to impose an added burden on customers of $10 to yield an added return to shareholders of $1. Something of this order seems to have been the case, for example, in many transit operations during the decline of the streetcar and trolley bus. The alternatives to such gross inefficiency under such conditions are either to inflict losses on investors or to bail them out through a subsidy of continuing operations. Unfortunately regulatory commissions have generally lacked the sophistication and foresight to give any specific undertaking as to what their reaction would be to such developments, and of course have generally lacked the power to promise any contingent subsidy, so that the uncertainty of governmental action is added to the basic uncertainty of the inherent economic position of the investment, which uncertainty must of course in the long run be paid for through added risk premiums included in the rate of return.

C. RESPONSIVE PRICING

i. *Needs and Difficulties*

Regulation brings in its train still another source of inefficiency; the difficulty of adapting prices to changing conditions. Marginal cost, indeed, varies from one moment to another or from one occasion to another in ways that are not always predictable far enough in advance to enable the regulatory machinery to operate effectively, at least in terms of present practices and procedures. In competitive markets, changes in the price of perishable produce, for example, tend to take place automatically so as to balance supply and demand and secure reasonably complete and effective use of the entire supply. In decreasing cost industries, where the pressures of competition are not available to induce such price variations, conscious attention must be given to the problem by those responsible for promoting the efficiency of the operations.

While situations vary in the degree to which customers can be informed of and react to changes in rates that would reflect these fine structure changes in marginal cost, there are many cases in which such changes can be reported at relatively low cost to consumers at the time they make their effective decisions, in such a way as to make possible a significant increase in the efficiency of utilization and consequently a lower average level of rates. Telephone service, for example, is technologically particularly well adapted to such rate variation. It would not be difficult, for example, to arrange for the dial tone to be varied, or even replaced by a short recorded message of one or two syllables, to indicate to the prospective caller the current level of rates

in effect for local calls, which in turn would be varied in accordance with the current load on the local switching network. The price variation could be made to reflect the current short-run marginal cost, which in this case is largely measured by the extent to which the making of an additional call by a given subscriber would increase the probability that the attempt to make a local call by some other subscriber would be frustrated by a no-circuit condition. For long-distance calls, the completion of the dialing, or even just of the area code, could be made to yield a further brief recorded message indicating the current rate of charge for the call, at which point the subscriber could either abort the call if the charge is more than he cares to pay, or simply allow the call to go through if the rate is one he is prepared to pay. To some extent, also, the subscriber could then adjust the length of the conversation according to the rate level quoted, whether long distance or local.

The use of such charging would make it less necessary to provide excess capacity to insure adequate levels of service during periods of peak demand: the rate variation would have the effect of leveling off peaks of demand, and, to a lesser extent, of filling in troughs of demand, resulting in higher utilization ratios and lower average rate levels. Such a rate variation scheme would be more significant in flattening peaks than in filling troughs, since much of the latter effect could be achieved fairly readily by regular scheduled rate reductions, though there is danger that if an attempt is made to fill in the troughs too completely by scheduled rate reductions, new peaks may be created that will produce unsatisfactory service conditions. Something of this sort has actually happened with respect to offpeak rates to resort areas during holiday periods.

The variation of rates in response to current traffic conditions is particularly important as a means of preserving adequate telephone service standards in periods of emergency such as a storm or other disaster. At present, and increasingly as the proportion of calls requiring human intervention continues to decline, when lines become overloaded about the only thing that can be done to insure that important calls get through is to disable all but selected classes of telephones from originating calls except through an operator, and perhaps to appeal to customers to use voluntary self-restraint. Even this, however, is likely to result in many emergency calls failing to get through because they originate with one of the disabled telephones, and even if such telephones remain capable of making emergency calls through operators, the dwindling staff of operators may be unable to meet the needs, particularly as additional time may be required to inquire whether the call actually qualifies as an emergency. With responsive rate variation, rates at such times could be made high enough to discourage frivolous calls and it would in most cases be possible for urgent calls to get through promptly without any need for special intervention. Responsive rates could thus result in a substantial improvement in the reliability of the service and a reduction in its average cost.

A government-owned utility would have no particular difficulty in adopting a responsive rate system and obtaining the corresponding benefits. With a regulated private utility, however, any schedule ap-

proved by a regulatory body that would establish a specific relationship between the degree of utilization of capacity at a given time and the rate that is to be chargeable would immediately create an incentive for the utility company to fail to install additional capacity to the extent that would be desirable, since this would increase its costs while simultaneously reducing its rates and probably also its gross revenues. Something would have to be done to counteract this perverse incentive, for if the regulatory body were to attempt to exercise control over the installation of capacity, the duplication of managerial effort, confusion of authority, and conflict of interests that would result would almost certainly lead to inefficiencies that would wipe out any advantage there might be in private rather than government ownership.

ii. *Escrow Funds for Responsive Pricing*

Actually it does seem possible even here to work out an arrangement that would be reasonably satisfactory. One approach would be to have both a responsive rate and a retention rate for each unit of service. The responsive rate would be the rate quoted to the customer, according to which his bill would be computed, and which would vary in some specified fashion according to the current degree of utilization of capacity. The retention rate would be a rate determined in the conventional manner, and would determine the amount of revenue that the utility company would be entitled to retain from the sale of a unit of service; it would be an internal accounting figure that would not concern the customer directly. Any excess of the responsive rate over the retention rate would be paid into an escrow fund, while on the other hand the utility would be entitled to draw on this fund to make good any deficiency of the revenue obtained from the responsive rate below the retention rate. Failure of the utility to expand capacity would cause the escrow fund to grow, while the retainable revenues would be held back by the limits of capacity and the high responsive rates. If the escrow fund were to grow beyond some reasonable size, provision would be made for the excess to escheat in some appropriate manner. If on the other hand the utility were in invest in excess capacity the escrow fund would be dissipated and after its exhaustion the utility would have to content itself with the actual revenues from the responsive rates as a maximum. This may appear to be a somewhat clumsy arrangement, but something of this order would seem to be well worthwhile in view of the greater efficiency to be achieved through responsive rates. If such an arrangement were deemed to be too clumsy, then greater efficiency achievable only through the use of responsive rates would have to be considered an argument for replacing private with public ownership.

Responsive pricing can also be useful in other fields. In electric power, for example, responsive pricing might have entirely averted the power failure that caused a blackout of a major part of Manhattan several summers ago, due to an overload during a heat wave, and would have greatly assisted in the recovery from the great Northeast power blackout of 1967. Here the technological changes required to put such pricing into effect is considerably more cumbersome, including the installation of meters capable of having their rates changed through

the transmission of special signals from a central source, as well as switches to turn certain types of apparatus off in response to those same signals. It seems likely, however, that here too this added expense would be well worthwhile in view of the greater efficiency of utilization secured and the greater reliability of the service, though this is not quite so clearly evident as in the telephone case.

In the case of services sold in advance on a reservation basis, such as long-haul air travel, responsive pricing can be applied through varying the price at which reservations are made in accordance with the ratio of previous sales for a given future date to the normally expected level of such sales. With computerized reservation handling methods, such responsive pricing becomes possible without excessive additional overhead, and for the longer flights, at least, the gains made possible seem impressive. It should be possible to maintain load factors of 90 percent or more, with a corresponding substantial reduction in the average fare. Last minute demands for space on particular flights could nearly always be accommodated, though possibly only at a relatively high fare; on the other hand passengers willing to shift their plans to match the availability of space would be able to travel at fares from one-half to one-quarter or even less of the current fares. And on routes over which substantial competition prevails, the escrow fund mechanism might not even be necessary.

D. EQUITY ISSUES UNDER REGULATION

The regulatory process is of course concerned not only with economic efficiency and with equality as between customers and shareholders, but with equity among different classes of customers. Such a concern sometimes affects the regulatory process in subtle ways. In the ingot molds case, in which a decision was handed down last summer by the Supreme Court, after 5 years of litigation, the Interstate Commerce Commission was upheld in deciding that even though the marginal cost of shipping ingot molds from Pittsburgh to Steelton, Ky., was $4.69 by rail as against around $5.19 per ton by a competing truck-barge route, the railroads concerned would not be permitted to lower their rate to $5.11, equal to the rate charged by the truck-barge route. The ostensible ground for this decision by the ICC, which the Supreme Court was not disposed to overturn on economic grounds, was that the fully allocated cost by rail, including a share in the intramarginal residue, was determined to be $7.59 as against $5.19 for truck-barge, and that the latter mode therefore had the "inherent advantage" and was entitled, according to the statute, to have its carriage of the traffic protected. This determination makes no economic sense in terms of the immediate issue at hand, for on the basis of the data presented the total cost of the transportation job would clearly be minimized by routing the shipment by rail, particularly as it appeared that if the rates were equalized, as the railroads proposed, shippers would generally have preferred the rail route. To an economist, at least, the normal interpretation of the term "inherent advantage" would have been one that gave this result.

However, there may have been a more subtle rationale for this decision lurking, consciously or otherwise, in the minds of those making the decision. The ICC was not at liberty to refuse the shippers the opportunity to ship by truck-barge at the $5.11 rate, as this would have in effect denied to them the advantage of access to the water route, which would have been contrary to the policy involved in the appropriations for navigational improvements. On the other hand to have lowered the rail rate to the level necessary to secure the traffic for the rails, which was the immediately efficient solution, would have involved introducing a locational discrimination between this rate and the rates by the same mode between other points over comparable distances for comparable commodities. While the locational discrimination is in fact there as soon as the low-cost waterway becomes available, whether or not the rail mode is allowed to compete, the discrimination seems to be more tolerable if it is generated by the actual use of a separate mode of shipment with different cost characteristics, especially if this mode appears to be covering its own costs entirely, than if the discrimination is between the rates charged for comparable services by the same mode and perhaps even by the same company, and is brought about by the mere threat of competition from a mode that is not actually used. As long as the actual movement of the ingot molds is by water, Sparrows Point, Md., let us say, is in less of a position to cry "foul" and clamor for a similar rate than it would be if the actual shipment to Steelton, Ky., were to take place by rail, even though the relative competitive position of the two localities would be substantially unaffected by a shift to rail shipment at the same rate, except, possibly to the extent of the minor difference in convenience to the shipper. Thus in this instance, the greater degree of felt injustice on the part of fourth parties when the shipment is by rail than when it is by water is in a sense irrational, in that the real relative position of the parties is substantially the same in either case. Nevertheless such feelings of injustice do generate a political pressure that is probably felt by the ICC in comparable cases, at least, if not in this one. The desire to avoid such pressure, together with the steady barrage of attacks by water transport interests on the discriminatory rate structures of the railroads and the charge somewhat loosely bandied about that the ICC had become the "captive" of the railroads may have generated the psychological atmosphere that led to this otherwise mystifying decision.

But while regulation may thus be prey to forces inducing decisions less related to economic efficiency than to notions of equity, even, in some instances, where the equity issue is a spurious one, it must be conceded that direct government operation is even more susceptible to the action of such forces. Strong demands are likely to be pressed for uniformity of nominal rates even where circumstances would call for substantially different rates on efficiency grounds and where the justification for uniformity on equity grounds is weak. A degree of discrimination that would be acceptable when administered by private enterprise may become intolerable in a government operation, and indeed to permit too wide a latitude for differentiation of price is to invite a more direct appeal to political pressures than is usually possible in the case of a private operation, even when regulated.

E. DECENTRALIZATION AS AN ESCAPE FROM INEFFICIENT UNIFORMITY*

In some instances the pressure for inappropriate uniformity of rates or service standards over a wide area can be averted by decentralization of the operation. In the TVA, for example, while the central agency acts as the wholesale supplier of power and exercises some control over the rates and policies of the local distributing organizations, considerable variation is permitted. One could well give consideration to the application of similar decentralization in other areas. For example the postal service might well be decentralized by turning over to local governments or to agencies financed and controlled locally the responsibility for delivery and collection of the mail, with bulk transportation of mail between local post offices performed by a national agency. This would make it possible, among other things, for an appropriate part of the cost of the delivery and collection service to be paid for through assessments against property owners, as suggested previously.

More importantly, such decentralization would make it possible for each locality to determine for itself what level of service it wants to maintain, in terms of frequency of delivery and collection, and the like, in relation to the corresponding cost, the density of the community, and other factors. The interface problem would be slightly more complex than with the distribution of electric power, as it would be necessary to arrange for the coordination of schedules and to determine who is to do what sorting, but such questions could presumably be solved. Similar considerations might also apply to telephone service, especially where, as in England, it is operated as a branch of the postal service. Decentralization of a sort already existing to a degree, for telephone service in the United States, but in general only down to a State or regional level, and even then a good deal of the effectiveness of this decentralization is lost because of constraints resulting from holding company relationships and the tendency of the pattern of decentralizations to follow political boundaries rather than to conform to functional requirements.

F. FACTORS OFTEN OVERLOOKED IN EVALUATING ECONOMIES OF SCALE

In addition to the many situations where the decline in costs with increasing volume is explicit and easily recognized, there are a number of situations in which the same principles of maximizing efficiency apply but where the economices of scale take forms that are not so obvious, or where the economies of scale tend to be underestimated. It is important that these less-obvious instances not be overlooked, and that the full extent of the existence of economies of scale be properly appreciated.

Thus for example the decreasing costs involved in a rail transit service are usually readily recognized in those cases where the minimum right-of-way provides more capacity than is likely to be needed

*Further discussion of this issue is found in the paper by Haldi in vol. 3 of this collection.

in any but the most exceptional cases. But even where the system is being operated at maximum capacity, as is the case, at least in terms of current operating procedures, with many of the central trunk routes in New York during the rush hours, the outlying routes are generally operated at less than capacity. The economies of scale involved in operating a busline, on the other hand, are not so apparent. The costs of operating a bus system may tend to vary fairly directly in proportion to the number of bus miles operated, at least as long as the relative daily pattern remains unchanged, and this may be thought of as indicating an absence of economies of scale. Nevertheless there may be substantial economies of scale in terms of passenger miles in that the higher density of operation may permit higher average load factors to be achieved. And even where load factors remain the same, there is a strong element of increasing returns resulting from the fact that as the volume of traffic goes up and with it the number of bus miles operated, the increased frequency and possibly also the increase in variety of routes and types of service offered in an improvement in the quality of service. If credit is taken for this increase in the value of the service to the former passengers against the cost of providing the additional bus miles that accommodate the additional traffic, the net cost of providing the service to the new passenger is less than the average cost per passenger of the entire operation.

Again, in transportation or communications generally, increased traffic leads to the establishment of more direct routes. Thus, in the freight transportation case, even though costs per ton mile were to remain constant, or perhaps to decline only very slightly with increased density of traffic along a given route, as more direct routes are established it becomes possible to ship twice the volume of traffic between the various origins and destinations according to the same overall pattern, but with less than twice as many ton-miles. If we attempt to measure the volume of service performed by the number of ton-miles over which the freight is actually carried, the economies of scale in accomplishing the carriage of freight from point to point may be significantly underestimated. For example, suppose that points A, B, and C are located at the vertices of an equilateral triangle 100 miles on a side, and that each point ships 10 tons of traffic per day to each of the other two. If at this level of traffic it is not found worthwhile to have a direct link between A and C in addition to the AB and BC links, total traffic will be 8,000 ton-miles per day. If then activity doubles and each point begins to ship 20 tons per day to each of the other two, and if now with this increase in traffic it becomes worthwhile to complete the link AC, then total ton-miles will increase to only 12,000 per day, an increase of only 50 percent resulting from a 100 percent increase in demand.

On another level, statistical studies that attempt to estimate marginal cost and the economies of scale by comparing the cost of operations in different areas involving different densities of traffic or intensities of operation often fail to isolate the element in the situation that

is of importance for efficient price-making purposes. For example, it is generally found that the cost per telephone connected is greater for large cities and metropolitan areas than for smaller towns and cities, from which it is sometimes concluded that local telephone service operates under conditions of increasing rather than decreasing cost. This statistical conclusion is sometimes buttressed by arguments based on the increase in the complexity of the switching equipment and in the number of stages of switching that each call must go through as the number of telephones out of which the desired one is to be selected increases.

But such an analysis leaves out of account the longer distances involved in the typical call in the larger city, as well as the fact that the telephone plant must be to a larger extent placed underground, with an increasing degree of difficulty of threading through a maze of other utilities as the city grows larger. The relevant question is not whether costs increase as the city gets larger, but whether average costs would go up or down if there were a significant increase in the number of telephones connected, possibly in response to a rate reduction, within a given area and with a given physical environment to contend with. Considered in this way it seems clear that the substantial reduction in cost per circuit-mile as more circuits are placed on a given pole or in a given set of ductways in a trench along a given route will more than outweigh whatever diseconomies of scale there may be in having to provide a more complex switching apparatus in the central exchanges.

Somewhat similarly, studies have been made purporting to determine the economies of scale in the railroad industry by comparing the costs of large railroads with those of small ones, without considering that it is not the size of the railroad as a whole, however, measured, that is the relevant variable, but the density of the traffic. Thus the Richmond, Fredericksburg & Potomac Railroad, with a relatively small total volume of operations in terms of ton-miles, can be expected to have a relatively low level of average costs because of the high density of traffic over its relatively short route, whereas the Southern Railway, though a much larger railroad, however measured, might not be able to achieve comparably low costs because its traffic is spread out over a much larger network of route miles. Comparisons which deal with entire railroads, accordingly, are unlikely to generate a relevant estimate of marginal cost or of the extent of economies of scale.

Economies of scale may thus be much more significant, in the sense here relevant, than might at first appear from a cursory study of the statistics. It is important that one not be misled into thinking that the problem is relatively unimportant, or that the efficiency problems are minor compared to the one of securing equity. Indeed, it is hardly too much to say that to date most of the practical regulatory and other pricing policies associated with decreasing cost industries have been much too concerned with questions of equity that often turn out to be of little consequence in terms of the way individuals are affected in the long run, and that the substantial possibilities for increased ef-

ficiency, lower overall levels of charges, and improved economic welfare that inhere in imaginatively conceived pricing policies have been largely disregarded. A thoroughgoing and imaginative attack on the problem of pricing the output of decreasing cost industries and activities for maximum economic efficiency along the lines suggested above is capable of yielding very substantial returns in increased output, better service, and a higher overall level of real national income.

UNCERTAINTY AND THE NEED FOR COLLECTIVE ACTION*

BY RICHARD ZECKHAUSER

Richard Zeckhauser is Assistant Professor of Economics at Harvard University.

Individuals frequently find it advantageous to undertake actions collectively when market breakdowns prevent the free market from leading to an efficient outcome. Usually such action is undertaken by a governmental unit. Here Professor Zeckhauser discusses how the presence of uncertainty, a lack of knowledge of which state of nature will obtain in the future, may hinder the working of the free market mechanism and create situations in which collective action is called for.

Individuals' future preferences for some classes of goods cannot be known with certainty. For example, an individual cannot know his future demand for firefighting services or various types of medical treatment. The efficient provision of such goods and services must take into account risk-spreading considerations and utilities in anticipation of use. The particular institutions available for providing these goods, perhaps taxes or user charges, may affect the decision whether the goods should be provided, and if so whether by a governmental unit. Insurance and markets for contingent claims have many related properties. They too might best be provided on a collective basis.

The collective provision of goods raises questions of equity if some individuals benefit at the expense of others. One can argue that if all would have agreed to a plan before the outcome became known, it would be equitable to undertake the plan after the fact. Government efforts such as aid to depressed areas and to the economically disadvantaged can be supported on this basis. Any such redistributional assistance schemes raise the incentives problem. Here it is examined with respect to an optimal income tax scheme.

In this paper, Professor Zeckhauser attempts to show that the presence of uncertainty creates special obligations and possibilities for collective or governmental action.

Introduction

In a perfectly competitive world it is well known that the free working of the market mechanism will lead to an efficient outcome, a Pareto optimum.[1] The neoclassical model of perfect competition requires perfect knowledge; there can be no uncertainty. The presence of uncertainty, a lack of knowledge of which state of nature will obtain in the future, makes the model inoperative. Ostensibly some corrective action will be required if an efficient outcome is still to be achieved. As with other forms of market breakdown such as externalities, public goods and declining marginal cost industries, some form of collective action might be an appropriate way to seek a remedy.

[1] A Pareto optimum is a situation in which it is impossible to increase the welfare of any individual without decreasing the welfare of some other individual. A situation that is not Pareto optimal is inefficient in the sense that we could find another situation (one which might be called Pareto superior to the first) which was better for at least one individual and worse for none.

*Robert Edelstein and Professors Robert Haveman, Howard Raiffa and Thomas Schelling made helpful comments on this paper.

Contingent Claims Markets

A number of recent papers have shown that the presence of uncertainty need not prevent the free-market mechanism from achieving an efficient outcome. It may be possible to expand the class of traded goods to include what are sometimes called contingent claims. A contingent claim is a right to a variable amount of a commodity, the amount to depend upon the state of nature which obtains. In real life, quite frequently, a positive amount of the good is received only in the event of a single state. An insurance policy is a good example of a contingent claim. Typically it pays the face value of the policy should the state "death of the insured" prevail. Similarly, a lottery ticket is a contingent claim. It might give one the right to $10,000 should the number 904521 be drawn.

Contingent claims markets need not lead to efficiency.* Transactions costs on these markets are far from negligible. The result is that the market price of a contingent claim may far exceed its actuarial value. Even if these markets were actuarially fair, other factors might prevent the achievement of a Pareto optimum. If there are substantial disagreements on probabilities or if probabilities are difficult to estimate, the free-market mechanism will not in general lead to an efficient outcome. However, the primary difficulty is not that existing contingent claims markets do not function effectively, but rather that these markets are established in but few areas. Many goods are difficult to trade on a contingent basis for reasons that will be elaborated later. In some instances it is impossible to make trades before the state of nature is known or at least a great deal of information has become available. This limits the risk-sharing advantages of trades of contingent claims. Once contingent claims trades have been made, incentive structures change, and the trading individuals may take actions that affect the probabilities that different states occur. For example, the individual who has insured his car may drive somewhat less carefully. On a realistic basis we cannot expect contingent claims markets to achieve optimality in a world with uncertainty. This paper examines the role of collective action, with the Government as the likely agent, in overcoming the inefficiencies that uncertainty engenders.

Varieties of Uncertainty**

Most of us are accustomed to look at uncertainties from the standpoint of the individual. Will there be good times? Will I prosper? Will a hurricane destroy my home? Will I be disabled in an accident? To the individual these questions may appear to be structurally similar. To the Government they may appear very different. If I am lucky in the first instance then so is everyone else. A hurricane will affect others, but hardly the whole country. An accident on the other hand is likely to involve just me or perhaps a few others. The Government is concerned with all citizens and would likely see these uncertainties in very different ways. The actions that it might or should take with respect to them might vary as well.

---

*Further discussion of this issue is found in the paper by Demsetz in this volume.

**Further discussion of this issue is found in the paper by Hirshleifer & Shapiro in this volume.

Collective Provision With Individual's Preferences Uncertain [2]

An individual will rarely be able to determine exactly his future level of demand for a good. This demand will depend upon the evolution of his preferences something that he may not be able to predict accurately. This need not reflect an inability to know his own future mind. Rather, events over which the individual has little control may significantly affect his future preferences. An individual would have to know whether his house would catch fire to predict his demand for firefighting services. He would have to know whether he would be involved in an accident before he could estimate his need for hospital accident emergency room.

In normal circumstances, the inability of individuals to predict their future demands for goods is not of great consequence. A good is supplied to the market, and is consumed according to the individuals' preferences at that time. Only if action must be taken at present to insure an efficient level of provision for the good is it important to make accurate predictions. Thus, for example, if there is a significant lead-time for production, it will be most useful to be able to predict aggregate demand with some accuracy.

This section considers situations in which individuals who are unable to predict their individual portions of future total demand may find it useful to make some decisions on the collective provision of a good before preferences become known. For the most part it is concerned with goods which by their very nature make accurate prediction of future preferences impossible. At best, for such goods, it will be possible to attach some probability assessment to possible sets of preferences for the individual. The acronym PIP derived from probabilistic individual preferences is used to refer to the goods of this nature. Most frequently, the probabilities attached to these sets of preferences will be determined subjectively rather than objectively. For the present purposes, this distinction is not of importance.

Option Values*

Consider a community that must decide whether to supply itself with a PIP good, perhaps a supply of plasma at a blood bank, or a fire engine to protect its homes. Following conventional efficiency dictates, the community should estimate future levels of demand for the good and decide whether perfect price discrimination against future consumers would yield sufficient returns to enable discounted revenues to cover discounted expected costs. If the decision is negative, we are told that the community should not provide itself with the good. This stipulation holds whether or not consumption by one individual reduces the amount available to be consumed by another, whether or not the good has any aspects of a collective-consumption good.

Many PIP goods serve a preventive purpose; for example, programs to deter crime or to immunize against disease. If such goods

[2] Part of this section is to appear in the May 1969, *American Economic Review*. It is taken from my essay collection, *Studies in Interdependence*.

*Further discussion of this issue is found in the paper by Steiner in this volume.

are provided, it may never be known who are the major beneficiaries, who would have been the victims of the crime or the sufferers from the disease. For such goods, discriminatory pricing questions could only be asked on a hypothetical basis. If you have been identified as a future target of a mugging or a future victim of polio, how much would you pay to prevent the unfortunate happening from taking place?

In an article dealing with commodities "the purchase of which is infrequent and uncertain," commodities which are in effect PIP goods, Burton A. Weisbrod argued that even if there will be perfect price discrimination, the option values of potential consumers would influence the decision whether such commodities should be provided. He states, "If these consumers behave as 'economic men,' they will be willing to pay something for the option to consume the commodity in the future." [3]

The key question for Weisbrod, then, is, What will be the magnitude of these option values? The amount that an "economic man" will pay for the right to consume a commodity in the future will depend upon the price that he will be charged for the commodity. The discriminatory price of a good for an individual is the maximum price that he would pay for the good, an amount just short of the one at which he would refuse to purchase. As a consequence, if an economic man will be faced with perfect price discrimination, should he be interested in consuming the good, he will be indifferent whether or not he purchases it. He will not be able to reap any consumer's surplus at the time of purchase; the availability of the good will in no way be able to increase his welfare. His option value to keep the purchase available will be zero. I disagree with Weisbrod and assert that in this regard the conventional efficiency dictate is not called into question.

RISK SPREADING

The option value, discriminatory price approach to the allocation decision problem is not sufficiently rich. It overlooks some important considerations, one of which is the risk-spreading property of PIP goods. Every potential consumer of a PIP good is in an uncertain situation. He does not know whether he will wish to consume the good in the future. The value in use of many of these goods can be very high.[4] If the actual consumer is to be faced with perfect price discrimination, then each potential consumer is in the position of having some probability of incurring a substantial charge.

However, it may not be necessary to charge the actual consumers even a small fraction of their discriminable price. The community of potential consumers as a whole can agree to share the costs of making the good available, with little or no additional cost to be borne by the actual consumers. In this way, each individual will bear a small certain cost rather than run the risk of a much larger, but considerably less

[3] "Collective-Consumption Services of Individual Consumption Goods." *The Quarterly Journal of Economics*, LXXVIII (August 1964), p. 472.

[4] This is particularly true of goods in this class which are primarily used to avoid or ameliorate unfortunate occurrences. Many goods with such a purpose fall into the class of PIP goods.

The expected time until a highly unlikely loss is incurred is great. All losses or expenses in this essay are assumed to be discounted. With uncertain occurrence time, the discounting process is complicated.

probable cost. If the individuals in the community are risk averters, this indirect form of risk sharing may be of value to them.

The possibility of risk sharing introduces a new consideration to the question as to whether the good should be provided. Let me illustrate with the fire engine example. Consider a community of 100 identical individuals, with identical utility functions for wealth, log (w), each with a house they value at \$10,000, and each with a nonhouse wealth of \$10,000. Every year one house, no more, no less, burns down. The community must decide whether to get a super fire engine that stops all fires instantaneously. The rental on the fire engine is \$12,000 per year in comparison to the \$10,000 that the individual whose house starts burning would pay for the services of the engine. (If the income effect for housing were positive, this would somewhat reduce the \$10,000 figure, and conversely.)[5] Following conventional efficiency dictates, the community would compare the amount that could be secured for the good through perfect price discrimination with the cost of provision. It would decide not to rent the engine.

However, the individuals within the community are risk averters. They would be willing to pay a premium to reduce the variance in their payoffs. If they share the costs of providing the engine, they can accomplish this. Let x represent the certainty equivalent dollar value for the lottery that gives the individual's payoff in the absence of the fire engine. We have, $\log(x) = .99 \cdot \log(20{,}000) + .01 \cdot \log(10{,}000)$. This gives, $x = 19{,}861.85$, and $20{,}000 - x = 138.15$. To avoid one chance in 100 of a \$10,000 loss, each individual in the community would be willing to pay \$138.15. This amount can be looked at as the individual's option value for the right to consume at zero cost the fire engine's services at any time during a 1-year period. The sum of the individuals' option values would be \$13,815, an amount well in excess of the rental price.

What, then, of the decision not to rent the engine? Each individual would rather pay his share of the rental rather than be faced by the lottery that confronts him if the fire engine is not rented. If the only alternative to renting the engine is to let the houses burn and have individuals suffer the losses, then the engine should be rented, and the rental cost shared equally. This will involve a loss in efficiency with regard to resource use, but there will be a more than compensating gain because, unlike the costs of a fire, the costs of fire control can be spread among all the members of the community.

But another, more attractive alternative may be available. It may be possible to spread the cost of the fires through an insurance pool in the community. If each individual agreed to pay one-hundredth of all fire losses in the year, his annual cost would be \$100. This would give him a considerable saving over paying his share of the engine rental cost. If there are no institutional constraints to prevent the operation of an insurance plan, then so long as administrative costs

[5] Income-type effects can be important if PIP goods are of exceptional value in use. For example, condition X has a 0.8 chance of being fatal. The discriminable amount an X-sufferer would pay for treatment B which would reduce this probability to 0 might not be much greater than that for treatment A (assuming B unavailable) which would merely reduce it to 0.4. If there were one unknown sufferer in a large community the result would be much different. The total discriminable payments from all members to make treatment A available to the unfortunate member would be approximately twice that for treatment B.
Where specific goods, such as houses, make up the potential losses to be prevented or insured, the amount that would be paid at the higher, before-loss income is the relevant amount for efficiency calculations.

are not prohibitive (below $2,000 in this example), this alternative will be the most attractive.

THE COMMUNITY INSURANCE DECISION WHEN THE SIZE OF LOSS IS VARIABLE

The assumption that a single house will burn each year is certainly unrealistic. There may be major conflagrations, and even if house burnings are independent events, there will be some variance in the number of houses burned each year. The fire engine has the advantage that it eliminates all such variance in total cost to the community. This improves its attractiveness relative to an insurance scheme. If there is significant variance in the number of houses that burn each year, the community might choose the fire engine over the community-wide insurance plan even though the latter would have a lower expected cost per individual.

The limiting case of complete contagion would have the only possible fires in the community destroy all houses, the annual probability of such a catastrophe being one in 100. In this case the insurance plan would serve no risk-spreading purpose. The lottery for each individual's payoff would be the same as the one in the absence of insurance. Following the analysis above, the fire engine should be rented.

Assume that there were no fire engines. We observed that a risk-spreading plan in a community will be the more effective the less positive correlation there is between significant losses for its individual members. This observation may be of import if each individual will be charged a proportional premium above his share of community losses if an insurance plan is instituted, if, for example, there are administrative costs that are some fixed percentage of total reimbursed losses. The individuals might find it desirable to establish an insurance plan that would reimburse losses if one or two houses were lost in a year, but not if a great many had burned.

By way of illustration, assume that there are but four possibilities. The number of houses that burn in the community with the probabilities of occurence are: 0, .75; 1, .15; 10, .07; 50, .003. The expected number of houses burned is one. Our community would find it desirable to reimburse losses from single-house fires if the premium over fair actuarial value were not more than 38.1 percent. It would be worth while to insure losses in 10 house fires as well if the premium were less than 33.9 percent. Only if the premium fell below 17.2 percent would the community wish to include losses from 50 house fires in its insurance plan.

UTILITIES OF ANTICIPATION VERSUS UTILITIES OF USE

There are some special characteristics of goods the future consumers of which are unknown that render the actuarial model of decision inadequate. The lottery model is essentially static. The decisionmaker is assumed to act as if the probabilistic outcome is to be determined immediately when his decision is made. Pleasure or pain derived from anticipation or anxiety over the outcome of the lottery does not enter into the model. But these feelings are not imaginary. We cannot afford to neglect these real and potentially significant utilities of sensation. They must be considered in addition to the familiar (what might be called) utilities of use.

Many individuals would gladly incur significant costs so that they could live in a community with a museum or symphony even though they are unlikely to make use of these cultural advantages. A shopper might pay to have a wide selection of goods available even though she knows she will probably choose her purchases from among a very few. We have in effect what might be called a utility of variety. Consumers like and get a positive utility from the fact that goods are kept available. From what we can observe, the amount that these individuals would pay to keep these facilities available is well in excess of the expected value to them, the probability they use them times the difference between the price they will be charged and the discriminatory price that they could be charged. For most individuals the discriminatory price would be low enough so that risk considerations do not play a major role. It would seem that another factor is present. It is what we might think of as the utility of knowing you can use something even though it is unlikely that you will use it. It is a utility that is defined independently of the use itself. In this way it is quite different different from the actuarial aspects of payments for PIP goods (or those that Weisbrod discusses). This distinction is reinforced to the extent that payments for utilities of variety are independent of the prices that will be charged for the goods or the probabilities that they will be consumed.

MISESTIMATION OF PROBABILITIES

There is the possibility that some individuals may misestimate the probability that they will wish to consume some good or service in the future. If there were a consistent bias toward overestimation, individuals would be willing to pay more than fair actuarial value to keep the good available. The converse is of course true.

This raises what might be more of a philosophical than an economic question. If individuals calculate their option values on an actuarial basis using expected consumers' surplus, and if they overestimate probabilities of use, should we correct for this bias? Let us say that the head of a community asks each individual how much he would pay to keep a museum available. Assume that the community head knows from experience that on the average individuals will overestimate the probability of future visits to the museum and will thus arrive at somewhat inflated option values. Has he done something useful if he deflates these values when he decides whether to provide the museum?

A number of considerations must be weighed in arriving at our answer. If all utilities are in use rather than anticipation, we would argue that the community head should modify the stated option values to reflect accurate probability estimates. If utility in anticipation is a meaningful concept, the question is more difficult and will depend on the specific parameters of specific cases.

Will individuals learn that their probability estimates were too high, that they made a mistake? We would surely agree that the more likely they are to find out, the more desirable it is to correct their misestimates. It is by no means clear by what means an individual would discover that his subjective probability estimate for an unlikely, nonreplicated event was in error. The less danger there is of becoming informed, the more blissful is ignorance. Taking the argument to the

extreme, we might even argue that we should provide a little-used museum if we can convince community members that they are fairly likely to visit it. If the pleasures in anticipation are sufficient, the provision of the museum could yield positive utility to each individual who shares in the cost of making it available. This will result in a real, though unanimously favored, misallocation of resources.

This is tricky material. This brief discussion is hardly adequate. The primary present purpose is to point out that in some instances there are benefits to be gained from fleshing out the conventional skeletal framework for expected utility calculations.

PIP Goods, Insurance, and Markets for Contingent Claims

Prospective and Retrospective Benefits

In our discussion thus far we have found a number of relationships and similarities between insurance mechanisms and PIP goods. An insurance policy gives one claim to future resources should some prespecified event occur. One takes out the policy because it provides an expected positive benefit. It cannot be known in advance whether one will actually benefit from having the insurance policy. Generally, in retrospect, one will not prove to have benefited. The events that we insure against usually have a low probability of occurrence. When one shares in the provision of a PIP goods so as to be included as a potential consumer, one is in a similar position. There may only be a small chance of benefiting from the consumption of the good, but the benefit may be sufficient to make the premium (sharing in its provision) worth while.

Quite frequently, PIP goods are provided by a political unit of a nonvoluntary contribution basis. In insurance, the equivalent arrangement is some form of mandatory policy. The social security system is a good example of a nonvoluntary contribution, risk-sharing arrangement implemented by a political unit. At present, there are discussions in progress within the Federal Government on a mandatory flood insurance programs for those holding property on the flood plain. This type of provision, whether for PIP goods or insurance, can be justified on efficiency grounds only if the gains of some individuals, as measured by the excess of expected consumer's surplus over share of provision cost, more than outweighs the losses of these who would not join voluntarily.

An insurance policy may provide a nonactuarial benefit that in some sense is parallel to the utility of variety that is associated with PIP goods. It is a utility that is not directly connected with the use of the goods or the money that is received should the insured against event occur. It is the pleasurable sensation that is derived from knowing that one has protected one's self (or one's family) in case an unfortunate contingency should arise.

Thomas Schelling kindly suggested the following demonstration of the value of this "peace of mind," what we might call a utility of security. A man with a medical ailment is fully aware of the seriousness of his condition, but does not know whether it will exclude him from life insurance. He asks his wife to find out and to pay the premium if his life can be insured. We would likely agree that she may be doing something useful if she can deceive her husband should he be uninsurable, and convince him that his life can be insured.

When one takes out an insurance policy, one sacrifices current assets in exchange for a contingent claim to be paid, if at all, in the future. An insurance policy represents a trade in a contingent claims market, that is restricted in two ways. (1) In an open contingent claims market, future as well as present payments can be used to purchase contingent claims. This restriction will not be of consequence if there are perfect capital markets and if borrowing and lending rates are identical. (2) In an unrestricted contingent claims market, one can exchange a contingent as well as a certain claim for a contingent claim.

If one wished to take out medical disability insurance, rather than pay some fixed amount, it might be far preferable to agree to pay an appropriate fixed percentage of future income that would result in the same expected payment. Carrying this reasoning one step further, it would likely be even better if one could work out an arrangement under which the percentage was higher the higher was future income.

An arrangement like this one might be worked out if the Federal Government decides to provide financial assistance to individuals seeking higher education. To fund this program the Government could charge these individuals a percentage of future income, or better yet, a variable percentage that increased with income. This repayment could be transmitted easily along with normal income tax collections.

There may be situations in which contingent claims markets or insurance are not available but where good substitutes can be found. A farmer who cannot find a future market in the particular commodity which he produces can in effect insure his price if he can sell short a related or derivative commodity which will have parallel price movements. In financial markets it is often possible to secure a form of insurance thruogh arbitrage-type operations; for example, one can sell short in a stock for which one holds convertible bonds. Some situations call for more inventive action. An individual who cannot buy storm insurance on his home in a hurricane belt might cover himself by investing in a roof repair company. In all these situations, the object is to diversify. The extensive literature on the subject of diversification is of consequence because perfect insurance and contingent claims markets do not exist. To the extent that balanced portfolios can be secured, a suitable substitute for these markets can be found.

If contingent claims markets are restricted and if opportunities for complete diversification are not available, there will be distortionary effects on many economic variables. This is particularly true where intertemporal allocations are concerned.

Uncertainty and Intertemporal Resource Transfers

The transfer of resources from present to future, what is commonly called savings, is both an intertemporal shift and a shift from a period in which income is known to one in which income is an uncertain variable. The greater is uncertainty about future income, expected income constant, the more a risk-averse individual will wish to transfer to the future. With fair and perfect contingent claims markets, uncertain incomes will not be of consequence. The risk-averting individual will be able to insure that he receives the expected value of his

future income.[6] There will be no additional transfers in response to risk considerations. The absence of complete, perfect, and actuarially fair contingent claims markets makes transfers to the future greater than they would be in a perfect world with all desirable markets established. Considering this factor alone, the level of savings in society will be supraoptimal.

The question as to whether the freely arrived at rate of savings will be socially optimal has recently received much attention in the theoretical literature. Other essays in this collection discuss it at length in relation to the appropriate rate of discount for social projects and to the responsibility of the Government to provide for the future.* To the extent that they neglect the influence of uncertainty in our imperfect world of incomplete markets, their answers for discount rates should be raised and those for provision for the future diminished.[7]

Collective Provision of PIP Goods, Insurance and Contingent Claims Markets

PIP GOODS

Throughout this paper, I have discussed, with good reason, community or collective provision of these goods or trading arrangements. There are at least three factors that might make it best to provide PIP goods on a collective basis. (1) If the probability is small that any one individual will consume a good in a given short period of time, and if it must be kept continually available, there will be considerable waste if each individual provides for himself. We find it efficient to store plasma in a common pool in blood banks rather than have each individual keep a personal supply. (2) PIP goods possess aspects of collective-consumption goods and for the conventional reasons are best provided on a collective basis. For example, the consumption of the services of a fire engine by one individual in no way reduces the amount that is available to be consumed by another.[8] An essential collective-consumption element of some PIP goods is that it is desirable to keep an inventory available. Thus, in our plasma example, the good itself is of the private consumption variety, but the existence of a stockpile upon which any member can draw has collective-consumption aspects, consumption of the "good" by one individual does not reduce the amount available to be consumed by another. (3) If PIP goods are provided on a collective basis, the costs of provision can be shared, and the as yet unknown consumers need not be charged any great amounts. In this way some useful risk-spreading can be accomplished.

EQUITY CONSIDERATIONS**

On an equity basis, we might object to the collective provision of collective-consumption goods that are not enjoyed equally by all contributors (assuming equal incomes and contributions to provision).

[6] Contingent claims markets will not be actuarially fair unless some individuals are risk neutral, or if there is zero correlation among individuals' incomes.

[7] This section greatly simplifies a complex problem. In my *Studies in Interdependence* I consider the effects of restricted contingent claims markets at greater length.

[8] In the unlikely event that the demand periods for the services overlap, there will be the congestion problem that is frequently associated with "nearly-collective goods."

*Further discussion of this issue is found in the paper by Baumol in this volume.

**Further discussion of this issue is found in the paper by Weisbrod in this volume.

It might seem that such an objection would apply with a vengeance to PIP goods, but such is not the case. Collective provision of PIP goods can only be called inequitable if on a prospective basis it is clear that some community members will have a higher expected benefit. Even then, there might be disagreement.

Final judgment may depend whether higher probabilities of use or higher values in use are the primary explanation of higher expected benefits for some individuals. If the former, it might be relevant to know why the probabilities are higher, whether they depended upon conscious choices. On an equity basis, we would be more likely to object to a subsidized public facility to treat those injured skiing than to a subsidized public facility to treat those who suffer from a particular inherited illness, even though we may be better able to predict who will be the beneficiaries of the latter.

The lessons of equity considerations in situations of uncertainty have not been thoroughly enumerated. These lessons may be applicable to the most basic of philosophical considerations. If we state that society should establish a mechanism that insures individuals against unfavorable outcomes in the lottery of genetic inheritance, against the possibility that they will have low capabilities and little opportunity to earn in a competitive factor market, is our statement based on considerations of equity or on considerations of efficiency? Is it not possible that we would have all agreed in advance to such a scheme, that the insurance mechanism gets us to the ex ante contract curve? These questions and others that relate in similar fashion to such matters as the desirability of nonfault compensation of accident victims, the justness of racial differences in income, or the obligation of the developed countries to aid the underdeveloped world merit further consideration. They are touched upon in the section that follows.

INSURANCE AND CONTINGENT CLAIMS MARKETS

For reasons similar to those relevant for PIP goods, it may be desirable to provide contingent claims markets, and their somewhat restricted offshoots, insurance markets, on a collective basis. To provide such a market is to provide a mechanism that has many of the characteristics of a collective-consumption good. The use of the mechanism by one individual not only does not reduce the amount available to another, it is likely to increase the utility in use of the mechanism for them. Each additional consumer (trader) extends the market and makes it more perfect, thus providing a benefit for all other consumers.

We would suspect that in the absence of some form of collective provision, markets for trades in contingent goods will not be established in many desirable areas. Those markets that are established we would fear to be insufficiently extensive. Unlike most markets, it is essential to have many parties to single trades. The marginal costs of accommodating addition users are likely to be low in comparison to the fixed costs of starting and maintaining the market. Efficiency considerations would dictate that transactions costs be kept low, and that the market be supported by charges assessed independently of use or the frequency of use. It is hard to see how a private entrepreneur would go about setting up a market with such features.

There are, of course, privately supplied mechanisms that approximate mechanisms for contingent goods. The best example is insurance companies. To make ends meet and then some, insurance companies find it necessary to charge amounts considerably over fair actuarial value. We would speculate that the excess charge over actuarial value is well above the nonactuarial costs of including another individual in the insurance pool. This is inefficient. It is equivalent to charging more than marginal cost to a man who uses a public park. I have not been able to find any abundance of privately provided mechanisms that would call into question my contentions relating to the provision of markets for contingent goods.

I have argued above that the Government may see fit to provide PIP goods, particularly in those instances in which adequate insurance mechanisms are not available. At times the Government may participate directly in the provision of insurance protection. At present governments engage in many programs of this nature, including crop insurance, unemployment and disability insurance, and social security. Recently it has been proposed that the Federal Government provide insurance to high-risk ghetto businesses, and to those who might suffer from certain types of natural disasters.

The Government need not limit itself to conventional insurances services. It may identify areas in which individuals would have purchased insurance if only mechanisms had been available. The question is what to do if the would-have-been-insured-against misfortunes have already occurred.

Governments generally step in to aid disaster areas. This form of relief can be justified on efficiency grounds quite apart from humanitarian or altruistic considerations. Disasters take tremendous tolls of public and private property. Public property is rarely protected by insurance. Disasters are highly unlikely events. As such, transactions costs to provide insurance would be quite significant in comparison to actuarial costs. It may be efficient for the Government to operate as if an implicit arrangement had been reached that all will pitch in to help after an unfortunte occurrence, rather than collect premiums to be paid to the losers in case of disaster. Without any explicit agreement, the Government can serve as an insurer. It collects premiums in its general tax funds and pays out claims as needed.

Aid to depressed areas can be viewed in a similar light. In the 1920's all areas of the country, including then coal-rich Appalachia, might have agreed to an assistance scheme to those areas of the country that turned out to be depressed in the sixties. The Federal Government might carry out such assistance now on the efficiency argument that on an expected value basis this arrangement would have been desirable if only it could have been formulated in advance.

A similar argument can be made for assistance to people who are handicapped to some extent, perhaps because of inadequate intelli-

gence or physical disability. Is it not the case that if none of us knew our future position in society that we would have agreed to a redistribution scheme from the more to the less fortunate? Contracts of this nature can never be made. By the time parties reach the age at which they could consent the die has already been cast. Ivy League men know that they are the blessed. Ghetto dropouts are aware of the bleakness of their future. However, as unidentified fetuses not knowing which they would be, these two groups might have agreed to a massive redistribution scheme. Arguments in favor of radically progressive tax schemes can be made on this basis as well. If only the agreements were drawn before the state of nature was revealed or at least substantially hinted at, all would have agreed to a substantial redistribution of income.

The principle here is a most simple one. If at some time during the play of a game it becomes clear that all players would have agreed to a particular rule in advance, then invoke that rule. This dictate is particularly important in the situations here, because, for the most part, these rules cannot be agreed upon ahead of time.

Trades in Contingent Goods and Incentives—An Income Tax Example

Unfortunately, trades in contingent goods that have a risk-sharing objective may reduce incentives for individual effort. Consider a completely levelling income tax scheme. In a community of n individuals, an additional dollar earned by any individual will yield him but $1/n$ of a dollar in extra income. If n is large, there will be no financial incentive to work.

The incentives problem can be examined with an income tax example. A group of individuals wants to develop an optimal tax scheme. They are able to contract before any of them know their earning capabilities. The possible wages are 1, 2, 3, and 4 dollars per hour. One quarter of the individuals earn each wage rate. The probability, p_i, that an individuals receives wage rate w_i is thus ¼ for all i.

Each of the individuals has a utility function in each 100-hour period. As agruments it has two variables: after-tax income, z; and leisure hours, x. The utility function, assumed to be common to all is:

$$U(z, x) = z^{\frac{1}{2}}x^{\frac{1}{2}}.$$

Each individual is allowed to follow his individual preferences to decide how many hours, h, he wishes to work leaving him with $100-h$ hours of leisure. It is this area of free decision that gives rise to incentives considerations. An individual's before tax income, y, equals his wage rate times the number of hours he works,

$$y = wh.$$

The individuals contract on a tax scheme $T(y)$ that tells the amount of taxes an individual must pay on a given before-tax income. After-tax income for an individual whose wage is w and works h hours is:

$$z = wh - T(wh).$$

The tax scheme that maximizes expected utility is the one that maximizes:

$$\sum_i p_i U(w_i h_i - T(w_i h_i),\ 100 - h_i),$$

subject to the constraint that:

$$\sum_i p_i T(w_i h_i) = 0.$$

To simplify the search task, I considered only quadratic tax schemes. These schemes are sufficiently diverse in shape to allow a close approximation to the optimal scheme of general form.

The optimal quadratic tax scheme, the one that maximizes expected utility, is:

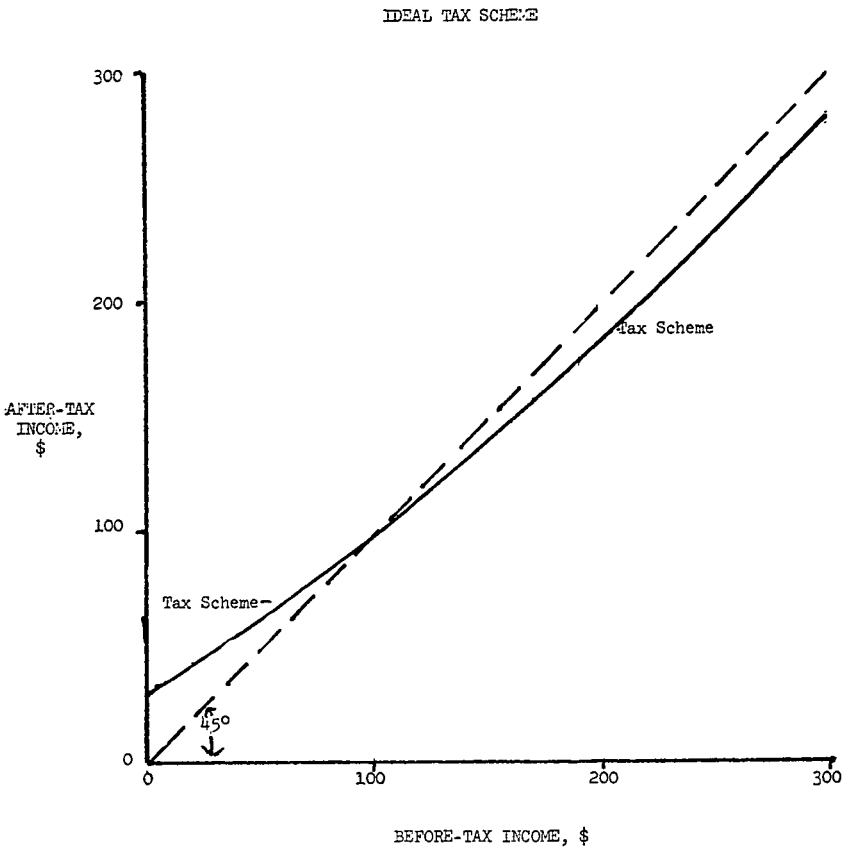
$$T(y) = -.0007y^2 + .372y - 30.125.$$

The shape of this tax function is rather interesting. The marginal tax rate continually falls over the relevant range although the average tax rate is rising. This tax scheme is progressive on an average though not on a marginal basis. At first glance this may seem a bit perplexing, but the explanation is quite evident. The large negative constant in the tax scheme represents a lump sum payment that each individual receives. This provides the progressivity. The decreasing marginal rate insures that high income people will maintain the incentive to work. The table below gives the outcome with the optimal scheme.

VALUES FOR THE OPTIMAL SCHEME

Wage rate	1	2	3	4
Leisure	72.178	58.500	53.316	50.347
Before-tax income	27.822	83.000	140.052	198.614
After-tax income	48.139	87.072	131.809	182.468
Average tax rate	−.730	−.049	.059	.081
Marginal tax rate	.343	.256	.176	.094

The diagram below illustrates the ideal tax scheme.



This result is only suggestive of the type of outcome that can be achieved with this type of analysis. It is interesting however, that with no special consideration of the problems of the poor or the lowly skilled, it produces results remarkably like those of the negative income tax at the bottom end of the income scale.[9]

The Government's Attitude Toward Risk*

In many instances a government will be in a position to insure individual's risks with very little risk to itself. This will be the case, for example, with disability insurance. The number of people disabled is a predictable figure with fairly little variance. However, in many

[9] The mathematics of this example has been omitted here. Other simulations have been run for other utility functions and distribution of earning opportunities without achieving markedly different results. A more detailed paper is available on request.
This same type of analysis can be applied to other related problems, such as the optimal plan for car insurance. The optimal result usually involves some form of limited risk sharing; it involves a tradeoff between risk sharing and the maintaining of incentives.

*Further discussion of this issue is found in the paper by Hirshleifer & Shapiro in this volume.

instances there will be significant positive correlation among the losses of a great number of individuals. Natural disasters can create losses ranging into the hundreds of millions of dollars. In periods of economic slack, unemployment compensation and welfare payments may increase significantly. Research and development efforts like that for the SST may be so risky that even an organization such as the Federal Government must take cognizance of the tremendous variability in potential cost. On the largest scale of all, the cost of a war is a variable that is difficult to predict within the many billions of dollars.

Some of these projects fall to the Government as part of the natural course of its activities. Others, such as the development of the SST, are taken on because the risks are too great for private entrepreneurs. How should the Government evaluate risky payoffs? The answer depends on the answer to the underlying question risky to whom? If a project is small in terms of the entire Government, and if its payoffs are relatively uncorrelated with other aspects of national income, the Government need not consider variability in its payoff. In this context the Government can be thought of as a mutual investment company that need not consider risks that are large in comparison to the portfolios of any of its individual investors but very small once parceled out among all of its investors.

There may be some Government projects whose uncertain payoffs are directed to specific groups. A small speculative area redevelopment project may be a small risk when looked at for the Government as a whole, but substantially risky for the potential beneficiaries. They might prefer a more conservative project with a lower but more certain expected payoff. The ideal situation in which the Government insures the risks from the speculative project may not be administratively feasible. Here, the Government, in its role as a sort of mutual investment company, must take account of its customers attitudes toward risk. If they prefer the safer project, that is the one that should be undertaken.

A project for which there are substantial uncertainties presents each affected individual with an uncertain set of benefits and costs. It presents him, in effect, with a lottery on net payoffs (benefits minus costs). The efficiency-oriented Government should evaluate the payoffs from uncertain projects in terms of the certainty equivalents of those who pay for and receive its benefits. A certainty equivalent of a lottery for an individual is the "for sure" amount that would afford him a utility equal to the expected utility he would reap from the uncertain payoffs provided by the lottery. It is the greatest amount that he would pay (if negative) or the least amount that he would accept (if positive) rather than participate in the lottery of payoffs that the project would present to him. If the sum of the affected individuals' certainty equivalents is positive, the project should be undertaken.[10]

With this efficiency approach, it need not be the case that the Government will undertake projects that yield positive expected benefits nor reject projects whose expected benefits are negative. A project that produces substantial benefits in times of general well-being will be

[10] This is a gross simplification. It neglects distributional considerations, and assumes that the project does not compete with others that also offer net positive benefits. Certainty equivalents defined at present incorporate discounting considerations.

much less attractive to individuals, will have lower certainty equivalents, than one whose payoffs are negatively correlated with other aspects of income. The latter type of project, for example dikes for flood control, viewed in isolation is a creator of risks. However, when viewed together with all other payoffs it helps to create what in effect is a more balanced portfolio for the entire Nation. As such, it is even to be preferred to a certain payoff equal to its expected value.

In some cases, the direct payoffs from a project itself may not vary greatly, but the project may be responsible for variations in other payoffs. The unexpectedly high cost of the Vietnam war has had the effect of making the Nation forgo much needed domestic social welfare projects. If Congress attempts to hold the Federal budget relatively constant, uncertainties on costs will make their risky aspect felt indirectly through substantial fluctuations in other benefits. If Congress lets the budget swing in response to fluctuating costs, variations in tax rates or the size of the Government debt will be the way the risk is transmitted to the individual.

The Government as a Source of Information*

The acquisition of information may enable us to ameliorate the effects of uncertainty, it may permit us to make informed predictions about future states of nature. We can then change our actions in response to these predictions so as to increase our expected payoffs. Accurate weather forecasting enables farmers to plant and reap at appropriate times. Perceptive forecasts of economic conditions enable corporations to make wise investment and production decisions. Information is a superb example of a public good. Its consumption by one individual in no way reduces the amount that is available to be consumed by another. (In certain competitive situations its value in use, though not its quantity, may be reduced.) As a public good, information is efficiently provided at zero marginal cost. It is natural then that information provision should often be handled by the Government.

In actuality, not only does the Government provide ordinary forecasting and information provision services, but it also carries out testing and grading operations that would be costly and expensive for the individual consumer. In the absence, let us say, of Federal grading of meat, quality would be highly uncertain and the market would function less well. The FDA tests new drugs and makes the results available to all. The argument for Government action here is that of economies of scale. It is much more efficient to have a single organization provide the information rather than have everyone provide it for himself.

The Government as an Economic Agent

Most of the discussion above relates to the role of the Government in overcoming the market imperfections that are created by the presence of uncertainty. In recent years, the Federal Government has come

*Further discussion of this issue is found in the paper by Davis & Kamien in this volume.

to play a conscious economic role in our society. The primary goal of this participation is to maintain the economy along a full employment, low inflation growth path. Ideally the Government acts as a stabilizing force, curbing demand when inflation is a danger, stimulating the economy when it shows a bit of slack. There can be much debate about the effectiveness of these actions, but one thing is clear, the Government's role is to keep the economy from swinging wildly.

It would be incorrect to conclude that the advantage from reducing these swings is primarily one of reduction of risk. The consequences of excess and inadequate demand are undesirable. However, the method in which the Government achieves this stabilization may be through a form of uncertainty reduction. As Keynes emphasized, convergent expectations can control the business cycle. If all businessmen independently would like to invest heavily, but all thought that others would not, a slack economy would be the self-fulfilling fear. The converse would be the case if all thought heavy spending would be the order of the day. However, if the Government assures everyone that it will step in to keep things on an even keel, then galloping or collapsing expectations are not likely to ever take over. The promise (or threat) of Government participation might be enough to prevent the very conditions in which the Government promises to participate. If everyone believes in the exceptional fiscal and monetary skills of the Government, they may never get a severe test.

Conclusion

In recent years, there has been an impressive expansion of knowledge on the subject of uncertainty in economics. For the most part the role of the Government in a world of uncertainty has not been stressed. This paper has attempted to show that the presence of uncertainty creates special obligations and possibilities for collective or governmental action.

CONTRACTING COST AND PUBLIC POLICY

BY HAROLD DEMSETZ

Harold Demsetz is Professor of Economics at the Graduate School of Business, University of Chicago.

Like externalities, decreasing costs, and uncertainty, the existence of transaction or contracting costs is a factor to be considered in resolving the question of the efficient mix of private and public sector activities. As Professor Demsetz points out, the primary test for determining whether the market system attains an efficient allocation of resources is whether or not in particular instances it can establish voluntary agreement among parties to an exchange.

Among the prominent costs of contracting are the costs of search and negotiation and the cost of insuring that voluntary agreements are honored. However, while there are significant transaction costs in the private sector, the allocation of resources to the Government also entails costs and inefficiencies. For one thing, allocation by Government substitutes compulsion by law for voluntary consent. Moreover, when the Government allocates resources, the criterion for decisionmaking may well not be consistent with efficiency. Professor Demsetz concludes that "It may be useful to give the Government the responsibility and incentives for allocating resources where contracting cost looms relatively large. But a bigger improvement * * * may be achievable if the Government * * * eliminates many of the legal restrictions and institutional procedures that presently raise the cost of contracting."

Introduction

This essay is related to the larger and more difficult problem of specifying the conditions under which it seems proper to rely on Government to allocate resources, a problem that is difficult for two primary reasons. First, persons who agree on the results that will be produced if either the Government or the market is chosen to resolve the allocation problem may yet disagree on the alternative they prefer because their personal values weigh important outcomes differently; in particular they may place different weights on the value of preserving a style of life in which voluntary agreements play a large role. Secondly, our knowledge of the workings of political systems, and to a much lesser extent of economic systems, is still too meager to produce widespread agreement on just what will be the outcome if either the Government or the market is relied upon to allocate resources.

We do have a better understanding of this important problem now than 20 or 30 years ago, and I trust that 10 years from now we will be still more confident that we know the answers. But the present state of the analysis is such that the most useful procedure is to discuss a few of the factors that now seem destined to play important roles in resolving the problem of the efficient private-public mix without presuming to know what is the proper resolution. Contracting cost, or as it is sometimes called, transaction cost, one of the more important and neglected of the factors determining this mix, is discussed in this paper.

Much of the discussion by economists of the proper role of Government implicitly accepts the notion that economic efficiency is the

objective sought, where efficiency is interpreted in a much broader sense than the commonly understood engineering concept with the same name. I shall not delve deeply into the technical fine points of the economist's notion of efficiency. For most practical purposes the criterion of efficiency can be defined as the maximization of the total value of economic activity. The components that make up the value of economic activity are not always *recorded* in the market place although they are *measured* through individual decisions in the market place. The distinction between recorded and measured deserves closer examination to show the weight given to the nonpecuniary aspects of life by the efficiency criterion.

An employer sometimes will seek to have some of his employees work overtime. If such work is not a part of the normally understood contractual relationships between the employer and his workers, he will offer an incentive, usually financial, to overcome their reluctance to work overtime. Some of his workers will place small value on the nonpecuniary advantages of enjoying additional leisure time and some will place high value, and, therefore, some will elect to accept the offered financial incentive and others will choose to forego the additional income in order to enjoy more leisure.

Those who choose not to work, place a higher pecuniary value on nonpecuniary leisure than on the financial incentive being offered by their employer. The economic value of the use of their time, as these workers see it, is higher if they reject their employer's offer. The fact that these employees voluntarily have chosen to spend this time at leisure is the only generally accepted standard of evidence that they believe they are better off. This *measurement* of alternatives offered in the market place coupled with voluntary decisions allow one to make deductions about the efficiency of resource allocation.

Can we reach the same conclusion if the problem is viewed from the employer's side? The employer could have secured voluntary consent from this subgroup of workers to work overtime by increasing the financial incentive he offers to them. If he values their overtime labor more than they value their leisure, he would have found it to his advantage to increase sufficiently the financial incentive offered to them to gain their voluntary consent to overtime. The fact that he has not succeeded in negotiating such an agreement is taken to mean that the workers value their leisure more than their employer values their extra work, and hence, that it is efficient for them to stay home even though this may reduce the *recorded* values of their earnings and output.

In the same way, the voluntary acceptance of overtime by the second subgroup of workers allows us to deduce that the value the employer attaches to their extra labor exceeds the value they place on their leisure; in this instance, the recorded value of wages (and output) increases.

In this manner, voluntary negotiations in the market place, whether an exchange is concluded, as with the second subgroup of workers, or is not concluded and not recorded, as with the first subgroup, lead to a maximization of economic value and to efficiency. Nonpecuniary aspects of the alternatives are not ignored but are directly included in this concept of efficiency.

It should be noted that the decisions made in the market place might be different if wealth were distributed differently, since different distributions of wealth give rise to different demands. In addition, the legal and social norms used by a society will influence the decisions taken in the marketplace. If the law throws bankrupts into jail, a different set of voluntary decisions will be made in incurring debt than will be made if the law merely lets bankrupts bear the mark of poor credit risks. But even if the good wealth distribution and the good social and legal norms are specified, the primary test for efficient allocation remains voluntary agreement.

But what can be concluded if the prospective costs of contracting loom large relative to the prospective gain from such negotiation? All that can be deduced within the framework of the usual criterion for efficiency is that it is inefficient to explore further through market channels these particular exchange opportunities.

The potentiality for beneficially altering the allocation of resources through nonmarket techniques is small the lower is the cost of contracting relative to the gains from trade because such gains are most likely to be exploited through private agreements to voluntarily alter resource allocation. Benefits from the use of nonmarket techniques potentially seem greatest where the cost of contracting is relatively large providing, of course, that the cost of using nonmarket techniques is lower than the contracting cost required by market negotiations.

The cost of contracting can be taken to include the costs of search and negotiation in the market place and the cost of insuring that voluntary agreements are honored. The cost of contracting is the value placed by markets on the resources used to make markets work; it is the cost of utilizing voluntary agreements to resolve the problems that arise from competing claims for scarce resources as this cost is measured in the market place.

Nonmarket allocation devices will, of course, have costs of their own.* If taxes are used, there is the cost of collecting and enforcing tax payments. Governmental costs of searching for and administering potentially benefical resource reallocations must be incurred. Nonmarket devices are subject to special errors, as discussed below, and these impose yet more costs on the use of such devices.

Probably the two most important differences between reallocation through the market and through the Government are: (1) The Government need not and does not incur costs to secure the consent of many who will be affected by the resource reallocations; if such costs were incurred, there would be little potentiality for Government reallocation cost to be smaller than contracting cost. (2) There is a greater likelihood of error from the viewpoint of efficiency in reallocating resources through nonmarket techniques. The two sources of this error are the absence of the requirement of voluntary consent on the part of many who are affected and the greater likelihood of a divergence between efficient decisions and politically viable decisions. These two sources of error are briefly discussed below.

The test of voluntary consent, we have seen, is the filter that separates and selects efficient resource allocations from inefficient ones.

*For further discussion of this issue, see the paper by Krutilla in this volume.

But if the Government seeks to realine resources through compulsion by law, the test of voluntary consent is largely forsaken. Thus, in the marketplace persons can decide on a voluntary basis whether or not they want to purchase automobile safety devices. If, however, a law requiring such devices is adopted, we no longer have the test of voluntary consent to help assure us that many individuals view the extra safety as worth the extra cost. Some persons may have preferred to economize on these safety devices and to divert the savings realized to other uses, many of which might increase safety in other directions, such as buying new automobiles more frequently than will be true if persons are forced to invest large amounts in safety devices. In effect, such a law is a tax levied to promote the purchase of certain types of safety equipment by all new-car drivers.

Precisely because the test of voluntary agreement is lacking it is desirable to undertake special investigations of the cost-benefit variety to help insure against errors. These investigations can provide much useful information to guide such Government decisions. Nonetheless, they still suffer from the absence of the test of voluntary agreement. Suppose that we are interested in determining how much the State should spend on automobile safety devices. To answer this question we can calculate the cost of an additional safety device and compare it to the value of the lives we expect it to save. If we are sophisticated, we can calculate this latter value by multiplying the expected decrease in deaths by the value of a typical live person. The value of a typical live person is frequently taken to depend on the discounted value of that portion of his earnings that an accidental death will eliminate.

Setting aside for the moment problems of externalities, the difficulty with this analysis is that the correct solution will be to equate the additional cost of safety devices to the price that persons are willing to pay for expected reductions in their accident rate. This *price* will be an individual matter. It will depend on a person's demand to live longer, on his income, on the prices of other things, and on his *taste for life.* The latter fact is knowable only to himself in principle and, although it will be revealed through negotiation in the marketplace, it is only approximated by a sophisticated cost-benefit analysis. A poor man may be willing to pay a higher price than a rich man for additional expected years of life, especially if he has a greater fear of hell.

The type of error most likely to filter through cost-benefit analyses is the prescription of too much uniformity. For example, the efficiency of having various kinds of safety devices on automobiles is likely to vary considerably between rural and urban areas and between expressway and local driving. Drivers whose driving circumstances differ will generally find different kinds and quantities of safety devices desirable. The cost-benefit approach could in principle distinguish between many such circumstances, but the finer the distinctions the more expensive will be the analysis and the implementation of its conclusions, and the smaller will be the saving associated with compulsion. It is difficult to see just how nonmarket allocation can be achieved at governmental cost less than contracting cost without relying on a more uniform treatment of individuals than is found in the marketplace. An outstanding example of this is the requirement to install smog-reducing devices in all automobiles. Persons residing away from a few major metropolitan areas are unlikely to benefit significantly

from these devices, yet all are required to purchase them with their automobiles.

The second reason that the likelihood of error is increased when nonmarket allocation techniques are selected is that the criterion for decisionmaking brought into play is not necessarily consistent with efficiency. Aside from problems of monopoly in Government or of errors in calculation, in a one-man, one-vote democracy, where votes are not for sale, the polling place will generate information that tends to be based on majoritarian principles rather than on efficiency principles. Thus, suppose some citizens prefer a stronger national defense but that a majority prefer a weaker defense. Left to a vote, the weaker defense will be our chosen policy even though the minority is willing to pay more than the additional cost required to bring defense up to the level they desire (and so, if possible, they may hire private police services). An error in the opposite direction is also possible. The majority of voters may approve of a large space effort even though they would not be able to bid high enough to acquire these resources for space in the absence of forced tax contributions. (Here, however, the minority cannot privately adjust.) It may be that to some extent coalition and logrolling guide majority voting toward more efficient results than would be true with simple majority politics.[1]

The importance of contracting cost in the determination of proper Government policy is revealed by considering the operation of an economic system in which contracting costs are assumed to be zero. Of course, this is an unrealistic assumption just as is the assumption that firms know for a certainty the cost and demand conditions they face. The assumption is useful only because it allows us to grasp the critical role played by contracting cost. That the usual sources of inefficiency fail to exist if contracting costs are zero is described briefly below.*

Externalities.—Persons subjected to harmful effects can, at zero contracting cost, collect payments for bearing these effects and this will lead to an appropriate economizing of the production of these effects. As has been shown by R. H. Coase,[2] divergencies between private and social cost cannot exist in a regime of zero contracting cost. Similarly, those who benefit from what might otherwise be external benefits will pay for such benefits.

In a world of zero contracting cost, the allocation of resources is left unaffected by the assignment of liability for costs. Table 1 shows the costs and returns from operating two enterprises if no damaging interaction is present. Suppose now that some change in the environment, such as a shift in wind, causes the proximity of the two enterprises to result in the suffering of damages equal to $50 by the marginal firm. If the marginal enterprise is left to bear this cost its revenues fall to zero and its inputs shift to other uses where, according to their

[1] On this point see James M. Buchanan and Gordon Tullock, *The Calculus of Consent* (The University of Michigan Press, 1962).

[2] R. H. Coase, "The Problem of Social Cast," *The Journal of Law and Economics* (1960), 1–44. Also, H. Demsetz, "The Exchange and Enforcement of Property Rights," *The Journal of Law & Economics* (1964), 11–26.

*The reader will find detailed treatment of the traditional sources of inefficiency elsewhere in these essays. This discussion proceeds on the assumption that the reader is familiar with these. See especially the papers in Part I of this volume.

opportunity costs, they will receive $50. The advent of the damage makes the marginal enterprise submarginal if it must bear the damage cost.

TABLE 1

Marginal enterprise:		Intramarginal enterprise:	
Opportunity cost	$50	Opportunity cost	$100
Revenue	50	Revenue	125

But if the intramarginal enterprise is made liable for the damage the same allocation consequences will follow. Given the $50 damage, if the marginal firm receives full reparation it will still just break even; it will be no better or worse than before the change in environment. Clearly, however, it will be in both parties' interest if the intramarginal enterprise paid some amount less than $50 to the marginal enterprise to induce the marginal enterprise's resources to shift to other work. For example, a payment of $5 certainly reduces the cost to the intramarginal firm and, since the marginal firm's inputs can earn $50 elsewhere, they will gain $5 more than could be made by remaining. Hence, whichever form of liability rule is adopted, the allocation of resources remains unaffected. Of course, this conclusion does not follow necessarily if contracting costs are positive.

Monopoly.—It is well known from the standard treatment of monopoly that output is kept below its efficient level in order to maintain price. But if contracting costs are zero it is possible for both the monopolist and his customers to negotiate to expand output to the efficient level. Since the efficient output level yields the maximum economic value of the resources employed by the monopolist, its production yields the biggest possible pie for all concerned to share. In a regime of zero contracting cost the problem of monopoly is not one of inefficiency but one of equity—what share of this pie is captured by the firm and what share by the customers.

Economies of scale.—The above argument applies equally well to "natural monopoly" situations—situations in which one firm can produce all relevant output rates more cheaply than two or more firms. But, in addition, if contracting costs are zero it is possible to negotiate costlessly with competing bidders for the contract.[3] Even though only one such bidder will win the contract under conditions of scale economies, his monopoly power is eliminated by the competition in bidding. Furthermore, zero contracting cost allows the use of sophisticated pricing techniques, such as multipart tariffs, in the bidding process. These will yield outputs consistent with marginal cost pricing.

Public goods.—If contracts can be costlessly enforced to prevent nonpurchasers from enjoying the benefits of what others have purchased, the implication of private production of public goods is not inefficiency but only that all who derive benefit can do so more cheaply if they contribute to the cost of producing the good. Anyone who does not pay for the good can be excluded from its use and, so, all who benefit will find it in their interest to contribute to its production. A distinction must be made between public goods subject to high contracting cost and those subject to low contracting cost. It is difficult to envision a situa-

[3] An extended discussion of the absence of a theory of natural monopoly can be found in my paper "Why Regulate Utilities?" *The Journal of Law & Economics* (April 1968), pp. 55–66.

tion in which it is possible to exclude one's neighbors from benefiting from the installation of an effective antimissile missile. The cost of excluding nonpurchasers from benefiting is so great that if the purchase of such missiles is left to private individuals all are likely to wait for their neighbors to make the purchase. But if the cost of excluding nonpurchasers is low, the case for allocation by Government is weakened. It may be possible for an additional moviegoer to enjoy viewing a film once that film is being shown to one moviegoer. The viewing of the film is a public good but, because the cost of excluding nonpurchasers is small, this public good can be produced efficiently through markets. All who wish to see the film can be required to pay to view it or forgo the benefits of being admitted to the theater, unlike all who benefit from the installation of a missile. By such exclusion the market can estimate the value, through the sum of the entrance fees collected, of diverting resources from other uses to the production of theaters and films. If the cost of contracting is zero, including the cost of enforcing the contractual agreements, it is possible to collect fees (which may differ from person to person) that give good estimates of the value on the margin of the benefits achievable by expanding the viewing audience and/or the theater.

What allocation of resources is efficient depends in part on underlying contracting cost conditions. One resource allocation, reached through voluntary agreement in the marketplace, will be efficient if contracting costs are zero. A different allocation, reached in the same way, will be efficient if contracting costs are positive, since under this cost condition it no longer is efficient to exploit all exchange opportunities. The gains from exchange that remain unexploited when contracting costs are positive must be of smaller value than the contracting costs or further trading would be undertaken for profit. Thus the reduction in economic value that might be associated with an externality or a monopoly must be less than the cost of contracting; if the loss in economic value were larger, agreements would be made between those involved to reallocate resources so as to reduce the loss in value to the cost of contracting.

If the Government can make some good guesses as to which voluntary exchanges fail to take place because contracting costs are positive, and if it can make these guesses and accomplish the reallocation at lower costs than the contracting costs involved, then it is possible for the Government to improve on the market's allocation of resources. The word "possible" is important in the previous sentence because the incentives provided by the political system may be such that the Government will fail to accomplish the desirable reallocations and may even substitute undesirable reallocations. A Government agency or department usually is made responsible for supplying goods through the Government. These agencies, for political reasons, may put too many or too few resources into the directions that resource reallocation would take if contracting cost were lower. That Government agencies will make the proper decisions does not follow from the fact that governmental reallocation costs are lower than contracting cost.

Little is known about the factors influencing the level of contracting cost. It does not seem to be true that the cost of contracting necessarily rises, per dollar exchanged, as the number of transactions increases.

There is evidence that the unit cost of contracting falls as the number of transactions increase for certain important goods—those traded on the organized stock and commodity exchanges.[4] In other cases the unit cost of contracting seems to be subject to constant return to scale conditions—retail trading. Where the cost of excluding nonpurchasers from benefiting from the purchases of others is high for technological reasons, such as in the provision of national defense or the establishment of a uniform currency and set of measures and weights, the case for allocation through Government is improved. But often the costs of contracting are high for institutional and legal reasons and not for technological reasons. Usury laws raise the cost of contracting for some loans above the costs associated with the technology of borrowing and lending. Minimum wage laws raise the cost to employers of contracting to employ low-productivity workers. Legal restrictions on entry into occupations and businesses raise the cost to outsiders of contracting with those who would find their services desirable.

Often, property owned by the Government is treated as though it were communal property—all citizens are given a right to use the resourse, say a lake or a road, and no one effectively has the right to exclude others from using the resource. The cost of contracting to prevent congestion and overintensive use of the resource is raised unnecessarily in these instances since a simultaneous decision to reduce usage is required of a large number of prospective users, none of whom can be excluded legally. No one can legally purchase the rights of use from others and, therefore, all are inclined to treat a scarce resource as if it were a free good. A State-owned forest for which the State collects no fee or only a small fee for cutting rights is likely to be cut over too quickly. The same would, of course, be true for a privately owned forest but the private owner's incentives are such that it is unlikely that he will allow much free cutting. *To achieve an efficient allocation of resources, it is necessary for someone*, the Government or private individuals, *to prevent others from using scarce resources as if they were free goods.* This can be accomplished by assessing a proper fee for the use of the resource, but for the cost of contracting to be reduced sufficiently to allow such exclusion to take place it must be possible legally to exclude a nonpurchaser from the use of the resource.

From the viewpoint of efficiency, it may be useful to give the Government the responsibility and incentives for allocating resources where contracting cost looms relatively large. But a bigger improvement in resource allocation may be achievable if the Government reconsiders and eliminates many of the legal restrictions and institutional procedures that presently raise the cost of contracting, sometimes by imposing prohibitions on productive activities.

---

[4] See, H. Demsetz, "The Cost of Transacting," *The Quarterly Journal of Economics* (February 1968), pp. 33–53.

Section D

The Distribution of Income: A Collective Responsibility

COLLECTIVE ACTION AND THE DISTRIBUTION OF INCOME: A CONCEPTUAL APPROACH

BY BURTON A. WEISBROD*

Burton A. Weisbrod is Professor of Economics and Staff Member, Institute for Research on Poverty at the University of Wisconsin.

The existence of market failure implies a role for collective (or governmental) action in achieving an efficient allocation of society's resources. While the desire to improve on resource allocation is surely an important motive for public sector economic activity, concern for the distribution of income in the society imposes at least as great responsibility on the public sector. In this paper, Professor Weisbrod discusses the equity and efficiency of social action influencing the distribution of income.

Following an analysis of the ways in which government action can affect the society's distribution of income, Professor Weisbrod discusses a number of criteria for choosing among the alternative means available to the public sector for redistributing income. These criteria include administrative costs, the efficiency of the measure in influencing the income of specified target groups, incentive or resource allocation effects, the existence of non-demeaning benefits, consumer versus taxpayer sovereignty, and flexibilty over time. Clearly, there are a number of conflicts among these criteria. It is these conflicts which make choices among alternative income redistributional programs especially intractable ones. Professor Weisbrod states "Overall, the objective of this paper is an improved understanding of why income redistribution is a goal of public policy, what alternative means exist for bringing about such a redistribution, and what problems exist in choosing among the alternative means."

Introduction

The public sector is generally conceived by economists as being concerned with three conceptually distinct functions: (1) income redistribution, (2) allocative efficiency, and (3) economic stabilization.[1] This essay focuses on the first function although, as will be explained below, the three are interrelated. The principal point of view is that of the Federal Government, but since State and local governmental units also act with respect to at least the first two functions, much of what is said will apply to all government levels.

The paper is divided into three parts. First is a discussion of the ways in which governmental activities affect—intentionally and unintentionally—the distribution of income in the society. Second is an examination of types of redistributional programs and of criteria for judging their desirability. Third is an analysis of conflicts among some of those criteria and of the resulting difficulty in choosing among alternative programs having a redistributional component. Overall, the objective of this paper is an improved understanding of why income redistribution is a goal of public policy, what alternative means exist for bringing about such a redistribution, and what problems exist in choosing among the alternative means.

[1] R. Musgrave, *The Theory of Public Finance* (New York: McGraw-Hill, 1959), ch. 1.

*I wish to acknowledge the helpful suggestions of W. Lee Hansen, Robert Haveman, Allen Kelley, M. Jack Lefcowitz, T. R. Marmor, and participants in the University of Wisconsin's Workshop on the Economics of Human Resources.

Emphasis throughout is on the problems associated with the "benefits" side of the redistributional process—that is, with helping the "needy." Much of what is said, however, will apply, *mutatis mutandis*, to the problems of deciding who should pay for these benefits and in what manner the required sums should be obtained. Moreover, it is assumed explicitly that income redistributions are intended to aid persons at the lower tail of the distribution of economic well-being—that is, to deal with "poverty." [2]

I. Income Distribution and the Economic Goals of Government

This section takes up each of the three economic roles of Government, examining, first, why Government is concerned with the distribution of income, and, then, how governmental activities in the allocative and stabilization spheres impinge on the income-distribution role.

Income Redistribution as a Goal*

The income distribution in the United States depends largely on the outcome of forces operating through the private market. At its most efficient "best," the private market rewards people in accordance with their contribution to marketable output which, in turn, reflects consumer preferences and incomes. This implies that persons whose productivity, in value terms, is low will earn little—regardless of whether the low productivity is attributable to lack of effort, lack of skill, or low demand for the skill.

The market rewards output, not effort. People who lack education and training, persons with limited intellectual ability or poor health, people producing goods and services for which the supply is so large relative to demand that very low prices result—all of these are likely to earn little. By contrast, those with greater ability, more and better education and training, and having skills that are in relatively short supply will tend to earn more, perhaps even when working shorter hours and under more attractive conditions. As consumer preferences change—partly in response to development of new products—and as technological advances alter production techniques, demand for workers with certain skills rise, and for those with others skills, fall. Given the absence of rapid supply shifts, earnings will behave similarly, as will the rewards to owners of capital equipment, land and other natural resources, the demand for which shifts.

This is all to note the degree to which incomes determined in the private market depend on changeable forces over which the individual has, at best, limited control. Thus, it is understandable that the levels of rewards offered by the market, and changes—particularly sharp drops—in those levels, have become relevant for social policy.

But the discussion above has assumed that the private market was functioning efficiently, actually rewarding those whose social contribution was large. Such a smoothly functioning system is, however, not the rule. The benefits from some activities, for example, in basic

[2] Not all income-redistributional proposals, however, have such a goal. Some may be designed to benefit a geographic region—e.g., the Southwest—or a group felt to be deserving whether or not needy—e.g., war veterans.

*Further discussion of this issue is found in the paper by Bonnen in this volume.

research, may be so broadly diffused that the individuals producing them cannot capture the bulk of their contribution to society. In addition, discrimination—against black, Puerto Ricans, Mexican-Americans, particular religious groups, etc.—has the effect of not fully rewarding people for their actual productivity, or not permitting them to obtain jobs in which their productivity would be maximized. Consequently, the resulting income distribution may be both inequitable and "inefficient"—that is, reflective of an inefficient use of resources.[3] For both reasons, social action to influence the distribution of income has a rationale.

Whatever the forces contributing to the existing distribution of income, there may be dissatisfaction with that distribution and, hence, attempts to alter it. The Federal Government is generally the most effective governmental level to undertake redistributions because the opportunities to circumvent Federal actions are most limited. By contrast, attempts by individual State or local governments to help certain people are likely to produce incentive effects resulting in an influx of the categories of persons to be assisted, while at the same time resulting in an exodus of persons and businesses who must pay for the aid being provided others.

The reference above to discrimination in the labor market suggests what should now be made explicit: collective (governmental) attention to the distribution of income need not involve merely the size distribution of income. There may also be concern about the distribution of income between whites and nonwhites, the distribution among geographic regions (witness recent efforts to develop Appalachia), and the distribution among age groups.*

The public sector's concern about the distribution of income should not be limited to the consequences of the *private* sector's concentration on efficiency and its essential disregard of income-distributional considerations. The public sector itself, as we shall see, is a significant contributor to undesired income redistributions. These occur as Government pursues its other goals of allocative efficiency and economic stability. That is, governmental activities, whatever their avowed objective, frequently if not inevitably affect individuals' income levels differently.

This observation is important for two reasons. One is that insofar as governmental activities designed to alter the allocation of resources—for example, highway and water resource programs—also have unintended side effects on the distribution of income, wise governmental policy should consider these effects when judging and designing such programs. The other is that since such "resource-allocation" programs actually do affect the distribution of income, they might be utilized deliberately for this purpose. The point is that a change in the income distribution—that is, in the absolute or relative positions of particular people or groups—is achievable in a variety of ways which include taking advantage of the redistributional "side effects" of programs that are normally regarded not as income-redistributional but as resource allocative.

---

[3] For an analysis of the necessary and sufficient conditions for achievement of economic efficiency see J. de V. Graaff, *Theoretical Welfare Economics* (Cambridge: Cambridge University Press, 1957); F. Bator, "The Simple Analytics of Welfare Maximization," *American Economic Review,* March 1957, pp. 22–59; or J. Henderson and R. Quandt, *Microeconomic Theory* (New York: McGraw-Hill Book Company, 1958), chapter 7 ("Welfare Economics").

*Further discussion of this issue is found in the paper by Levine in vol. 3 of this collection.

ALLOCATIVE EFFICIENCY AS A GOAL

When Government attempts to influence the production of particular goods and services—housing, highways, dams, schools, rivers and harbors, air traffic control systems, etc.—even though the motivation may be to enhance allocative efficiency, the fact is that these activities do affect the distribution of income, sometimes in subtle though powerful ways. Increased spending on highway construction is likely to send upward the price of cement companies' securities, as rising profits are expected. The automobile industry and its employees would also benefit, but officials of commercial airlines are likely to be less than enthusiastic. Water-resources projects in the arid Southwest enhance land productivity in that region; but even if it were shown that such projects were efficient for the economy as a whole, it would remain true that agricultural producers and workers in other regions would be hurt by this new competition.

This essay, concentrating on the rationale for governmental concern about the income distribution, is not the place to delve deeply into the justification for governmental activity as a provider of goods and services or as an influence on the level and forms of private provision of particular goods and services. Justifications have been offered—involving the provision of "collective-consumption" goods (such as national defense), for which consumption by one person does not reduce the consumption opportunities available to others, or involving, more generally, goods such that external effects are significant and are not taken into account by private decisionmakers (anti-air and water pollution activities illustrate most clearly this type of commodity.) [4]

When "allocative efficiency" is being considered, whether with respect to private or governmental activities, it is vital to understand how that term is defined—particularly in the context of our attempt to examine the rationale for governmental attempts to alter the distribution of income. The key point is that there is no single "efficient" allocation of the economy's resources; instead, there is an efficient allocation corresponding to each initial distribution of income. This is so because the income distribution affects consumer demand patterns and resource-owners' supply patterns which, in turn, determine the pattern of production that will maximize the economic welfare of the society, given those demand and supply patterns. The efficient allocation of society's resources is not defined in abstraction from the distribution of income. What is efficient for the economy to produce with one income distribution may be quite inefficient if the distribution of income were altered.

It has been pointed out above that when the Government alters its purchases or production of goods and services, it benefits (or hurts) particular groups of people and in this way it alters the distribution of real income. Whether the Government's primary objective in undertaking a given program is redistribution, increased allocative efficiency, or economic stabilization is immaterial; one effect of a change

[4] At the analytic level see W. Baumol, *Welfare Economics and the Theory of the State* (Cambridge: Harvard University Press, 2d edition, 1965), and P. Samuelson, "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, 36 (Nov. 1954).

in either the composition or level of Government spending is a change in the income distribution.[5]

These distributional effects extend beyond the initial impact. More is involved than simply whose land is made more productive by a Government-financed flood control project or who saves time and expense by traveling on a new segment of interstate highway. For in addition to the distributional impact of these real benefits from the projects, there are many secondary effects that are income redistributional in character. Property values near the new highway will leap as opportunities are sought to construct new motels, restaurants, and automobile service stations. These increases in land values have their counterparts, however, in the decreases occurring elsewhere, particularly along the roads from which traffic is being diverted. Urban renewal projects bring about similar pecuniary or income-restributional results. Not only are property owners affected—with some gaining and others losing—but residents, too, feel the consequences. Some receive relocation subsidies although they planned to move even without them; and others suffer because a reduced supply of housing units pushes rents upward, at least temporarily.

Beyond the direct and the secondary distributional effects of a program, additional distributional effects will be felt through the respending process. Those whose incomes have increased may well have marginal expenditure patterns different from those whose incomes have decreased. As a result, the respending effects, operating through the conventional economic "multiplier" process, will bring about further distributional changes.[6]

So far our attention has focused largely on governmental *expenditures*, but, of course, decisions regarding the method of *financing* governmental services also affect the distribution of real income in the society. A benefit-principle of taxation assumes implicitly that Government provision of goods and services should be financed in such a way as to leave unchanged the distribution of income (somehow-defined).[7] The ability-to-pay principle, by contrast, explicitly contemplates income redistribution, although whether a redistribution actually occurs depends on how the benefits of the governmental goods and services are divided among income groups. If benefits are correlated positively with ability to pay—as may be the case with many publicly-provided services such as interstate highways, waterway improvements, and art museums—then adoption of the "ability" principle of taxation is equivalent to adoption of the benefit principle. Unfortunately, little is known at present about how the benefits of public services are divided among persons at various levels of ability to pay.[8]

---

[5] It should also be noted that a change in the composition of Government expenditures may alter the amount and distribution of aggregate unemployment in the economy, and this, too, has clear implications for the distribution of income. For further discussion see R. Haveman and J. Krutilla, *Unemployment, Idle Capacity and the Evaluation of Public Expenditures.* (Baltimore: Johns Hopkins Press, 1969.)

[6] For a discussion of this effect in the context of geographic income distributional consequences of water-resource programs, see S. Marglin, *Public Investment Criteria* (Cambridge: MIT Press, 1968), pp. 83–84.

[7] A "benefit principle" of taxation could be defined in various ways and although all would involve beneficiaries paying the total cost of the public service, the cost burden could be shared in various ways. Hence the net effect on the distribution of income might well not be "neutral."

[8] For an examination of this matter in the higher education area see W. L. Hansen and B. A. Weisbrod, "The Distribution of Costs and Direct Benefits of Public Higher Education: The Case of California," *Journal of Human Resources,* (Spring 1969), pp. 176–191.

The stabilization efforts of Government involve changes in tax revenues, expenditure levels, and/or in monetary variables that alter interest rates and credit terms. Whenever tax revenues are altered by adding or abolishing taxes, raising or lowering tax rates, or adding or eliminating deductions, exemptions, or exclusions, the result is that particular groups of persons benefit or are hurt. There is no escaping this fact and, indeed, it is sufficiently evident that lengthy debates on the merit of proposed redistributions inevitably accompany consideration of legislative change in taxes.

When and insofar as fiscal policy for economic stability involves changes in expenditure levels the income distributional consequences are also powerful. These consequences are possibly somewhat less evident to the society in general, because changes in the level of any one expenditure program are likely to have a pronounced effect on only a small segment of the population. At any rate, the fact remains that different people benefit from Government expenditures on elementary-secondary education, higher education, medicare, medicaid, interstate highways, water resources, and urban renewal. Thus, the distribution of *real* income—of the economic well-being of various people—is inevitably affected by Government efforts to stabilize the economy by varying expenditure levels. The point is that when fiscal policy is being utilized for stabilization, it is not an amorphous category of "expenditures," that is varied, but expenditures on specific programs, each of which affects some persons more than others and in different ways.

To summarize this section: Income redistribution is only one of the three primary economic goals of Government; yet as the Government seeks to achieve the other goals of efficiency in the allocation of resources and employment stability of the economy, its actions inevitably affect the income distribution. Thus, governmental concern about the distribution of income reflects—or, at least, it ought to reflect—both the undesirable distributional outcome of private sector activities—for that sector is preoccupied with efficiency and profits, not with distributional consequences—and also any undesirable side effects of the Government's own activities in pursuit of the goals of economic efficiency and stability.

If income is to be redistributed, two questions must be answered: (1) *What* (and how much of a) redistribution should be brought about—that is, to whom and from whom should the redistribution be made—and (2) *how* should the redistribution be brought about—that is, what method should be used? These are perfectly analogous to the questions involved in ordinary production decisions by a firm: (1) What (and how much) should the firm produce, and (2) what technique of production should it employ? These questions are not really independent, in either the production or the redistribution context. The answer to question 1 depends partly on costs which, in turn, depend on the choice of production or redistribution technique (question 2). And the answer to question 2 depends in part on the type of commodity or redistributional program that is being considered (question 1). Thus, in principle these questions should be dealt with simultaneously.

The emphasis of this paper, however, is on questions of the second type. Granted that a tentative decision has been made to redistribute from group A to group B, and to provide a given level of benefits (in some form) to group A, which redistributive technique—among the alternatives available—should be used?

If a number of criteria are deemed relevant to program selection—and this is doubtless the case—the problem is, first, to identify the criteria; second, to develop operational measures of how fully each criterion is met by any particuar program; and, third, to determine the relative importance of each criterion—that is, the weights that should be attached to a given level of achievement of each criterion. These weights imply normative judgments regarding what is "desirable" public policy in the income distribution area. The summation of these weighted criteria-measures would give a "score" for each redistributional technique, the highest score determining the technique to be used.

Once the highest score technique for bringing about a given redistribution has been determined—a most difficult task—the desirability of that redistribution may be called into question. It might be decided that the best program available for producing a given redistribution is so poor that plans for the particular redistribution should simply be dropped. Some other redistribution, involving groupings other than A and B (though perhaps overlapping with them), might be preferable. In other words, while the A–B redistribution might be thought preferable to one involving groups C and D, the best technique available for the latter might be considerably better, and so the C–D redistribution might be chosen. This would be analogous to case involving some ordinary commodities in which one commodity was more valuable than another (at the margin) but the cost was so much greater that it was deemed best (more efficient) to produce the less-valuable good.

The tasks of (1) stating decision criteria for choosing among redistributional programs, of (2) devising operational measures, and of then (3) determining weights, constitute a large order. The emphasis in this paper is on the first step, and although there is some attention to the second step, the paper stops short of tackling the third.

II. Criteria for Choosing Among Alternative Means for Redistributing Income

A number of alternative approaches exist for a government that desires to alter the distribution of income. In general the alternatives involve programs to transfer money, to transfer income *in kind*, or to enhance earning power for the group to be aided. Transfers in money form may take such diverse forms as a "negative income tax"—which might be addressed to all low-income persons—or agricultural price supports—which are available only to farmers. The money-transfer program may allow the recipient to spend the added income as he wishes (negative income tax) or it might constrain the use (rent supplements). Transfers in kind occur via distribution of surplus food or subsidies for production or consumption of specific goods such as public housing. Redistributions can also take forms aimed not at expanding consumption opportunities directly—as the programs mentioned above do—but aimed at increasing the individual's produc-

tive capacity and earning power, and in this way increasing his consumption opportunities indirectly. Education and training programs are the principal means by which this approach is currently being pursued.

We turn now to the criteria by which the desirability of any particular approach can be judged. Since a number of avenues are available for influencing the income distribution, the first step in choosing intelligently among them is the specification of standards or norms.[9] The following is a list and discussion of six such standards. It is not a logically exhaustive list. Hopefully, however, it is one that can be supported widely as a basis for program choice. The emphasis here is on the "transfer" side of the tax-transfer process, although much of what is said can be applied also to the tax portion of the redistributional mechanism.[10]

1. *Administrative cost.*—One consideration relevant to program selection is the cost of its administration. A family allowance, for example, involving systematic payments to all families with dependent children,[11] requires smaller administrative costs per dollar of benefits than does a conventional welfare program with its elaborate machinery for determining eligibility, preventing cheating and monitoring beneficiary behavior. To be sure, a principal reason for the divergence of administrative cost is that the welfare programs are intended to make a great effort to assist only the very needy, and clearly it is more costly to identify them than, for example, simply to identify a family with a dependent child.

Implied by this last statement is a second criterion for program selection—that the "deserving," and only the deserving should be aided. To this we turn next.

2. *Target efficiency.**—By this term I mean the degree to which the actual redistribution coincides with the desired redistribution. For example, a manpower retraining program may be intended to benefit the hard-core unemployed—those who cannot find regular employment even under full-employment conditions—but as it is actually administered it may (1) miss many in the hard-core group, while at the same time it (2) aids a number of less needy persons. Some would argue that these effects detract from the desirability of the program. Similarly, insofar as other antipoverty programs turn out to benefit the nonpoor while bypassing many of the poor, they may be regarded as inefficient means of redistributing income.

---

[9] This amounts to stating those elements in a social welfare function that are likely to be affected by the choice of a particular redistributional mechanism.

[10] The usual intention of such mechanisms is to aid some needy group, not to penalize any other group, and so the choice of an aid program normally is considered separately from the decision regarding how to finance the program. Indeed, the financing is customarily from general revenue, no explicit attention being given to the question of which group should pay for a particular income-redistributional program. The distributional effects of alternative tax policies have received considerable attention. See, for example, Musgrave, *op. cit.*, chs. 4, 5, and 8; L. Johansen, *Public Economics* (Chicago: Rand McNally & Co., 1965), ch. 7; and J. Pechman, *Federal Tax Policy* (Washington, D.C.: The Brookings Institution, 1966).

Although the tax and transfer (benefit) structures are often considered separately, they are, to some extent, substitutes. Insofar as the needy are taxed they have more need for transfers. (See V. Tanzi, "Governments' Approaches to Income Redistribution: An International Comparison," *National Tax Journal*, XXI (December 1968), pp. 483–486.)

[11] C. Green, *Negative Taxes and the Poverty Problem* (Washington, D.C.: The Brookings Institution, 1967), pp. 46–50; and J. Vadakin, *Family Allowances* (Miami: University of Miami Press, 1968).

*Further discussion of this issue is found in the paper by Schmid in this volume.

Two issues are involved, having to do with the accuracy of the program in assisting *only* the "target" group, and the comprehensiveness of the program in assisting *all* of that group. The first matter—the degree to which a program intended to benefit a low-income group A also benefits a higher income group B—might be termed the program's vertical target efficiency (or, more simply, vertical efficiency). The second is the degree to which a program intended to benefit group A reaches all members of this group. This concept of comprehensiveness may be termed the program's horizontal target efficiency (horizontal efficiency).

The vertical efficiency of a program might be defined as the ratio of benefits received by the intended beneficiaries to total benefits. A ratio of unity would thus indicate that all resources of the program are being devoted to the target group and none to any other group.

Horizontal efficiency may be defined in the two dimensions: (1) as the ratio of the number of beneficiaries in the target group to the total number of persons in the target group, and (2), as the ratio of benefits going to the target group to the total benefits "needed" by that group. Whereas the vertical concept involves the reaping of benefits by persons outside the target group, the horizontal concept involves the absence of "adequate" benefits to some or all persons within the target group. The concept of horizontal efficiency is related to the equity norm of equal treatment of equals,[12] for the lower is the value of ratio 1 the smaller is the proportion of the target group that is being treated. But the horizontal efficiency concept is also related to the equity norm of adequacy, for the lower the value of ratio 2, the smaller is the proportion of the target-group *problem* that is being treated. All persons in the target group might be treated equally but inadequately. Only in the limiting case in which (1) everyone in the target group received benefits, and (2) the benefits were at an adequate level, would full horizontal efficiency—in both dimensions—be achieved.

Governmental efforts to redistribute income—whether by transfering money, income-in-kind, or investment opportunities for raising earning power—often fail to reach all of those whom it is desired to aid (or tax), and aid (or tax) others unintentionally.[13] A desire to assist the "poor" may become operational in a program that disqualifies some of the poor on grounds of an inadequate period of residency. At the same time the program, if based on current income alone, may provide aid to persons whose low income is temporary or who possess considerable wealth (net worth) even though having very limited income.[14]

Governments sometimes provide goods and services at subsidized or even zero prices in an effort to raise the real income of certain persons. But the actual beneficiaries are frequently not the same as the intended beneficiaries. The provision of medical attention in a "free"

---

[12] In the context of tax policy the analogous concept is "horizontal equity"—equal taxation of people with equal ability. See R. Musgrave, *op. cit.*, ch. 8.

[13] As Kenneth Boulding put it, discussing the California water plan:

> "It would be well to be quite sure
> Just who *are* the deserving poor,
> Or else the state-supported ditch
> May serve the Undeserving Rich."

[14] For an analysis of effects of considering net worth when estimating the prevalence and distribution of poverty, see B. A. Weisbrod and W. Lee Hansen, "An Income-Net Worth Measure of Economic Welfare," *American Economic Review,* December 1968.

clinic does, no doubt, benefit many of the poor. Yet the frequently long waiting time required discriminates against the "working poor"—those for whom the opportunity cost of waiting is high because it involves the loss of work time and earnings.[15] (This is the horizontal efficiency criterion.) At the same time, the clinic may be utilized by some persons outside the target group—some of the nonpoor—those for whom the opportunity cost of waiting is low, because they forgo little by waiting.

In a recent study of the distribution of public higher education subsidies in California, W. Lee Hansen and I discovered that the amount of subsidy actually received is correlated positively with the level of parental income of students.[16] A key factor is the presence of a very low, heavily subsidized tuition rate—which applies to all students, regardless of their ability to pay. Students from higher-income families benefit most from this subsidy because they are more likely than their lower-income counterparts to go to college, are more likely to go to a high cost (high public subsidy) school, and are more likely to remain in college until graduation. Thus, a public higher-education system that many feel is, or at least should be, a means of redistributing economic opportunity in favor of children of the poor is having quite the opposite effects, in the aggregate. That is, public higher education performs rather badly by the vertical efficiency criterion. To be sure, the public subsidies are permitting some low-income students to attend college when they would otherwise be unable financially to do so, but many of the poor receive little or no subsidy through the higher education system—thus causing the program also to perform badly by the horizontal efficiency criterion—at the same time that many nonpoor reap the largest subsidies.

The vertical efficiency criterion poses a thorny problem of the importance of distributional as compared with efficiency goals of government tax-expenditure policy. The question is whether a program such as compensatory education for the disadvantaged or retraining for the hard-core unemployed—which are aimed at particular kinds of people—should count any benefits (outputs) that accrue to persons not in the target group, and, if so, whether these benefits should be given the same weight per dollar as benefits to the intended beneficiaries. The importance of this issue is indicated in a recent study by Thomas Ribich of a number of compensatory education programs.[17] He argues that since many children from poor families will not grow up to become poor adults, compensatory education, considered only *as an anti-poverty device*, need not be provided to *all* poor children. Whatever other arguments may be made in favor of added government spending on education in general or on education specifically for poor children, the point is that some of it will be provided to the nonneedy. The issue, then, is how to weigh the benefits from compensatory education programs that accrue to the needy versus the nonneedy.

Vertical efficiency may be desired for either or both of two reasons. For one, taxpayers may have a psychological aversion to seeing nonneedy persons receiving benefits. For another, vertical efficiency can be

---

[15] This point came out in a discussion with Eugene Smolensky.

[16] See W. Lee Hansen and Burton A. Weisbrod, *Benefits, Costs, and Finance of Higher Education* (Chicago: Markham Publishing Co., 1969), especially chapter IV.

[17] Thomas Ribich, *Education and Poverty* (Washington, D.C.: The Brookings Institution, 1968).

viewed as a proxy for cost effectiveness in the sense that the greater is the vertical efficiency ratio the smaller is the cost per unit of benefit to the target group. Thus, as between two programs that bring equal benefits to a given target group the program having the higher vertical efficiency ratio, *ceteris paribus*, will be the least costly.

This latter interpretation implies, however, that benefits to nontarget persons should be given a weight of zero, and this is a questionable procedure. Even though the benefits to nontarget persons are agreed to be valueless from the standpoint of redistributional programs, they may represent real benefits which can be justified on other than redistributional grounds—for example, allocative efficiency. For example, a compensatory education program that is aimed at reducing the future prevalence of poverty among today's poor children may increase learning and subsequent labor market productivity of some who would not be poor even in the absence of compensatory education; this increase in productivity, while possibly valueless with respect to the redistributional goal, is quite relevant to the overall benefit-cost evaluation of the program—an evaluation that ought to reflect allocative efficiency as well as distributional-equity goals of Government economic policy.[18] This leads to the conclusion that some "unintended" benefits of a redistributional program should receive a positive weight—when they contribute to nondistributional goals.[19]

Another objection to giving a weight of zero to those benefits from redistributional programs that accrue to the nontarget groups, relates to the observation above that there may be an aversion to seeing windfall gains to the "undeserving." Thus, such unintended benefits to the nontarget group should receive a weight not of zero but of less than zero. This aversion is especially likely to be expressed when pure transfer programs are involved, for then it is clear that benefits to nontarget persons are unlikely to be justifiable on allocative efficiency grounds. Thus, one of the objections to the various negative-income-taxation (NIT) plans for transferring income to the poor is that in order to provide a meaningful floor on income and to avoid very high "tax" rates on incremental income—reductions in NIT payments per additional dollar of the unit's income—it is necessary that sizable payments go to persons who clearly are not poor.[20] This fact would be disclosed, of course, by the vertical efficiency ratio, though possibly not in a fully satisfactory manner.

3. *Allocative efficiency**.—At issue here are the effects of the redistributional scheme on the allocation of resources. Principally involved are the effects on incentives of those who benefit from the redistribution and of those who are made worse off by it. Conceptually, incentive effects are avoidable by the utilization of "lump sum" trans-

[18] For further discussion of appropriate relationships between efficiency and equity in Government decisionmaking see B. A. Weisbrod, "Income Redistribution Effects and Benefit-Cost Analysis," in S. B. Chase, Jr. (ed.), *Problems in Public Expenditure Analysis*, Washington, D.C.: The Brookings Institution, 1968), pp. 177–209. See also the paper by Freeman in this volume.

[19] Similarly, the favorable (unfavorable) distributional effects of programs undertaken for allocative-efficiency reasons should receive positive (negative) weight.

[20] C. Green, *Negative Taxes and the Poverty Problem, op. cit.*, especially chs. V and VI; C. Green and R. Lampman, "Schemes for Transferring Income to the Poor," *Industrial Relations*, February 1967; J. Tobin, *et. al.*, "Negative Income Taxation—Practicality and Cost," *Yale Law Review*, November 1967; and W. Klein, "Some Basic Problems of Negative Income Taxation," *Wisconsin Law Review*, summer 1966, pp. 1–25.

*Further discussion of this issue is found in the paper by Zeckhauser in this volume.

fers to the target group and lump-sum taxes on those who are to finance the transfers. The essence of the lump-sum taxes and transfers is that, while leaving some persons richer and some poorer, no one could avoid the tax or increase the benefit he receives by altering his behavior.[21] Thus, no incentives would be established for resource suppliers or consumers to alter their behavior except as a consequence of having more (or less) income.

The fact is that incentive effects cannot, in practice, be avoided. We have discovered no way to tax or to subsidize without producing incentives to avoid paying the taxes and to qualify for receipt of the subsidies. Thus, the current controversy over negative income tax or other schemes to help the poor centers largely on the (unknown) magnitude (but not the direction) of incentive effects on the recipients who, as a consequence of an income guarantee, might decide to work less, or to have more children. Whether such consequences are "bad" is obviously a value judgment, but they appear to be fairly widely deplored in our society at this time. The negative income tax experiment now underway in New Jersey—designed and administered jointly by Mathematica, Inc. and the University of Wisconsin's Institute for Research on Poverty—has as its principal research objective the assessment of work-incentive effects of various levels of income guarantees.[22]

Another form of incentive effect—one that is believed to be caused by AFDC (aid for families of dependent children) transfer programs—is the breakup of families. In some States a low-income or unemployed father can increase his family's income by deserting it, thereby permitting the family to qualify for AFDC payments. The quantitative significance of this behavior is not known, but such an incentive exists, and was a factor leading to the adoption by more than 20 States of an AFDC-UP (unemployed parent) program under which a family with an unemployed father could qualify for welfare aid.

For decades, economists have generally faulted the agricultural price support programs because of their adverse incentive effects. Since the subsidies even though joined with land-use diversion programs, are available only to farmers, the effect is to establish incentives to remain in farming. Thus, the movement of capital, land, and labor out of agriculture and into more productive activities has been slowed.*

Because we are unable to devise redistributional devices that produce no incentive effects at all, all redistributional programs fall short of the ideal, but, clearly, some are better than others. Most welfare programs are especially weak in this respect, because welfare recipients who increase their earnings are frequently penalized by cuts in welfare payments that are as large, and sometimes even larger, than their additional earnings.[23] By contrast, manpower training programs produce an incentive to take advantage of new skills and thereby to increase earnings.*

---

[21] There would occur shifts in demand through the income effects, of course, but there would occur no substitution effects, given the absence of any initial change in relative prices.

[22] For additional information see H. Watts, "Graduated Work Incentives: An Experiment in Negative Taxation," *American Economic Review*, Papers and Proceedings, May 1969 and the papers by Rivlin and Levine in vol. 3 of this collection.

[23] C. Green, op. cit., especially p. 119.

*Further discussion of this issue is found in the paper by Bonnen in this volume.

*Further discussion of this issue is found in the paper by Mangum in vol. 3 of this collection.

As the manpower-training program illustration makes clear, allocative-efficiency effects are not limited to pure transfer programs. Whenever the redistributional technique involves the use of real resources (e.g., in manpower training or highway construction) rather than simply transfers, the question arises of whether and to what extent the resources are being used in an economically efficient manner—efficient in the sense that the value of outputs (better trained men, new roads) exceed the costs of production. If the equity effects of a particular program are sufficiently favorable, it may be worthwhile to sacrifice some economic efficiency, but in any event, account needs to be taken of both types of program consequences.

4. *Nondemeaning benefits.*—This criterion reflects the norm that benefits should not be provided in a manner that tends to stigmatize or to destroy self-respect. This criterion can be said to reflect a concern not only about how much income people have, but also as to what they must do to qualify for a redistributional aid program, and in what manner the aid is provided. Receipt of a check with the imprint, "Department of Welfare," might, by this standard, be regarded by recipients as inferior to an equal amount of money received from some other agency "for services rendered." A price-support program, in which a subsidy is paid as an above-market price for a product, would, in this context, be preferred, at least by recipients, to an outright cash dole.

Even when it is agreed that a particular redistributional program is demeaning to its beneficiaries, there may not be concensus that this is undesirable. "The degree of stigma associated with a program may be viewed as one means of rationing Government program utilization." [24] In addition, the stigma may be desired (by taxpayers) because of its incentive effect, encouraging recipients to move out of the needy category.

At the same time it is worth considering that what taxpayers regard as demeaning or stigmatizing may not be so regarded by beneficiaries. Recent research has disclosed that welfare programs—which seem to be quite demeaning in their requirements, particularly for proving an individual's eligibility—are not regarded by recipients as especially demeaning.[25]

5. *Consumer versus taxpayer sovereignty.*—Some would argue that if the distribution of income is viewed as unsatisfactory, then taxes and transfers should be made in money form, thereby permitting each taxed person to decide how to reduce his spending, and permitting each transfer recipient to decide for himself what to do with his added purchasing power. This is the principle of consumer sovereignty, embodying the liberal ethic of freedom of choice. Given the consumer's set of preferences he can maximize his economic welfare if the tax he pays or the transfer he receives occurs in an unconstrained form. (In fact, this is a necessary condition for efficiency in the consumer sector.)

If consumer sovereignty is extended to all consumers, the beneficiaries of an income-redistributional program would be free to do as

---

[24] T. R. Marmor, "Income Maintenance Alternatives and American Politics," University of Wisconsin, mimeo, n. d. (circa February 1969), p. 3. (Paper presented at the Poverty Seminar of the American Academy of Arts and Sciences, Boston, Mass., May 16, 1969.)

[25] J. F. Handler and Ellen J. Hollingsworth, "How Obnoxious is the 'Obnoxious Means Test'? The Views of AFDC Recipients," University of Wisconsin, Institute for Research on Poverty, Discussion Paper 31-69, January 1969.

they wished with their added income. Those who make the payments, however, may be willing to give up income only if their sacrifices produce added purchasing power that is used in particular ways—"wisely"—by the recipients. Of course, conflicts can and do arise over the meaning of wisely, and so taxpayers may wish to constrain the freedom of recipients in the uses to which their increased income may be put. Thus, "taxpayer sovereignty" may supersede consumer sovereignty. When this occurs, a rather strong value judgment is implied regarding who knows best what a consumer ought to have—the individual himself or someone else? Traditional liberalism has opted for the former—as has modern theoretical welfare economics [26]—but for reasons discussed below, actual redistributive policy has often followed another course.[27]

Nowhere is this issue of who should decide how additional income should be spent seen more clearly than in current debate over governmental antipoverty policy. The case for a negative income tax, a children's allowance, or a guaranteed income rests firmly (though not entirely) on the foundation of freedom of choice—consumer sovereignty. By contrast, programs to provide goods and services for the poor—public housing, compensatory education, medical care clinics, or training programs—suggest taxpayers' preferences for minimum standards in consumption of housing, health services, and education, but apparently not in clothing, entertainment, or automobile transportation. It is not my objective here to urge one or the other of these viewpoints, consumer versus taxpayer sovereignty; rather it is to note that the choice between them is rooted in a value judgment regarding who ought to determine the pattern of consumption by beneficiaries of a public program.

The issue of transfers in kind versus transfers in cash (or, better perhaps, constrained versus unconstrained transfers) may be viewed from another perspective—allocative efficiency. If consumption of certain goods and services produces external economies (favorable effects on outsiders) while consumption of other goods produces external diseconomies (unfavorable effects on outsiders), it is both understandable and economically efficient that consumption of the former class of goods be encouraged (e.g., subsidized) while consumption of the latter class of goods is discouraged (e.g., taxed).

Taxpayers' interest in providing better housing for the poor may thus be explainable as a response to the external diseconomies that slum housing causes for outside parties—fires, crime, personal despair.[28] Similarly, the interest in providing better health care for the poor may be justified as at least partially a response to external diseconomies—in particular to the dangers that contagious disease may spread.[29]

---

[26] See, for example, E. J. Mishan, "A Survey of Welfare Economics, 1939–59," in Mishan, ed., *Welfare Economics* (New York: Random House, 1964), p. 6; A. Bergson, "Consumer Sovereignty and 'Real' Income Distribution," in Bergson, ed., *Essays in Normative Economics* (Cambridge: Cambridge University Press, 1957), ch. I.

[27] An alternative way of viewing this issue is that while consumer sovereignty is accepted generally, a question exists as to who is the "consumer"—the program beneficiary or the taxpayers?

[28] For further analysis see Jerome Rothenberg, *Economic Evaluation of Urban Renewal* (Washington, D.C.; The Brookings Institution, 1967), especially pp. 54–56.

[29] See Burton A. Weisbrod, *Economics of Public Health* (Philadelphia: University of Pennsylvania Press, 1961), especially pp. 18–20.

6. *Flexibility over time.*[30] Redistributive programs are seldom once-and-for-all events; they are continuing efforts. As such, their effects need to be examined over a "suitable" time period. The length of this period is itself a policy variable which is manifest in discussions about the intergeneration consequences of various antipoverty programs and about the likelihood that a program will contribute to increased self-sufficiency and decreased dependency. The point to be emphasized here, however, is this: Even if the program were fully in accord with the criteria of desirability at a given point in time, there remains the question of whether it will remain satisfactory as conditions change. For example, as technological advances have increased economies of scale in agriculture, price supports have increasingly benefited not the small, low-income family farmer, but the large, profitable corporate farm enterprises. (The extent to which price supports for one crop, cotton, are benefiting primarily the largest producers is documented in a recent study by James Bonnen.)[31] Perhaps such programs are less justifiable with the passage of time, but once instituted they are politically difficult to eliminate. It is important, therefore, that redistributional programs be flexible and adaptable to new conditions, either through an automatic mechanism that brings about needed shifts or a more responsive political system.

III. Policy Choices

Two principal types of problems arise in applying the six criteria presented above. One is the lack of information as to how well any specific redistributional program will work, in terms of each criterion. Second is the conflict among criteria.

Consider first, the information problem. With regard to criterion 3, the incentive issue, how substantial are the likely work-disincentive effects of a negative income tax for the poor? How large are the external diseconomies resulting from the greater fire hazards in slums? Regarding criterion 5 (consumer sovereignty and taxpayer sovereignty), to what extent is there a difference between the uses to which poor people would put additional income if they had it, and the uses to which the nonpoor wish to allocate additional income?

Turning to the target efficiency criterion (2): How much of the benefits of a program will accrue to the intended beneficiaries and how many of the intended beneficiaries will receive little or no help? While the answer to this question is frequently unavailable because the matter has not been studied, a contributing factor is that it is often not entirely clear precisely whom the program aims to benefit. The issue is how one moves from a conceptual definition of the "needy" (or "deserving," or "poor") to an operational definition. For example, recent analyses of the "poor" population of the United States have disclosed rather different numbers of poor families and different characteristics of those families, depending on which operational definition of poverty is used—e.g., at least $3,000 annual income per

[30] This general point came out in conversation with W. Lee Hansen.

[31] See his "The Distribution of Benefits from Cotton Price Supports," in Samuel B. Chase, Jr. (editor), *Problems in Public Expenditure Analysis* (Washington, D.C.: Brookings Institution, 1968), pp. 223–248; and J. Bonnen, "Distribution of Benefits of Selected Price Support Programs," in *Rural Poverty in the United States*, A Report of the National Advisory Commission on Rural Poverty, USGPO, 1968. See also the paper by Bonnen in this volume.

family,[32] at least $3,000 of combined annual income and annuity value of net worth,[33] or a variable minimum annual income that depends on family size and rural versus urban location.[34] When wealth as well as income is taken into account there is a particularly striking effect on the prevalence of poverty among the aged; although the proportion of poor families among all families headed by a person 65 years or over remains high, it drops from 47 to 32 percent.[35]

The point is that before a determination can be made as to the target efficiency of a project the target group must be defined operationally, not merely conceptually. And our operational measures are always far from perfect. Thus, programs to help the poor have to struggle with the facts that low income in a single year does not "really" constitute poverty, that actual money income is likely to be greater than reported money income,[36] and that total income, including nonmoney income, often exceeds money income considerably. Yet reported money income for a single year may be the firmest information available for determining program eligibility and for defining operationally the target group for a program.

The choice of a precise specification of the target group may also have implications for the decision criteria discussed above. Aid in a form that is demeaning to one group may not be to another. Aid in a form that produces significant efficiency (incentive) effects for one group may not do so for another. Aid that would be spent in a specific way by one group may be spent quite differently by another. (This is germane to the consumer sovereignty/taxpayer sovereignty criterion.)

Having pointed up a few of the informational requirements for applying our six criteria in making choices among alternative governmental redistributional programs, we can now raise additional questions: (1) Do these decision criteria conflict? (2) Whether they do or not, if programs are differentially good according to each criterion, how much weight should be given to each criterion in order to come up with an overall "mark" or a ranking for each alternative redistributional program?

The decision criteria surely do conflict. Before investigating some of the conflicts, however, a brief digression will be useful in order to consider the case for lump-sum taxes and transfers as an ideal. In the tradition of modern theoretical welfare economics, such taxes and transfers have appeal because they cannot be avoided. Thus, they would produce no incentive effects and no changes in the market allocation of resources other than those which resulted from the fact that some people were better off financially and others worse off. But whether such a system would be desirable even if it were feasible is not clear. For its justification rests on the premises that the allocation of resources was efficient to begin with, and that no incentive effects are desirable. In an economy with widespread deviations from

[32] See Council of Economic Advisers, *Annual Report*, January 1964, ch. 2.
[33] B. A. Weisbrod and W. L. Hansen, "An Income-Net Worth Measure of Economic Welfare," *American Economic Review*, December 1968.
[34] Mollie Orshansky, "Recounting the Poor—A Five-Year Review," *Social Security Bulletin*, April 1966.
[35] B. A. Weisbrod and W. L. Hansen, "An Income Net Worth Measure of Economic Welfare," *op. cit.*, table 3.
[36] For further discussion see Edward C. Budd and Daniel B. Radner, "The OBE Size-Distribution Series: Methods and Tentative Results for 1964," *American Economic Review, Papers and Proceedings*, May 1969.

perfect competition in both output and input markets, and with numerous Federal, State, and local taxes and subsidies that also produce deviations between marginal production costs and market prices, there is little reason to believe that the economy is organized in a near perfectly efficient manner prior to any income redistributional program that is being considered. Consequently, it is difficult to accept the proposition that redistributions of income should be made only through lump-sum taxes and transfers.[37]

This issue of the economic efficiency (incentive effects) of alternative courses of Government action has been debated for decades in the context of efficient tax policy—in particular, regarding the choice of "direct" (income) taxation versus "indirect" (commodity) taxation.[38] Regardless of the reason for collecting revenue—whether to redistribute income or to finance a national defense force or for any other purpose—the basic efficiency argument is that taxes should be levied in a manner that minimizes their misallocative effects in the economy. As the analysis and debate developed it became clear that all of the taxes that are actually utilized—as contrasted with the theoretical lump-sum variety—do produce incentive effects. The income tax, for example, distorts the choice between work and leisure by taxing income from the former but not enjoyment from the latter. Moreover, given the preexisting distortions of relative prices, an additional "distorting" incentive effect might actually improve the allocative efficiency of the economy.[39] The choice of a redistributional mechanism that minimizes all incentive effects is not necessarily desirable. Lump-sum taxes and transfers, even if implementable, may not be suitable norms.

Whether lump sum or another form of redistribution is considered, it is likely that an effort will be required to identify the target group and devise a mechanism for aiding it. The conflict between the desire to minimize administrative costs (criterion 1), and to maximize vertical and horizontal target efficiency (criteria 2a and 2b) immediately develops. In order to achieve 2a—that is, to limit benefits to the target group—it is likely either that some members of the group will be missed—which conflicts with 2b—or that large administrative costs will be required, which conflicts with 1. In general, it appears that to perform well by any of these criteria requires a sacrifice in at least one of the other two.

It has been implied by the discussion above of criterion 3, incentive effects, that a redistributional program should avoid producing "adverse" incentive effects. One might argue that the program ought not to provide an incentive to escape poverty, for example, by a mechanism that involves high resource costs when a lower cost alternative does, or at least, could exist. Thus, it might be argued that if it is more costly in resources to alleviate poverty through compensatory education programs than by, say, cash transfers, then the latter mechanism should

---

[37] For an analysis of the general issue of what constitutes efficient behavior when some of the necessary conditions for efficiency do not hold, see R. Lipsey and K. Lancaster, "The General Theory of Second Best," *Review of Economic Studies*, 24 (December 1956).

[38] M. F. W. Joseph, "The Excess Burden of Indirect Taxation," *Review of Economic Studies*, June 1939, pp. 226–231; and I. M. D. Little, "Direct versus Indirect Taxes," *Economic Journal*, September 1951, pp. 577–584 (also in American Economic Association *Readings in Economics of Taxation*).

[39] For a survey of the debate over direct versus indirect taxation, see D. Walker, "The Direct-Indirect Tax Problems: Fifteen Years of Controversy," *Public Finance*, November 1955, pp. 153–177.

be used.[40] To provide an incentive for the poor to utilize compensatory education programs would be undesirable—because it would represent an inefficient allocation of resources.

Such an adverse (allocatively inefficient) incentive effect can be avoided, but to do so could conflict with the objective of avoiding demeaning forms of redistributional programs. Even if a compensatory education, manpower retraining or Job Corps program were not efficient in the sense of raising earning power by as much as the (marginal) cost of the program, they might raise incomes in a more self-respecting (less demeaning) manner than would a pure transfer-payment program. The point is that a program of investment in increased labor productivity—through training, education, relocation, etc.—might score weakly in terms of allocative efficiency while earning a high mark for its contribution to the recipients' self-respect.[41]

Of possibly greater significance is the conflict between the criteria of non-demeaning benefits and of vertical target efficiency. If all of the benefits of a redistributional program—especially a transfer program—go to the target group, such as the poor, the benefits are quite likely to be viewed as demeaning. If only poor people (or say, poor farmers) receive aid, then anyone who receives that aid is clearly identifiable—by himself and by others—as being poor. If, however, benefits go partly to the non-poor, then receipt of a benefit is likely to be less demeaning, for the (target group) recipients would not need to think of themselves as recipients of poverty aid, and outsiders would not be able to distinguish easily the poor from the non-poor recipients of the program's benefits.

Thus, the very process of pursuing the criterion of non-demeaning benefits tends to lead to the choice of redistributional programs that perform badly by the vertical efficiency criterion. For, by design, the tendency for well-targeted redistributions to be demeaning can be offset by deliberate non-targeting—by redistributing to some of the non-needy. The appeal of agricultural price supports may be understood in this context, as may programs to develop depressed regions such as Appalachia. While proportionately more farmers and residents of Appalachia are poor than is the case for the Nation as a whole, it is nonetheless true that the majorities of both groups are not poor, using Government yardsticks. Similarly, a Headstart program of compensatory education for the disadvantaged (the target group) is likely to be more appealing (less demeaning) if the groups include non-disadvantaged as well as disadvantaged youngsters.[42] Broadening the group of participants, however, runs directly into conflict with the vertical efficiency criterion and also with the horizontal efficiency criterion insofar as a budget constraint necessitates the exclusion of some members of the target group in order to accommodate some outsiders.

Another illustration of the conflict between the objectives of vertical efficiency and nondemeaning benefits involves public higher educa-

[40] T. Ribich, *op. cit.*

[41] B. A. Weisbrod, "Expenditures on Human Resources: Investment, Income Redistribution, or What?" in Joint Economic Committee, U.S. Congress, *Compendium on Human Resources*, vol. I (USGPO, 1968), especially p. 82.

[42] This is apart from the issue of whether the cognitive effectiveness of the program for the target group is influenced by increased heterogeneity of the student body.

tion.* There is in this country a well established system of private higher education, but costs of attendance have been too high to permit large numbers of youth from low income families to attend. Partially in response to the essentially income-distributional objective of broadening access to higher education, State supported colleges and universities have been established. The criterion of vertical efficiency would dictate that public subsidies in higher education be restricted to the needy.[43] Either the nonneedy would be excluded from the State universities, or else the nonneedy would have to pay a larger tuition, thereby losing some or all of the public subsidy.[44]

In practice, publicly supported institutions of higher education have neither prohibited entrance by the nonneedy nor have they charged differential tuition rates according to ability-to-pay.[45] They have preferred, instead to keep tuition as low as possible for all students, regardless of ability to pay. As a result, large subsidies have gone to the affluent via the higher education system. As noted earlier, a study of public higher education in California revealed that the affluent benefited disproportionately from the public subsidies because they were more likely to attend a public college, more likely to attend a high-subsidy school and more likely to remain until graduation.[46] The following table shows the estimated distribution of subsidies by level of family income. The difference between the pattern indicated and the pattern that would exist if vertical efficiency were being maximized—all subsidies going to the poor—is obviously great.

	California families without children in California public higher education	California families with children in California public higher education—children in—		
		Junior college	State college	University of California
Average income	$7,900	$8,800	$10,000	$12,000
Average subsidy	0	$1,050	$3,810	$4,870
Subsidy as percent of income	0	12	31	41

Source: W. Lee Hansen and Burton A. Weisbrod, "Benefits, Costs, and Finance of Public Higher Education" (Chicago: Markham Publishing Co., 1969), ch. IV.

A number of justifications could be offered for this vertical inefficiency. Perhaps low-income students learn more when high-income students are present; if so, this externality would support a subsidy to the nonneedy, on economic efficiency grounds. More important, possibly, is the desire to avoid the demeaning nature of the higher education subsidy that would exist if the subsidy were available only to those who "passed" a test of financial plight.

Turning to the criterion of flexibility suggests that as among alternative redistributional programs the choice should be made in favor

[43] Other considerations such as external benefits might have different policy implications.

[44] For further discussion see Alice M. Rivlin, *The Role of the Government in Financing Higher Education* (Washington, D.C.: The Brookings Institution, 1961), especially pp. 144–145.

[45] In 1968 Michigan State University did institute a flexible tuition scheduled under which tuition varied somewhat with family income, from a low of $118 per term for students with family income of $11,800, or less, to a high of $168 per term for students with family income of $16,800 or more.

[46] W. L. Hansen and B. A. Weisbrod, *Benefits, Costs, and Finance of Public Higher Education, op. cit.*

*Further discussion of this issue is found in the paper by Brandl in vol. 3 of this collection.

of maximum visibility and recognition as a subsidy.[47] For unless subsidy programs are recognized as such, there may be little likelihood of their being altered or dropped in response to changing circumstances.[48] Visibility of subsidies conflicts directly, however, with the non-demeaning-benefits criterion.

The conflict between the criteria of consumer and taxpayer sovereignty—involving who should decide the use to which benefits from redistributional programs may be put—was alluded to above and need not be belabored here. It is noteworthy, though, that the conflict between these criteria reflects two rather distinct views of "society." The taxpayer-sovereignty criterion implies a view of society not as a unit but as a bifurcated organization consisting of "we"—the "haves"—and "they"—the "have-nots." From this standpoint, income redistribution is not an activity which benefits "society," but rather one that benefits "taxpayers." Normally, everyone decides for himself how to spend his income, subject only to modest legal constraints that apply, however, to everyone. In invoking the taxpayer-sovereignty criterion, however, taxpayers appear to be saying that they should decide not simply to whom income will be transferred but how the funds will be spent by (or for) the recipients.

Granted that the criteria for choosing among redistributional programs are frequently in conflict, the next question is what relative weights should be given to each of the criteria—how much is "society" willing to sacrifice in terms of each criterion in order to obtain more in terms of each of the other criteria? If, for example, a retraining program for the long-term unemployed increased the (present value of) lifetime earnings of trainees, but only by an estimated 95 percent of the program cost, would this be "preferable" to a pure transfer program? If yes, if it is worth a 5-percent loss on investment in order that people can earn their way out of poverty, is it worth a 50-percent loss? The answer obviously depends on value judgments. So far, little research has been done either on what prevailing value judgments are with respect to these multiple goals of social policy, or on what the policy implications are of alternative "tradeoff" rates.[49]

IV. Conclusion

The income-redistributional effects of governmental actions pervade most, if not all aspects of governmental economic activity. Some activities are intended to be redistributive (e.g., welfare payment or a negative income tax); other activities may or may not have such an intent but are clearly recognized to have that effect (tax policy) There are, however, many governmental programs that ostensibly have little or nothing to do with redistributional policy, yet they do bring

[47] In the context of the use of the Federal tax system to provide subsidies a similar point has been made by Stanley Surrey, former Assistant Secretary of the Treasury. See "What Is the Impact of Those Tax Breaks?", in *Business Week*, Feb. 1, 1969, p. 62ff.

[48] James Bonnen, in commenting on the present paper, noted that over the past 8 years the agricultural program mix has shifted away from price supports toward direct payments. This, he points out, is a movement from a not very visible subsidy to an extremely visible one that now constitutes over three billion dollars a year in direct payments. He has speculated as to whether the Kennedy-Johnson Department of Agriculture deliberately shifted the program emphasis so as to place the programs in greater political jeopardy—to exert indirectly some constraining influence that the Department itself was unable to exercise.

[49] See, however, the citations in footnote 18, p. 187.

about distributional effects. Economic stabilization policy and resource-allocation policy—ranging from education to highway to flood-control programs—have direct as well as indirect or secondary consequences for the distribution of income. Recognizing the widespread effects of Government actions on the income distribution is the first step in learning to anticipate these effects, taking them into account, and deciding how to deal with them.

If government has such widespread effects—some intended, others not—on the income distribution, it is important that an effort be made to establish criteria by which to judge the desirability of alternative means for redistributing income. Thus, part II of this paper presented six criteria by which all income-redistributional effects of government actions might be assessed. Conflicts among a number of these criteria were considered in part III, but, in general, not resolved. The principal point is that for a variety of reasons, people—both taxpayers and program beneficiaries—care not only about how much income is transferred, but about how it is transferred and to whom other than the intended target group it is transferred. Further study of income distributional considerations in the context of government decisionmaking is warranted in order to develop more rational public policy. Needed particularly is consensus on criteria for choosing among programs that have a significant redistributional *effect*—whatever their *intent*—and on the relative importance of each criterion.

Part II

Institutional Factors Affecting Efficient Public Expenditure Policy

Section A

The Making of Public Expenditures Decisions and the Implementation of Public Policy

THE ROLE OF INCENTIVES, PENALTIES, AND REWARDS IN ATTAINING EFFECTIVE POLICY

BY CHARLES L. SCHULTZE

Charles L. Schultze is Senior Fellow at the Brookings Institution and Professor of Economics at the University of Maryland.

He here discusses the important role of incentives in designing and implementing effective public policy. Dr. Schultze observes that because the attainment of national objectives creasingly depends "on the joint action of many independent decision makers, private as well as public; and . . . because the growing complexity and geographical diversity of public programs requires decentralized decision making within the public sector itself" some structure of incentives, "rules, organizational structures, performance measures, and penalties and rewards" is necessary to ensure that the results or public programs coincide with their original objectives. He examines the different stages of public expenditure policy at which incentives can play a major role: "incentive considerations should enter into the formulation of public objectives, the design of public programs and the allocation of budget resources, equally as much as in program execution." The different types of incentive problems which may be encountered in the consideration of public policy are analyzed as well as the possibility of channeling private actions toward public objectives by removing negative incentives, supplying positive incentives, and properly applying Federal user charges. Dr. Schultze argues the importance of encouraging efficiency within public programs by providing appropriate market-type penalties and rewards.

Some of the major problems of concept and political feasibility which will be confronted as incentives are employed in formulating effective policy are also discussed. The definition and measurement of public sector outputs is cited as one of the major difficulties. While application of incentives to individuals as well as to institutions is essential to assure program effectiveness, further study of the feasibility of applying such incentives is needed. It is suggested that the nationwide functional basis of the Federal budget should be accompanied by a complementary system of regional budgeting to chance political incentives at the local level. Dr. Schultze stresses that "the problem of incentives is not a discrete and separate part of public expenditure theory relating to how programs are carried out, but an aspect of social behavior which should be taken into account at every stage of public policy formation."

Introduction

The traditional theory of public expenditures concerns itself primarily with such questions as the proper role of Government, the evaluation of the benefits and costs of public programs, and the optimum allocation of national resources to the various public programs.* It deals with *what* ought to be done and *how much* of our resources ought to be devoted to doing it.

In a similar vein, designers of social legislation in the past decade have concentrated primarily on what ends should be sought. Most Federal programs are characterized by two features—(i) a *grant-in-aid* of funds to State and local governments or private institutions, conditional upon the submission and approval of (ii) a *plan* setting forth the objective for which the funds will be used.

*Further discussion of this issue is found in the papers by Steiner and Arrow in this volume.

But objectives, plans, and budgets are not synonymous with actions and results. Promises are not performance. The adoption of an urban development plan does not guarantee that urban investment will fit the plan. The establishment of water quality standards does not clean up pollution. The formulation of a model city neighborhood plan will not itself produce a model neighborhood. When a particular public program in health or education or pollution control or urban development must be jointly carried out by a complex governmental structure and a host of private decisionmakers, it becomes crucial that those who execute the program have incentives or inducements to act in directions which are consistent with the objectives of the program.

It is becoming increasingly important to recognize the difference between objectives and plans on the one hand and performance on the other. The great expansion of public spending in recent decades has primarily been devoted to programs which involve the private as well as the public sector of the economy. While programs which produce purely public goods (like national defense) or simple income redistribution (like veterans' pensions) are still important, the cutting edge of the recent expansion in Government activities involves such matters as urban rehabilitation, control of air and water pollution, the provision of medical services, on-the-job manpower training, the development of depressed areas, and the like—programs in which public actions and decisions cannot alone determine results. Moreover, even where programs are predominantly confined to the public sector—as is the case with elementary and secondary education—the major instrument of Federal policy is not direct action but joint action with State and local governments through the grant-in-aid mechanism.* To a growing extent, therefore, public program performance depends upon the behavior of a large number of *independent* decisionmakers, public and private. Actions cannot be commanded. There is no hierarchy of officials in a single line of command who can be directed toward a set of predetermined objectives. In such cases the careful specification of plans and objectives by a public agency will not suffice to guarantee effective programs. The program must also be explicitly designed to provide incentives or inducements for the relevant decisionmakers outside the public agency to act in directions which are consistent with program objectives.

Even within the public sector itself the problem of incentives is taking on growing importance. Social goals have become more ambitious, program objectives more complex. The model cities program is infinitely more difficult to execute than the disbursement of veterans benefits or the management of an agricultural conservation program. Designing, procuring, and operating a strategic nuclear weapons systems is a far cry from buying and maintaining the mounts for a cavalry regiment. Moreover, one of the chief characteristics of recent social legislation is that it seeks to realize national objectives in thousands of diverse communities across the land. Highly centralized managerial system cannot cope with the sheer number and diversity of the day-to-day decisions which have to be made. Decentralization of decisionmaking is not only desirable, it is unavoidable. But decentralized decisions

*Further discussion of this issue is found in the paper by Mushkin & Cotton in this volume.

should be compatible with central goals. And that in turn requires a system of rules, organizational structures, performance measures, and penalties and rewards, which induces decentralized decisionmakers in public programs to act in ways consistent with overall program plans and objectives.

For two reasons, therefore, the problem of incentives deserves particular attention in the formulation of public expenditure policy: *first*, because national objectives increasingly depend for their realization on the joint action of many independent decisionmakers, private as well as public; and *second*, because the growing complexity and geographical diversity of public programs requires decentralized decisionmaking within the public sector itself.

Incentives have a role to play in public expenditure policy which goes well beyond the problem of program execution. Incentive considerations should enter into the formulation of public objectives, the design of public programs and the allocation of budget resources, equally as much as in program execution. Or to put the matter another way, the problem of incentives is not a discrete and separate part of public expenditure theory relating to how programs are carried out, but an aspect of social behavior which should be taken into account at every stage of public policy formulation.

The first section of this paper considers the role of incentives at the various stages of public expenditure policy—program design, budget allocation, and program execution. The next part of the paper discusses the various types of incentives which appear to be relevant in formulating public policy. A final section examines some of the theoretical and political problems which arise in applying incentive concepts to public expenditure programs.

I. The Role of Incentives at the Different Stages of Policy Formulation

> Webster defines the word "incentive" as: "something that incites or has a tendency to incite to determination or action;" looking further, we find the word "incite" defined to be: "to induce to exist or occur."

In the case of public programs, considerations of incentives arise when we concern ourselves with the problem of inducing individuals and groups, either public employees or private decisionmakers, to undertake actions which produce results desired by public policy. There are three stages of public expenditure policy to which considerations of incentives apply: program design, program execution, and budget allocation.[1]

1. PROGRAM DESIGN

Many of the Federal Government's most important expenditure programs involve—or at least ought to involve—a mixture of public and private actions. More specifically, in perhaps the majority of public programs, the public sector is not producing a pure public good, but is attempting to take account of external costs and benefits in the production of private goods. Public programs seek to modify, in quality

[1] This paper does not, except peripherally, deal with public regulatory policies, where questions of incentives also arise.

or quantity, the outcome of private production and investment decisions. Urban development programs, air and water pollution controls, and flood protection are examples. Yet too often, by concentrating solely on the public sector side of a joint public-private problem, public programs produce distinctly inferior solutions.

Flood protection is a case in point.* Since the Flood Protection Act of 1936, the Federal Government has spent some $7 billion on flood protection projects. Expenditures on such projects currently run about $500 million per year with an additional $100 to $150 million per year spent on disaster relief to flood victims. While estimates of national flood damages are very approximate, they exceed $1 billion a year and are increasing. Extensive economic and engineering literature has been developed on the optimum design of flood control projects and on the techniques of benefit-cost measurement for such projects.

National policy toward flood losses has been quite straightforward—build flood protection and prevention works primarily at public expense and assist States, localities, and individuals to recoup against large flood losses. Where it could be shown that potential projects would prevent losses whose value exceeded the cost of the project, then those projects became eligible for Federal financing, subject to the overall availability of funds in the Corps of Engineers budget and the normal vicissitudes of pork barrel politics.

But, as a matter of fact, public policy ought not to be expressed solely—or even primarily—in terms of criteria for the construction of public works for flood prevention. Rather it should be formulated in terms of encouraging rational use of flood plain lands. We should be seeking a policy which induces public and private investment in the flood plains only if the advantages of locating there are greater than those of alternative sites by an amount which exceeds the expected value of flood damages or the cost of preventing those damages. The present policy, which concerns itself almost solely with public projects, not only fails to consider the establishment of incentives for economic private investment in flood plain lands, it sets up a series of monetary and political incentives which induce distinctly uneconomic investment decisions.

Once the flood plain is developed, the standard cost-benefit calculation will often show that the construction of flood prevention or protection works is worthwhile in terms of expected damage avoided. However, in all too many cases the preferred alternative would have been a much less intensive development of the flood plain land or no development at all. In other words the differential advantage of the flood plain over the next best alternative is worth less than either the cost of flood protection works or the expected value of damage. Since States, local communities, and individual beneficiaries typically contribute only a fraction of the cost of Federal flood protection works (ranging from 5 to 60 percent and averaging 25 percent) there has developed a set of incentives for uneconomic use of flood plain lands. Development occurs in flood plains. Either in response to or in anticipation of floods, strong and often successful pressure is brought to bear for Federal flood protection. In many cases, flood-proofing of individual buildings would be much cheaper than building

*Further discussion of this issue is found in the paper by Knetsch in vol. 3 of this collection.

flood control public works. But the costs of floodproofing are borne by the individual owner; the cost of public works is not. Once Federal works are constructed, further development occurs, beyond the protected areas. The resulting encroachment on the flood plain itself raises unexpected flood heights, increases the expected flood damage to prior investments and leads to still further flood protection works. Studies of flood plain use "show that some flood plain enroachment is undertaken in ignorance of the hazard, that some occurs in anticipation of further Federal protection, and that some takes place because it is profitable for private owners even though it imposes heavy burdens on society." [2]

In earlier years most flood control projects were justified on the basis of protecting existing developments. More recently, however, an increasing proportion of flood control projects have been justified on the basis of protecting land for future development. Most often the economic and engineering surveys upon which construction authorization is based do not examine alternative sites for the projected development. As a consequence, benefits are calculated on the basis of the absolute value of the site as a location for the potential development, rather than its differential value compared to the next best alternative. This tends to accelerate still further the "cycle of losses, partial protection, further induced (though sub-marginal) development, and more unnecessary losses." [3]

As the Task Force on Federal Flood Control Policy pointed out, an effective approach to national flood damage policy would "alter the price signals received by potential flood plain developers." The full costs of flood plain occupancy would be shifted to prospective occupants through an occupancy charge equal to estimated annual damages plus any external costs which occupancy causes others. These payments would, in turn, be used to compensate those suffering flood plain damages. Flood control works would lower the annual occupancy charges and the costs of the necessary public works would be charged to the beneficiaries, whose annual costs had been reduced.

As initial steps in this direction, the report recommended the careful experimentation and development of a flood risk insurance program, with premiums on future investment in flood plain land related to flood damage risk. Ultimately such insurance would be a requirement for any investor in flood prone lands in order to be eligible for Federal loans, loan guarantees, flood protection investment, or other similar assistance. The development of such a program must necessarily be gradual, since premiums seriously out of line with actuarial risk would invite uneconomic location and heavy costs. As a corollary to this recommendation, the report urged a sharply expanded program for determining flood hazards, flood frequency, and unexpected flood damage. Without such information economically meaningful premiums cannot be developed—and, to stress again—an insurance program with a poor premium structure is worse than no program at all. In 1968 the Congress enacted a flood insurance program. While premium rates on existing residential dwellings and small business establish-

[2] *A Unified National Program for Managing Flood Losses,* a report by the Task Force on Federal Flood Control Policy. (Printed by the Committee on Public Works, Aug. 10, 1966), p. 11.
[3] *Ibid.,* page 12.

ments will be subsidized (up to a specified dollar value limitation) premium rates on all new construction and major improvements will be set to cover actuarial risks.

In addition to flood insurance the task force report recommended the development of a cost-sharing formula for flood control works which would more nearly assign the costs to the beneficiaries.* Should an insurance program be developed, with premiums related to risks, the existence of a heavily subsidized means of reducing premiums would continue to generate excessive development in the flood plains. And requiring beneficiaries to pay for the cost of public works would tend to neutralize the present bias against investment in the flood-proofing of individual structures.[4] Where flood control projects are justified on grounds of benefits from future development, as opposed to protecting existing investments, the report urges that only *net* locational benefits be taken into account—that is, only the excess value of flood plain location over the next best alternative be taken into account.

The point of this summary of flood control policy is that the problem of incentives must be considered in the initial specification of public policy objectives in those cases where joint private and public action is involved. And joint action is likely to be involved in a high proportion of cases where the basic objective of public policy is to take into account the external benefits or costs which arise in private decision-making. And "incentive-oriented" point of view would recognize that present flood control policy is deficient on at least three counts:

It fails to provide incentives for private decisionmakers to consider flood hazards sufficiently in their investment decisions.

By its lack of user charges, it positively encourage uneconomic investment coupled with political pressure for subsidized flood protection works.

By its lack of user charges it also sets up a false set of incentives to minimize floodproofing and maximize flood protection.

Other examples are numerous, in which failure to consider incentives leads to an inferior specification of objectives and inferior program design. In the field of water-pollution control, for example, public policy emphasizes the subsidized construction of waste-treatment plants, dams for low-flow augmentation, and the separation of storm drains from sewers, as means of treating pollution once created. But it generally fails to consider means of altering the price signals received by pollutors through the mechanism of user charges and effluent charges.** Through such charges, industrial pollutors would be assessed the social and economic cost of pollution, and in many cases would find it profitable to change their internal processes to reduce the amount of pollution they create. In general, it is cheaper to improve the quality of our streams by combination of prevention and treatment than by treatment alone. But because the private sector is primarily responsible for prevention and the public sector for treatment, public policy excessively concentrates on the latter aspect. And to the extent it does deal with the prevention aspect of pollution

[4] See footnote 13, p. 217.

*Further discussion of this issue is found in the papers by Krutilla and Milliman in this volume.

**Further discussion of this issue is found in the papers by Davis & Kamien, and Kneese & d'Arge in this volume.

control, it does so by attempting to enforce, through the police power, a set of water quality standards rather than providing economic incentives to individuals which would induce them, in their own interests, to take action to improve water quality. Again, the question of incentives arises in the specification of objectives and the design of programs—not merely in program execution.

More generally, public policy must often concern itself not with the provision of public goods which can only be handled by the public sector, but with the problem of external costs and benefits in the private sector. In such cases market prices and costs do not reflect "true" social benefits or costs. As a consequence purely private decisions do not produce desirable results. Pollution is a cost to society which the pollutor does not bear. The full costs of flood damages are often not foreseen and even less often borne by those who locate in flood plain lands. As a consequence, public action is needed. But that public action need not be simply the provision of public facilities (waste treatment plants or flood protection works) to offset the economic losses caused by private actions. Rather the objectives of public policy, in such cases, should include a modification of the "signals" given and incentives provided by the marketplace so as to induce private actions consistent with public policy. Excessive concentration upon the purely public part of public policy may result in poorly specified objectives and ineffective programs.

2. PROGRAM EXECUTION

In the 1920's, expenditures of the Federal Government outside of the traditional functions of defense, post office, veterans' benefits and interest on the debt represented less than 1 percent of GNP. In the budget for fiscal 1969, such expenditures will equal 10 percent of GNP. The Federal Government today manages large enterprises—the space program, the Atomic Energy Commission, the Forest Service—whose very size and complexity require highly decentralized operations. The establishment of objectives and plans at the top level of the agency concerned does not guarantee that the vast number of decisions which must necessarily be delegated to subordinate officials will result in effective and efficient program execution.

The problem of providing incentives for effective program execution is closely related to the problem of devising measures of performance for subordinate decisionmakers. It is impossible to provide incentives without knowing what to reward and what to penalize. Two major consequences flow from failure to provide performance measures related to program objectives. The first is the growth of detailed regulations which rigidly specify what is "acceptable" behavior by subordinate decisionmakers. Standard contract provisions multiply, and are required to be included in all contracts regardless of their suitability to a particular situation. Tables of organization are centrally established and carefully monitored. Elaborate procedures are developed to control the purchase of supplies, the use of long-distance communications, travel, and the like. Since subordinate decisions cannot be controlled by judging them in terms of their effect on output, they are controlled by a rigid specification of inputs.

The second consequence of failure to provide appropriate per-

formance measures, is that individuals and institutions often become avid risk averters. Overall success cannot be recognized, but individual "mistakes" can be singled out for punishment. A few examples will help. In 1966 inflationary pressures mounted rapidly in Vietnam. As one means to counteract those pressures, the Agency for International Development launched a commodity-import program, designed to soak up some of the excess purchasing power with American-furnished commodities. Two options were open in running the program:

A. A carefully controlled license program, in which every possible step was taken to insure that import certificates did not fall into the hands of black marketeers or into the hands of the Vietcong. One consequence of such a program would be a mountain of redtape and a very slow trickle of imports into the Vietnamese commodity markets.

B. A less tightly controlled program whose main objective was introducing rapidly a large volume of commodity imports into the country. One early foreseeable consequence of such a program was that a significant proportion of imports would end up with black marketeers and some part with the Vietcong.

Given its overall objective, AID quite properly chose the second course of action. And inevitably, a year later, it found itself subject to sustained and violent attack in the Congress for the easily identifiable consequences of its policy, namely the appearance of AID-financed imports in the hands of black marketeers and small amounts in the possession of the Vietcong. The obvious reaction of program operators to the performance criteria which implicitly underlay this attack, is in the future to accept a lower expected value of program accomplishment in exchange for a smaller proportion of "mistakes."

A similar situation commonly arises in the loan programs of the Federal Government. Many of these programs, rightly or wrongly, have the supposed objective of providing loan capital to small enterprises which are too risky for investment by commercial lenders. The Small Business Administration is a case in point. Measures have not been developed, however, which can be used to judge the performance of various regional loan offices in terms of overall program objectives. Defaulted loans, on the other hand, are easily identified, and a significant default rate is sure to invite congressional questions. Loan officials, therefore, tend to avoid risky loans. As a consequence, far from meeting their original objectives, the programs end up, in many cases, simply in making loans of commercial quality at less than commercial rates.

The problem of providing incentives for effective program execution is even more difficult when the Federal Government itself does not directly operate the program, but provides grants-in-aid to other governments or to private organizations.* The major growth in Federal expenditures in recent years has been devoted to complex social programs in such fields as health, education, manpower training, urban development, and the like. These programs deal with problems which, while national in scope, require very particular solutions in

*Further discussion of this issue is found in the paper by Mushkin & Cotton in this volume, and the papers by Grosse, Brandl, Mangum, and Ross in vol. 3, of this collection.

thousands of individual communities. The Federal Government provides the financing and requires the submission of plans which specify the objectives for which the funds are to be used, but it does not have direct responsibility for operating the programs. The current widespread pressure for decentralization and participatory democracy arises, in part at least, from a recognition that highly centralized decisionmaking is virtually impossible in these programs. Yet without meaningful measures of performance and effective incentive systems, is it possible to decentralize decisionmaking while still making progress toward the generally accepted high priority national goals toward whose realization the Federal grant programs were originally established? I believe that in selected cases, at least, an incentive approach can help to resolve the national goal versus decentralization dilemma.

3. BUDGET ALLOCATION

It is not customary to think of budgetary allocation from the standpoint of incentives. We tend to view the budget allocation process as one in which the Government seeks to adjust program levels so that equal marginal returns are realized from the last dollar spent on each program. Massive problems of evaluating benefits and costs confront decisionmakers in trying to approach this ideal result. But in what sense do incentive considerations arise? In two ways, I believe.

In the first place, with few exceptions, the Federal budget is developed on a nationwide functional basis—education, health, defense, space, manpower training, and so forth. There is no mechanism, however, by which tradeoffs among alternative Federal programs are possible at the local or regional level.* A mayor or Governor has almost no means of negotiating a tradeoff between a submarginal flood control project and a highly needed hospital; between an urban freeway and a waste treatment plant. Hospital, highway, pollution control, and water resource budgets are decided nationally. This fact has two major consequences:

First, it sets up incentives for local communities and their Congressmen to lobby in the Executive and the Congress for almost any project they have a chance of getting. Incentives are established for inefficient political bargaining; the relevant set of tradeoffs has no way of appearing in the bargaining process.

Second, many functional decisions involve both gainers and losers. Water pollution control programs help downstream communities and often hurt upstream. Location of low-income subsidized housing in suburban communities is often viewed by that community as entailing positive costs. Concentration of assistance to depressed areas in potential growth centers is, quite naturally, viewed with some hostility by hinterland counties who receive no immediate help from this approach, but whose poverty may be much greater than the community actually receiving the assistance. As a consequence, losers are often able to veto potentially high priority programs. But considering the entire range of potential Federal programs, it may often be possible to compensate losers in one functional program with perfectly appropriate assistance of another kind. Yet so long as budget alloca-

*Further discussion of this issue is found in the paper by Schmid in this volume.

tions are made solely on a nationwide functional basis, there is no mechanism by which compensatory tradeoffs can be negotiated in particular cases.

Third, theory tells us that funds should be allocated among various Federal programs so that the benefits from the marginal dollar in each program are the same. But many Federal programs have objectives which are reached through particular projects in individual communities scattered across the Nation (for example, hospital construction grants, flood control projects, and so forth). In such cases an evaluation of the merits of one program relative to another is much more difficult to make on a national basis than it is in a specific community. The relative desirability of allocating more Federal funds to hospital grants, flood control, or to slum rehabilitation can often be determined more reasonably for New York City than it can for the Nation as a whole. This is not to say that considerations of national objectives are unimportant in making such a determination. Rather it is to say that, at the margin, allocation of Federal funds among various programs can often be done better on a regional or community basis than on a nationwide functional basis.

In short, for certain types of programs purely functional budgeting sets up political incentives for inferior bargaining, suppresses incentives for a kind of compensatory bargaining which might hold great promise, and fails to provide for the kind of relative benefit comparisons which efficient allocation demands.

For these reasons experimentation with a limited form of *regional budgeting* could be very worthwhile. Tentative functional budgets could be drawn up on a national basis, just as they are now. But in each locality, Governors and mayors could be given the right, up to some limit, to propose reallocations among particular Federal aid funds flowing into their own jurisdiction. They might propose, for example, an increase in funds for education and a decrease in highway grants. In effect, the final allocation of Federal budgetary funds would arise out of a joint set of considerations—national allocations based on nationwide objectives, modified by reallocations based on conditions and preferences in particular communities. Functional budgeting would be supplemented where appropriate and feasible by regional budgeting. Such an approach has much to recommend it, but it is not without its problems. I shall return to a discussion of those problems, and some suggested means of overcoming them, at a later point in this paper.

Another type of incentive problem in budget allocation arises when major elements of cost are not charged to the decision unit responsible for making decisions which involve those costs. Examples range from major areas of program policy down through detailed management decisions. The U.S. merchant marine is supported with a massive subsidy program on the primary justification that national security demands it. Yet if the costs of this decision were charged to the Defense Department, it is quite likely that defense strategists would opt for a smaller subsidy program and for use of the savings in other defense purposes. Whatever the result, the appropriate program tradeoffs could at least be considered. On a smaller scale, if military personnel are not charged to the budget of the commander of a military

installation, he has every incentive to substitute this free good for those other inputs which do enter his budget.

More generally, budgetary structure affects the incentives of decisionmakers. Programs whose costs do not enter the budget of the decisionmaker, being a free good, are easily recommended. Free resources are overused. In other cases desirable programs are blocked because the budgetary structure provides no means by which "losers" in a particular decision can be compensated. As a consequence they have no incentive to agree to a reasonable compromise. In still other cases State and local officials have no incentive to develop and negotiate an effective program for the use of Federal assistance, because the excessively functional character of Federal budget decisions leaves them no leeway to negotiate transfers among functions.

So far we have examined the role of incentives under a set of classifications based upon the various stages of public policymaking—the specification of objectives, program design, program execution, and budgetary allocation. Further insight can be gained by considering a classification based on the various types of incentive problems which arise in the consideration of public policy.

II. The Various Types of Incentive Problems

1. INCENTIVES DESIGNED TO CHANNEL PRIVATE ACTIONS TOWARD PUBLIC OBJECTIVES

A. Removing or modifying current incentive structures which lead to actions with large social costs or prevent the achievement of social objectives.

If we review public programs designed to modify private behavior in socially desirable directions, we find that failure to consider the problem of incentives has often led to very ineffective or inefficient solutions. In some cases ineffective solutions have emerged because public policy has attempted to impose on the private decision system detailed plans which require actions running directly counter to those called for by the existing system of private rewards and penalties. Enforcing the plans by use of the police power may often be far less effective than eliminating or modifying the private incentives which run counter to the plan.

In other cases, public programs have specific subsidy structures which themselves set up highly inefficient responses. Approximately the same subsidy objective could be reached with a subsidy designed to encourage efficiency.

Let me cite some examples:

Both the Federal Government and city governments have struggled for years with the problem of planning urban development. Almost every form of Federal assistance to municipalities is conditioned on some kind of a planning requirement—comprehensive plans, functional plans, planning processes, social renewal plans, workable programs, and so on down the litany. But, all too often, the plans are more breached than observed. And in large part this occurs because the system of rewards and penalties at work in connection with urban investment is not merely neutral to but runs precisely counter to the goals and objectives of the plan.

One major aspect of urban development plans is zoning for differential intensity or other differences in use.* The purpose may be esthetic, it may be to control development along lines conducive to efficient urban transportation, or it may have other purposes. But the principal characteristic of zoning is that it reduces the potential rent on specific parcels of land below the rent which could be earned without zoning restrictions. Under the most favorable conditions, this characteristic would make zoning hard to enforce. But this problem is substantially increased by the present tax system. The return to investment in physical improvements will not generally vary with zoning changes—the landowner will capture the gains from such changes. And while physical improvements yield an annual return subject to normal tax rates, the potential rewards from securing a change in zoning can be realized as capital gains, and will be taxed at much more favorable rates. As a consequence, the energies and capital of real estate developers are channeled into land speculation and into massive efforts to secure favorable changes in zoning codes. Other things being equal, this kind of activity yields returns which pay less than half the tax securable by investment in physical improvements. Small wonder that "year 2000" plans in most metropolitan areas quickly succumb to the relentless pressure of land developers.

In a similar vein, the owner of slum property hoping for a rise in land prices, has every incentive when faced with a tradeoff between improving his property and extending his holdings of existing property to favor the latter. Improvements yield a return subject to normal tax rates. The yield from the acquisition of additional property can be taken as capital gains. The availability of capital gains tax treatment on capital gains from the sale of land, shifts the margin of tradeoff between improvements and extension of holdings in favor of the latter.[5]

More generally, the availability of highly favorable tax treatment for those who speculate in land, tends to work counter to most of the objectives contained in urban plans. A change in the tax system would not itself automatically channel urban investment in socially desirable directions—there are a host of other factors which influence such investment. But certainly the system could be made more neutral with respect to planning objectives, rather than being highly counterproductive as it now is.

The system of payment by which hospitals are reimbursed under medicare, medicaid, and most private insurance plans represents another example of an existing incentive system which produces highly undesirable results.

During the past several years, hospital costs have been rising at a highly accelerated rate. In 1967, hospital costs rose by more than 16 percent for the second year in a row. Average per diem hospital costs now approach $60, and have been projected to rise to $100 within 5 years.[6] If means could be found to reduce the rate of increase in hos-

[5] The tax treatment of depreciation on buildings complicates matters, but does not invalidate the basic proposition. In fact the particular rules relating to the tax treatment of existing buildings compared to the treatment of major improvements tends to discriminate in favor of the former and against the latter. See Richard E. Slitor, *Tax Approaches to Low Income Housing Problems:* A study prepared for the National Commission on Urban Problems.

[6] *Secretary's Advisory Committee on Hospital Effectiveness, Report.* U.S. Department of Health, Education, and Welfare, 1967; p. 10.

*Further discussion of this issue is found in the paper by Ross in vol. 3 of this collection.

pital costs by only 2 percent per year, the resulting savings would amount to $5 billion annually by 1975. Not only have per diem costs at hospitals been rising, but hospital utilization, in terms of hospital days per thousand population, has also been increasing rapidly. Per capita utilization of hospital services rose by 47 percent over the 1955–65 decade.[7] There is an accumulation of evidence that some part of this rise reflects an excessive use of hospital services relative to patient's needs. If a 25-percent reduction in hospital utilization rates could be effected, without reducing the quality of patient care, an annual savings in medical costs of $7.5 billion could be realized by 1975.[8]

The Federal Government's medicare and medicaid programs now pay for the hospital care of a significant part of the population. Built into these programs is the concept of fully reimbursing hospitals for the costs associated with medicare and medicaid patients. Through these programs the Federal Government can have a major impact on hospital costs, both directly and through its influence on the practices of Blue Cross and commercial insurance carriers.

At the present time Federal reimbursement formulae undoubtedly contribute to hospital inefficiency. Essentially each hospital is reimbursed by the Federal Government for the "reasonable costs" of delivering services to patients under medicare and medicaid programs. Payment is matched to the individual costs of each hospital. There are virtually no incentives for efficiency. Any savings from more efficient operations result in lower Federal payments; any increased costs are fully passed on. To the extent that larger staffs bring prestige and promotion, there are positive incentives for inefficiency. Moreover, since policies in most hospitals are controlled by the physicians serving it, and since the hospital provides, in effect, a free workshop for those physicians, there are powerful incentives to upgrade the workshop, when the costs are reimbursed, insofar as most patients are concerned, either by Government program or private insurance carriers.

All of the evidence indicates that there is substantial room for improving efficiency in the delivery of hospital services. For example, recent data show that average per diem costs in voluntary short term teaching hospitals in New York City ranged from a low of $50 to a high of $87. The range among 42 New York City community hospitals was $34 to $61. Among a group of 12 hospitals carefully chosen for high quality care, costs per patient-day ranged from $46 to $96, after wage scales had been adjusted to a common basis.[9] The American Hospital Association compiled operational data on 431 hospitals of different sizes throughout the Nation and found substantial economies of scale: small hospitals produced an average 3.7 laboratory procedures per man-hour, large hospitals averaged 8.9; small hospitals annually served 1,800 patient-days per one administrative employee, large hospitals 4,1000; small hospitals produced 3.9 meals per man-hour, large hospitals 5.9.[10] While some of this huge variance in costs among hos-

[7] Derived from data in the *Report of the National Advisory Commission on Health Manpower*, U.S. Government Printing Office, November 1967; app. I, table 4. The per capita increase in average daily hospital beds used (11 percent) was combined with the increase in the deflated value of service per patient day to arrive at the 47-percent figure used in the text.

[8] *National Advisory Committee on Health Manpower, report*, vol. 1, p. 68.

[9] All of these data are cited in vol. I of the *Report of the National Advisory Commission on Health Manpower*, p. 55.

[10] Cited in Herman M. Somers and Anne R. Somers, *Medicare and the Hospitals*, Brookings, 1967.

pitals many indeed be associated with differences in the quality of care, much of it is undoubtedly traceable to differences in efficiency.

A number of schemes have been suggested for "incentive reimbursement" as a technique for reducing hospital costs. Payment might be based on a regional average cost. Hospitals with higher costs than the regional average would have to absorb part of the excess; hospitals with lower costs would be allowed to share part of the savings. Incentives would thus be introduced for each hospital to reduce costs. Over the longer run, hospitals of more than average efficiency would be able to accumulate internal funds for expansion and to demonstrate to lenders that their cash flow could amortize borrowings. Inefficient hospitals could not. An alternative scheme is to reimburse each hospital initially on the basis of its own costs, but to reward it for reducing costs toward the regional average and penalizing it in contrary cases.

Incentive schemes, however, will ultimately prove viable only to the extent we can distinguish change in cost for a constant quality of care from changes in cost associated with changes in quality. We seek a means to reduce hospital costs per unit of output. We do not seek a reduction in per diem costs achieved by lowering the quality of care provided. The National Advisory Commission on Health Manpower has suggested the establishment of peer review panels—groups of physicians, hospital administrators, and other professional personnel who would review the case provided and make judgments with respect to the quality and utilization of services.[11] While this approach appears to have much merit, and may become an indispensable part of any incentive reimbursement scheme, it needs to be backed up by the development of criteria for evaluation purposes. The range of services provided by a hospital are too complex and diverse to be evaluated as a lump, particularly when the evaluation is for the purpose of establishing payment for those services. Rigorous analysis of hospital care directed toward establishing output oriented functions or categories is clearly a prerequisite for the establishment of meaningful evaluation criteria for peer review panels or for any other quality evaluation mechanism. In short, to apply incentive reimbursement we must first make progress toward measuring the "output" of hospitals.

Other examples of inappropriate or positively harmful incentive systems are numerous. The American merchant marine is a case in point. We spend about $500 million per year to subsidize the American merchant fleet. The justification is twofold: to provide a carrying capacity in case of war and to have a weapon against discrimination by foreign-owned fleets in this highly cartelized business. While the case for subsidy is dubious on both grounds, let us accept its necessity and look at the subsidy system itself. The basic operating subsidy essentially makes up the difference between American and foreign operating costs. Any productivity gains result in lower subsidies—inefficiency leads to higher subsidies. Not only does the overall subsidy remove any incentives for efficiency, its detailed composition discourages the kind of merchant fleet which is most appropriate for the American economy. The subsidy brings each element of American costs in line with foreign costs—labor, repairs, and maintenance, etc. But the *comparative* advantage of American ships lies in higher speed—i.e., American ships' fuel costs are no higher than

[11] *Report*, op. cit., vol. I, pp. 46–48.

foreign costs, but labor costs are much higher. Consequently, American ships should be designed for high speed, quick turnaround time, catering to high-value cargo. But the subsidy system removes any incentive for pursuing this comparative advantage. To make matters worse, while the subsidy is partly justified as a means of providing a competitive weapon against potential foreign discrimination, our cargo preference laws have resulted in the fact that more than half of the export cargoes carried by U.S. ships are preference cargoes, leaving less than half of our capacity to compete in the world market. Even accepting the need for a subsidy program, we have designed one which is guaranteed to produce decreasing relative efficiency in the American merchant marine compared to its foreign competitors.

We have so far concentrated on the negative aspect of the incentive question; i.e., those situations in which the existing incentive structure produces private actions which run counter to the public interest. But in addition to removing "negative" incentives, public policy can also provide "positive" ones:

B. Providing new incentives designed to channel private actions toward public purposes.

Two important examples have already been given of areas in which incentives for private action might be created as a means of achieving quite specific public objectives:

Mandatory flood insurance, with premiums adjusted to risk, as a means of encouraging economic investment in flood plain lands.

Effluent charges as an efficient means of reducing air and water pollution.

In both of these cases, the necessity for a public program arises from the existence of external costs and benefits in the private market. But the possibility of creating an incentive system which channels private actions toward public goals is not limited to programs which primarily deal with externalities. The massive new Federal housing program, passed by the Congress in 1968, proposes to construct 6 million low-income housing units in the next decade, and to subsidize part of the rental or ownership costs. The primary objective of this program is presumably income redistribution—redistribution through transfers-in-kind, but redistribution nevertheless. (Anthony Downs has eloquently pointed out the host of problems connected with achieving this ambitious goal—problems whose solutions will require massive changes in urban institutions.)[12] Consideration of this program from an incentive standpoint, however, raises questions which go well beyond the matter of income distribution. In particular, the availability of authority to the Federal Government to contract for large blocs of low-income housing—both sale and rental housing, both multifamily and single unit—may make it possible to test various devices to encourage innovation in the development and construction of low-cost housing. Will the market provided by large-scale multiyear contracts induce new kinds of firms to enter the industry, encourage the development of specialized materials, promote substantial R. & D. expenditures by building materials and

[12] Anthony Downs, "Moving Towards Realistic Housing Goals," in *Agenda for the Nation* (The Brookings Institution, 1968).

construction firms? The Defense Department's proposals for the fast deployment logistics ships (FDL's) was based on a positive answer to these questions in the case of the shipbuilding industry. In the case of housing we do not know how important economies of scale and the security of long-term markets are. But the new program does make it possible to test whether large-scale markets and long-term contracts will provide incentives for significant cost reductions.

Another example of potential "incentive" creation arises in connection with manpower training programs.* The Federal Government in its manpower training programs has begun to shift emphasis away from numerous small-scale, publicly operated institutional training programs, toward subsidizing large-scale on-the-job training programs with private industry. In effect, Federal retailing of training programs proved relatively expensive and ineffective; hence the switch toward wholesale operations. The question arises whether incentive type contracts can be developed, in which private industry is automatically driven toward achieving the kind of results desired by public policy. For example, in training the hard-core unemployed one of the major problems is persistence of effort. Absenteeism is high; motivation is often low; accepting work discipline does not come easy; and all of this is sometimes exacerbated by the hostile attitude of existing employees. To the extent that these obstacles to increasing the employability of the hard-core unemployed can be eliminated with enough time on the job, then program benefits (measured by increases in the long-term employability and productivity of trainees) are not linearly related to hours of on-the-job training. Presumably, up to a point at least, the marginal benefit from on-the-job training rises with length of training. If subsidy contracts do not recognize this fact, but, for example, treat a half year's training for two men as equivalent to a full year for one man, the program will be less effective than it could be. This argues for incentive contracts—analogous to the incentive contracts in military procurement—under which the return to the contractor depends in part on persistence of effort.

In a similar vein, the Government should be interested in the "job mix" into which trainees are placed. If a contractor places all of them in menial and unskilled positions, and suffers no monetary penalties, his incentive for testing and motivating trainees is reduced. Again, incentive contracts which include as a performance measure, the job content of training positions, might well be devised to overcome the added costs of sophisticated testing and training for higher skills.

I have been able to devise no theoretical framework which could be used to help policymakers determine which kinds of public programs and public objectives lend themselves to the development of new incentives for private actions. The examples given simply show that the range of possible applications is quite wide. I can offer nothing more sophisticated than the proposition that if policymakers are continually aware of the importance and potentialities of the incentive approach, applications will suggest themselves in large number.

---

*Further discussion of this issue is found in the paper by Mangum in vol. 3 of this collection.

C. Improvement of Federal user charge policy as a means of more efficient resource utilization.*

User charges are a special aspect of incentive policy. Most of the potential applications are not new to economists. The use of prices as a rationing device, however, often seems alien to the public policy-maker.

One of the most dramatic examples is the current agitation over congestion and delays at major air terminals. It is easy to predict that limited facilities will be rationed by congestion, when the prices charged for facilities are either extremely low (as in the case of general aviation) or unrelated to the degree of congestion. I will not pursue this example in detail, except to note that in the case of air traffic, the administrative problems of levying congestion charges are far less than in the more classic case of highways (despite Professor Vickrey's ingenious suggestions for the application of congestion charges to highways).**

There is one major area of Federal programs in which user charges might well be profitably employed to change current *political* incentives for the better. At the present time, a large part of Federal water resource investment is devoted to projects whose major beneficiaries are identifiable, but who pay little if any of the cost of the projects. In the case of navigation projects, there are no charges levied on the barges which use the waterway. Beneficiaries of Federal irrigation and flood control projects typically pay only a small fraction of the cost. Moreover, there is little evidence that the benefits are distributed primarily to lower income groups, and much evidence to the contrary. But the availability of substantial subsidized benefits, primarily available to the "establishment" in the affected communities, sharply increases the political pressure for the authorization of marginal projects, and for the maintenance of very liberal benefit-cost criteria (low interest rates, liberal rules for defining and evaluating benefits, and so forth). Adoption of more stringent rules on user charges would probably reduce sharply the political incentive to maintain uneconomic project evaluation criteria and to lobby for very marginal projects.[13]

We have considered, to this point, the problem of publicly instituted incentives for *private* action. The second major type of incentive problem deals with the provision of incentives for *public* decisionmakers.

2. INCENTIVES FOR PUBLIC OFFICIALS DESIGNED TO IMPROVE THE EFFECTIVENESS OR EFFICIENCY WITH WHICH PUBLIC PROGRAMS ARE EXECUTED

In dealing with large scale programs, top level public executives labor under a major handicap compared to the situation facing their counterparts in large business concerns. Large complicated programs

[13] I am aware, of course, that where projects have declining average costs, marginal cost pricing rules may dictate low user charges. Yet I am convinced that the resultant shortrun allocational efficiency is often far outweighed by the lack of investment signals and the political incentives for poor project selection that large project subsidies entail. See also the papers by Krutilla and Milliman in this volume.

*Further discussion of this issue is found in the papers by Vickrey, Krutilla, and Milliman in this volume.

**Further discussion of this issue is found in the paper by Vickery in this volume.

must be carried out by a hierarchy of subordinate officials. Decentralization is necessary. But given the lack of any measurable performance criteria for subordinate officials, the top level public executive is often forced to specify in detail the set of permissible and nonpermissible actions of his subordinates. This leads both to excessive rigidity and often to poorly conceived program plans, imposed uniformly from the top on a variety of differing situations.

An analysis of this problem from the standpoint of creating incentives for decentralized operators to pursue public objectives, suggests two ways of attacking this problem:

- providing market competition for public programs.
- imitating market conditions in public programs.

These approaches may not prove viable in many situations; but at least in selected cases they may prove very useful in improving the design and execution of public programs.

A. Introduce market competition into decisions about the production of public goods.

The fact that public programs produce public goods does not imply that they need be completely sheltered from the competition of the marketplace.

Public elementary and secondary education is a case in point. In the inner cities of the Nation the public school system is virtually a complete monopoly, with a captive market, since in practical terms the private school alternative is open to few ghetto residents. Not only is the system a monopoly, but it usually tends to be fairly well isolated from control by the community, except as that control is expressed in overall budget limitations.

Much emphasis has been placed in recent years on educational research, and Federal funds have begun to flow into this field. But there is little use in inventing new or more effective approaches to compensatory education, if there is no incentive in the various school systems to adopt the more promising changes, and to evaluate alternative educational improvements seeking the most effective. Change in established routines and procedures threatens the security of the existing order, introduces uncertainty and tensions, and is inevitably painful on those who are subject to it. Without strong incentives, therefore, a monopolistic structure is unlikely to be very receptive to innovation.

In recent years a number of proposals have been made, designed to introduce a competitive element into the system, and thereby provide incentives for higher performance. These include the radical proposals of Friedman and Jencks which would in effect completely replace the present public school system with a private competitive model financed indirectly through school support grants made to individual parents. A more moderate suggestion has been made by James Coleman under which school districts would contract out, on bids, for specific parts of the public school curriculum—remedial reading, science courses, etc. A similar suggestion, adapted specifically for neighborhood-controlled school boards, has been proposed by Henry Levin.

The literature on this subject is growing rapidly. And I do not intend to try to summarize it. But analysis and exploration of this approach should not be confined just to the elementary and secondary education systems. Decisions about financing higher education—whether through grants and loans to individual students or through

aid to institutions of higher learning—should take into account the incentive effects of the decision. And we can go beyond the field of education. There is no reason why the public health facilities for the poor need be solely run by local or State public health agencies. Cannot private institutions (profit or nonprofit) be allowed to bid for a contract to provide such services?

B. Imitate the market more fully in public programs.

This category primarily deals with changes in organizational and budgetary structure designed to provide incentives for public officials to see *efficiency* in the administration of governmental programs.

The recommendations of the President's Commission on Postal Organization are principally directed to this end.* Creation of a public corporation, with power (subject to review) to set prices and wages, with authority to borrow from the public for capital investment, and with a directive to cover costs with revenues, would, in effect, make the Post Office an analog to a private utility. While this solution represents no panacea, it is far superior to the present arrangement in which there are few incentives for efficiency, and in which the basic variables of managerial control—prices, wages, investment, location of facilities, etc.—are decided by a 535-man "board of directors" primarily on political grounds.

Other examples, of a less dramatic kind, are not hard to find:

- Federal agencies could be charged in their budgets for the full costs of all the resources they use—rent for building space, full costs of employee retirement benefits, interest on capital equipment used in internal operations, etc. At the present time some resources are free and, therefore, either overused or controlled by arbitrary central regulations.
- Charge Federal construction agencies interest on funds during the construction period. In evaluating alternative bids on construction projects, agencies now have no incentive to put any value on time. In addition, use of an interest charge might provide incentives to reduce the seasonality of construction, which in turn should help to moderate the rapid increase in hourly wage costs of construction workers.**
- Charge Federal agencies for services provided by other agencies. The major case in point is the AEC's nuclear weapons production, which is now transferred to the Defense Department at a zero transfer price. While nuclear weapons costs are given shadow prices in the Defense Department's systems analysis, an actual charge to the DOD budget should strengthen the motivation to consider all costs fully in making decisions. Moreover, by providing the DOD with the funds for nuclear weapons and having them contract for production with the AEC, decisionmaking power would be transferred from the producer to the consumer, avoiding the natural tendency of producer-controlled decisions to result in excessive output.

The suggestions made earlier, to charge the Defense budget for subsidy programs primarily justified on national security grounds,

*Further discussion of this issue is found in the paper by Haldi in vol. 3 of this collection.

**I owe this suggestion to Professor David Martin, Indiana University.

is an extension of the "full cost" principle enunciated above. Principal examples are the merchant marine subsidy, and the commodity stockpiling progam. An even more radical extension would be to charge the Defense budget for the economic costs of those "protectionist" programs which are justified in national security grounds, but do not show up as budgetary expenditures. Examples would be oil import quotas and some part of the oil depletion allowance. The likelihood of being able to make these changes, of course, is so low that the suggestion should be treated as an interesting application of the basic principle rather than a serious proposal.

C. The problem of additivity and substitutability in Federal grant-in-aid and transfer-in-kind programs.

A very particular kind of incentive problem arises in the case of Federal grant-in-aid programs, which constitute a very large part of recent Federal social legislation. Presumably the purpose of the grant is to increase the resources devoted to a particular objective. (If the purpose of the program were to ease State and local overall financial burdens, revenue sharing or some form of tax credit would be much more appropriate.) But, as a matter of fact, little attention is paid to the problem of whether Federal grant funds, designed to achieve a particular purpose, add to the resources currently being spent for that purpose by State and local governments or simply substitute for funds that otherwise would have been spent by those governments.

Various "maintenance of effort" provisions have been written into Federal grant programs, but we know little about their effectiveness. In some cases, there may be other appoaches to maximizing "additivity." Title I of the Elementary and Secondary Education Act, for example, provides funds to *school districts* to be used for increasing the resources devoted to the education of poor children. But the larger and more diverse the income levels in the school district, the more difficult it is to determine the extent to which Federal funds have actually increased resources devoted to the purpose of the program. Allocation of Federal funds to units smaller than the school district (neighborhood areas, individual high schools with their associated junior and elementary schools, etc.) might help to increase the additivity of Federal funds.

But my point is not so much to describe "solutions" as to point to a major problem. To the extent that the Federal grant-in-aid system continues to be the major tool of social legislation, we need to do substantial research on the factors which determine additivity and to experiment with various devices to maximize additivity.

A similar problem arises in the case of Federal transfers-in-kind to individuals and institutions. To what extent does the transfer-in-kind increase the consumption of the particular good by the recipient, or merely substitute for funds which otherwise would have been spent on that good, thereby freeing up income for other consumption. More accurately expressed, we need to know the extent to which transfers-in-kind change recipients' consumption patterns from what they would have been if the same resources had been transferred through a cash grant.

Evidence seems to suggest that low income housing subsidies and food stamps do increase recipients' consumption of the goods in ques-

tion beyond what they would have been under cash grants.[14] But there are other cases which are distinctly questionable. The Federal Government's college housing program is a case in point. Under this program about $300 million in loans, at 3-percent interest, are made to public and private universities for the construction of college housing. For public universities, who have access to tax-exempt bond financing, the value of the interest subsidy is relatively modest, amounting to about $6 to $7 per month for each student housing in the newly constructed facility. Some 45 to 50 percent of the students at publicly supported universities come from families with incomes of $10,000 or more. Of those who board on campus, the proportion from upper income families is probably higher yet. At these income levels, it is hard to believe that a subsidy of $60 to $70 per year has any significant effect on the quantity or quality of education demanded, although it might have some small impact on the number of those who choose to board rather than commute. In the case of public universities, therefore, the transfer-in-kind probably has little effect in raising the consumption of the particular good, and is, in effect a disguised *cash* subsidy heavily favoring upper income groups.

More generally, Federal grant-in-aid and transfer-in-kind programs need to be reviewed in terms of the additivity question, and techniques devised to maximize incentives for grant recipients to use the funds for the purpose intended.

III. Conceptual and Political Problems

The first two sections of this paper concerned themselves with describing and classifying the problem of incentives, first, in terms of the various stages of policymaking and second, in terms of the various types of incentive problems. We now turn our attention to several major problems of concept and political feasibility associated with the use of incentives in public programs.

1. The definition and measurement of output

There is a major danger associated with the introduction of incentive systems into public programs. In the case of a private good, its various characteristics can be evaluated by the market and reflected in a single measure—*price*. Although buyers are not always perfectly rational, and sometimes lack all the relevant information, in most instances market prices are a reasonably good measure, at the margin, of the value of private output. There is usually no such single measure which commensurates all of the various aspects of the output of a public program—indeed that is why it is a public rather than a private output. The measurement of the output of public programs, therefore, is usually extremely complex. Yet, in devising incentive programs, we must know what is the set of outcomes that we wish to induce. If we are not careful, we may end up producing large unwanted side effects. This was the basic problem of early socialist production systems. The manager of a nail factory, whose quota was set in terms of the *number* of nails, and who was rewarded as he made or exceeded

[14] In the case of food stamps, additivity arises from the fact that recipients are required to purchase the stamps at an aggregate cost which equals what they had previously been spending on food. Barring a black market on stamps, this should guarantee additivity.

his quota, was inevitably driven to producing large numbers of small nails, regardless of market requirements. With production quotas specified as a certain *weight* of nails, the same manager would necessarily concentrate on producing a smaller number of very heavy nails—again, regardless of market demand.

More to the point, examine the problems of measuring output and defining objectives in a few of the incentive systems used as illustrations earlier in this paper. One suggestion was the establishment of incentive reimbursement schemes for Medicare and Medicaid, designed to induce more efficient design and operation of hospitals. Reimbursement on the basis of regional average cost was suggested, with low cost hospitals retaining some part of their "savings" and high cost hospitals being forced to absorb some fraction of excess costs. The discussion pointed out, however, that we are interested in reducing costs *per unit of output*, not in reducing costs through a sharp deterioration in quality of service rendered. Consequently, any incentive reimbursement scheme must be accompanied by some measure of and control over the quality of output produced. I suggested a possible approach: *first*, research directed toward establishing a usable classification of hospital "outputs", and *second*, the use of peer review panels which would rate hospitals on the basis of these classifications, with the rating results incorporated in the reimbursement formula.

Another example is the use of incentive contracts in Federal manpower training programs. Without a careful specification of the multiple objectives sought by the program, an incentive contract might well result in producing one kind of output (e.g. retaining trainees for long periods) at the expense of other aspects of output (e.g. training in a useful skill). A mandatory flood insurance program in which premiums were not reasonably adjusted to risk might increase rather than reduce uneconomic investment in the flood plain. A system of effluent charges on water pollutants, not devised to reflect a reasonable approximation of the costs of pollution, would result in either too much or too little effort being devoted to removing pollution and too much or too little industrial relocation. Introducing competition into the public school system (e.g. by allowing specific types of services to be contracted out to private bidders) may produce little of value unless there is some means of evaluating the various bids and some method of measuring performance.

In short, the use of incentive systems puts a premium on the careful specification of objectives and output. This is both an advantage and a disadvantage. An advantage because it forces program designers to be more specific in their statement of objectives and in the relative weights they attach to various aspects of output.[15] A disadvantage, because we often are literally unable, given the present state of knowledge, to specify objectives and assign weights to various aspects of performance with sufficient confidence to warrant the introduction of incentive systems.

What all of this suggests, is that the introduction of incentives into public programs calls for two things: *first*, extensive experimental projects, testing various alternative incentive schemes; and *second*,

[15] The introduction of incentive contracting into DOD and NASA had the side advantage of forcing decisionmakers to make explicit judgments about tradeoffs among various aspects of contractor performance—promptness of delivery, operating characteristics, costs, etc.

substantially increased research in the area of specifying and measuring output.

2. INCENTIVES TO INSTITUTIONS AND INCENTIVES TO INDIVIDUALS

Many of the examples suggested in this paper, particularly those related to the provision of incentives for *public* action, take the form of incentives to institutions: the inclusion of all relevant costs in agency budgets so they do not have access to "free" resources; incentive reimbursement for hospitals; reorganization of the Post Office, et cetera. But if decisionmakers in these institutions are not themselves judged and "rewarded" on the basis of criteria which are consistent with the incentives provided the institution, then little good will come of incentives to institutions: the inclusion of all relevant costs in budget of an agency or an installation, is to present the decisionmaker with the full costs of his decision. If he is interested in minimizing costs, he will presumably seek the lowest cost resource mix. But if he has no interest in cost minimization, then we should expect no results from the full cost approach.

It has been suggested that incentive reimbursement for most hospitals will accomplish little. Most hospitals are really run by the staff physicians, who are reputed to have little concern for the financial condition of the hospital itself.[16] To the extent this is true, then incentive reimbursement will not be very effective in lowering costs. Introducing competition into the public school system may change performance very little if the individual decisionmakers who let the contracts and evaluate performance themselves have no incentives to seek significant changes and improvements in the school system.

In one sense, these considerations suggest the obvious: that incentives for more effective and efficient performance of public programs cannot be considered apart from the structure of motivations, rewards, and penalties which determine the attitudes and actions of the bureaucracy. To the extent that public employees are themselves judged and "rewarded" by criteria which relate to the effectiveness and efficiency rather than to the mere size of the programs under their control, individual and institutional incentives can be made consistent.

3. POLITICAL PROBLEMS IN REGIONAL BUDGETING

In an earlier section of this paper, I pointed out that purely functional planning and budgeting by the Federal Government provides little or no scope for the consideration of tradeoffs among alternative Federal investments or projects at the local level. I then suggested that some form of regional budgeting, as an adjunct to functional budgeting, might serve this purpose, providing a framework within which bargaining about meaningful choices might take place.

As soon as one begins to consider specific means of carrying out this proposal, one major problem emerges. A complete regionalization of the budget allocation process would surely lead to major power strug-

[16] Personally, I do not put much store by this argument. Under incentive reimbursement schemes continually inefficient hospitals will eventually be squeezed out of existence. And to say that those who make hospital decisions are not overly concerned about the financial condition of the hospital does not square with the vigorous efforts of the American Hospital Association to obtain more favorable reimbursement formulas from Medicare and Medicaid.

gles among the various States, regions, and cities over the division of the budgetary pie. The current struggle over functional budgetary shares is widely diffused among a constantly changing set of alliances and factions. No Congressman or Senator need have his entire constituency at stake in any one functional budget decision. The divisiveness of the conflict is muted because of its complexity. But with purely regional budgets, the struggle would be etched in sharp outline, and the regional allocation of a given budget total would be a zero sum game to the participants.

To derive some of the advantages from regional budgets, however, it is not necessary to construct the "pure" model whose consequences were sketched above. There are a number of ways in which a partially regional approach might be grafted onto the existing process:

A. There are in existence, at the present time five federally sponsored regional development commissions. The commissions are composed of the Governors of the affected States and a Federal representative. Except for the Appalachian Regional Commission, the commissions are still in relatively embryonic form. Rather than become vehicles to lobby for "special" Federal authorizations and appropriations, the commissions might be organized to participate in the budgetary planning and allocation of those Federal projects which have special relevance for the economic development of the region: water resource projects, highways, economic development assistance, pollution control facilities, and so forth. The commissions could draw up several investment budgets—representing alternative overall constraint levels. These recommendations could then be used as guides in the formulation of the relevant nationwide functional budgets.

B. Functional budgets could be formulated, proposed to the Congress, and appropriated as is now done. In turn, for a selected number of investment and project-type programs[17] the relevant agencies could make allocations to major cities, or metropolitan areas.[18] Those functional allocations could be combined into consolidated budgets for each such area. In turn, the mayor (or a council of chief executives in a metropolitan area) could be given authority to reallocate funds, within prescribed limits, among the functional components. Constraints of various kinds could be imposed—for example, certain minimum sums for particular functions. Depending upon the nature of the project involved, the reallocation might be subject to joint approval by the local chief executive and the governor. Individual projects or grants would still have to be approved in terms of the relevant statutory and administrative requirements.

Several facets of this last suggestion deserve mention. First, it would substantially mute the regional battle over "shares", since the determination of those shares would depend on a complex set of functional decisions, and a host of separate regional allocations. Second, it would retain in Federal hands the authority to approve individual grants as a means of carrying out national objectives in a local context (for example, Model City grants would still be subject

[17] I.e., "project" as apposed to "formula" grants.

[18] Agencies, in effect, make such allocations now, as they approve individual projects or grants. Under the proposed scheme they would have to "budget" for such allocations in advance.

to the requirement of neighborhood participation in decisionmaking, hospital grants would be subject to minimum standards and planning requirements, et cetera). Yet, at the same time, it would transfer to State and local hands some additional authority over budget allocation—that is, authority to determine marginal trade-offs. The *composition* of the local budget would be less completely dictated by Federal budgetary decisions than is now the case. Finally, it would tend to transfer power from local department bureaucracies to State and local chief executives. Under today's purely functional budgeting, bureaucracy deals with bureaucracy—the Federal Public Health Service with its State counterpart, the Office of Education with State departments of education, and so forth. These relationships, coupled with the rapid growth of categorical Federal grant-in-aid programs, have tended to take a large part of the power over budget allocation out of the hands of State and local chief executives. Introducing some elements of regional budgeting into the Federal structure might help to strengthen the central authority and planning capability of State and local chief executives while retaining in Federal hands sufficient control over the use of funds to accomplish basic national goals expressed in functional terms. One final point on this matter. Any move toward "regionalizing" parts of the Federal budget will surely be vigorously resisted by many Members of Congress. Regionalization transfers some of the power over budget allocation from congressional subcommittees to Governors and mayors. Decentralization of power from the executive branch to State and local governments is one thing—presumably a good highly to be desired. Decentralization of congressional de facto powers is quite another matter.

I cannot pretend to have thought through the full consequences of budget regionalization. I am sure that as any particular proposal is worked out in detail, important difficulties and problems will be uncovered. But I believe that the basic concept of budget regionalization warrants further exploration, as a means of changing political incentives and motivations in the direction of making better allocation decisions at the local level among alternative Federal investment programs.

A MODEL OF PUBLIC DECISIONS ILLUSTRATED BY A WATER POLLUTION POLICY PROBLEM

BY ROBERT DORFMAN and HENRY D. JACOBY*

Robert Dorfman is Professor of Economics at Harvard University. Henry D. Jacoby is Assistant Professor of Economics at Harvard University.

In this paper Professors Dorfman and Jacoby adopt a unique point of view in discussing governmental decisions. Instead of making recommendations for public sector action which would attain specified objectives, they recognize that public decisions are indeed political decisions. They develop a framework or model of analysis which allows all of these political considerations to interact in the decisionmaking process. This framework, they argue, enables one to determine the likely range of outcomes of a political process in which decisions are reached through the interplay of a diverse set of interest groups with varying motivations.

In presenting their model of how political decisions get made, they have taken an imaginary problem from the field of water pollution control. They build into the model reasonable assumptions about politics, economics, hydrology, and sanitary engineering. Their analysis sketches out the range of alternative decisions which would be forthcoming from the assumed constellation of facts and relationships. Professors Dorfman and Jacoby assert that their view of the political decision process is important in understanding the workings of the process and may be "a practical tool of political analysis."

Merely corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative.

—W. S. Gilbert

1. *Introduction*

Governmental decisions may be approached from either a normative or a descriptive point of view. The normative approach accepts well-defined objectives for governmental undertakings and recommends specific policies and actions for attaining them. It will not be followed here. The descriptive approach accepts the facts of life, including the nature of governmental agencies and the purposes of diverse interest groups in the community, and attempts to provide insight into what will happen in the circumstances. That will be our approach.

Our method will be to construct a mathematical model of a political decision problem. The model will contain room for a great many data, ranging from the technological features of the problem that technical experts have to take into account to the political ob-

*This paper is based upon a chapter in Robert Dorfman, Henry D. Jacoby, et al., *Models for Water Quality Management*, now in preparation. The work has been financed by grants from Resources for the Future and from the Federal Water Pollution Control Administration. The responsibility for all opinions expressed in this paper lies entirely with the authors.

jectives and pressures that responsible officials have to evaluate. One of the advantages that we claim for the model, indeed, is that it provides a systematic framework for assembling such diverse data.

From these data the model will produce, mechanically, some predictions about the outcome of the political decision process. These will not be unambiguous predictions like an astronomer's prediction of the moment that an eclipse will take place. Rather they will take the form of stating a range of likely outcomes of the process, perhaps a fairly wide range but still much more limited than all the decisions that might be conceived of in advance. The power of a scientific theory, it has been said, is measured not by what it asserts but by what it precludes. This theory will preclude a great deal.

Within the range of likely outcomes the theory (or model) will provide some valuable information. It will highlight the political alinements that make some of the outcomes more likely than others and will indicate, rather specifically, the changes in the configuration of political influence that will tend to shift the decision from one outcome to another. Furthermore it will express vividly just how interests oppose and how different decisions affect the welfares of different participants.

Concede, for the moment, that such a model is possible. Whether this is so or not is for the sequel to determine. Then, surely the practice of politics should take it into account. (Note that we are being normative temporarily.) For, any political decision should be made in the light of a realistic assessment of its consequences, which requires a prediction of how things and people and political bodies will react to it. In the case of water pollution control, for example, the Federal Government's policies are implemented by the States and by river basin commissions; the States and river basin commissions work through local agencies and authorities, individuals, and business firms. Any decision at the Federal level has to be based on a prediction about how the States and basin commissions will respond to it; the States and basin authorities must similarly predict the reactions of subordinate units and of the public at large. Many other programs of Federal and State agencies are similarly affected by the responses of other governmental bodies. All legislation that concerns the powers and composition of Government agencies is influenced by predictions of how those provisions will influence the behavior of the agency. In short, a predictive model of political behavior can help improve political decisions.

Constructing a predictive model of political decision processes is an ambitious, indeed presumptuous, undertaking. There is, in fact, such good reason to believe that it cannot be done that we have construed our main task to be to convince ourselves and the reader that it is possible. Our method of proof is what mathematicians call "constructive." That is, rather than arguing the matter in the abstract we have considered a political decision problem taken from the field of water pollution control (not a real problem, but one that catches the essence of real problems) and have constructed a theoretical model that predicts the outcome of the political decision process in that instance. This construction shows that such a model is a possibility and exhibits its main features. It does not show that this model is a practical tool of political analysis. That showing would be a major research undertaking.

In order to show that this political model or any political model is empirically valid it would be necessary to apply it to several real political decisions and to compare its predictions with the observed political behavior. This would be a laborious and expensive task for it would necessitate ascertaining all the physical, technical, economic, and political data that the model requires and then performing elaborate computations.

We have not undertaken this mammoth empirical enterprise and not merely out of laziness. This large task is not worth undertaking unless the analytic model to be tested shows at least fair promise of success. A necessary preliminary to serious empirical testing is trial experimentation under favorable circumstances to see whether the model can be implemented in principle and whether it behaves sensibly. If the model passes the preliminary screening, then it pays to go further with it. This essay reports on such a preliminary testing of our conceptual model of political decision.

To carry out this test we have conceived an artificial basin with a pollution problem, known as the Bow River Valley. It is small, it is simple; but it could exist, in the sense that it does not violate any known principles of hydrology, sanitary engineering, economics, or politics. We have populated this valley with a large industrial source of pollution, two moderately sized cities, and a recreation area and placed it under the jurisdiction of a pollution control commission organized under the Clean Water Restoration Act of 1966. We have provided everyone concerned with such data and information as he might actually have under the circumstances and no more.[1] We have simplified the problem by suppressing much detail that would obscure the essential conflicts and issues that would arise. For example, we have reduced the specification of water quality to a single dimension, namely dissolved oxygen concentation (DO), and we have limited our description of the waste content of the various municipal and industrial effluents to the number of pounds of biochemical oxygen demanding material (BOD) that they carry.[2] In addition, we have limited the powers of the regulatory authority essentially to a single decision, namely, regulation of level of treatment by each polluter. We have ignored hydrologic uncertainty and other probabilistic complications and have made any number of simplifying approximations to facilitate computation. In short, we have loaded the dice heavily in favor of the model, as is perhaps appropriate for a preliminary test. The question now to be put is: Can this problem be expressed by a formal analytic model, and if so, does the model provide sensible and useful insights?

2. The Test Problem

The situation on which we are going to try out the conceptual scheme is sketched in figure 1.[3] As can be seen the Bow River flows generally from north to south. It is a respectable stream with a flow of 800 cubic feet per second (cfs) during the summer months. But it is not a very high quality stream. Because of the residual waste from up-

[1] Or, anyway, not much more.

[2] The description of the problem necessarily involves some technical concepts, such as these. Most of them will be explained below and in the Appendix.

[3] Prof. Harold A. Thomas, Jr., acted as consultant engineer for this study. We are indebted to him for formulating and analyzing the hydrological and engineering aspects of our sample basin.

stream cities the river, as it passes under the Gordon Bridge, has a dissolved oxygen concentration of only 7 milligrams per liter (mg/l). Without the influence of effluent discharges upstream, one could expect the level of oxygen in the water to be near saturation (8.5 mg/l at summer water temperatures).

The northernmost installation in the region under consideration is the Pierce-Hall Cannery. This is a large but somewhat outmoded vegetable cannery with an annual production of slightly over 7 million equivalent cases a year, concentrated in the summer and autumn months. The cannery adds an ultimate BOD load of about 1 pound per case to the river, after primary treatment, to which it is already committed. The cannery is not very profitable. Its net operating revenues, allowing for the cost of primary waste treatment, are only about 7.5 percent of stockholders' equity. It employs about 800 workers, many of whom live in Bowville (population 250,000), 10 miles downstream.

Bowville and the other riparian city, Plympton (population 200,000), are both fairly large centers supported by varied light manufacturing and commercial establishments serving the surrounding agricultural region. Both have waterfront parks of which they are proud, and Plympton in particular has some aspirations to being a tourist center because of its proximity to Robin State Park. Plympton draws its water supply from the Bow River; Bowville brings its water from a better quality tributary. Both cities discharge their wastes into the river after primary treatment. For simplicity we shall assume that neither city anticipates that its population or its waste load will grow significantly in the foreseeable future. This simplification will save us from having to forecast growth rates and from having to consider the possibility of "building ahead of demand." In fact, it enables us to neglect all dynamic considerations.

Robin State Park, between Bowville and Plympton, has woodland recreational facilities. All concerned would like to develop its waterfront for boating, fishing, and, if at all possible, swimming. It is used by the inhabitants of both cities, by the neighboring farm population, and by some tourists and day-trippers from outside the valley. Everyone is agreed that the quality of the water in the neighborhood of the park should be improved.

Thirty miles below Plympton the river crosses a State line and flows out of our ken.

The current quality of the stream at critical points under low-flow conditions is shown in figure 2. From just below Bowville down to the State line water quality is very poor during summer droughts. For long stretches the river is anaerobic (i.e., the DO level falls to zero), and it is unfit for recreational or other use. In response to a generally felt need to improve the river, especially near the park, and in response to some pressure from the State water commission, the Bow Valley Water Pollution Control Commission has been established, with the editor of the *Bow Valley News* as chairman and with membership drawn from the city councils of both cities and including the deputy State commissioner of parks and recreation. One member, of course, is Mr. Pierce (of the Pierce-Hall Cannery Pierces), who is one of the most respected businessmen in the valley and also a dedicated fisherman.

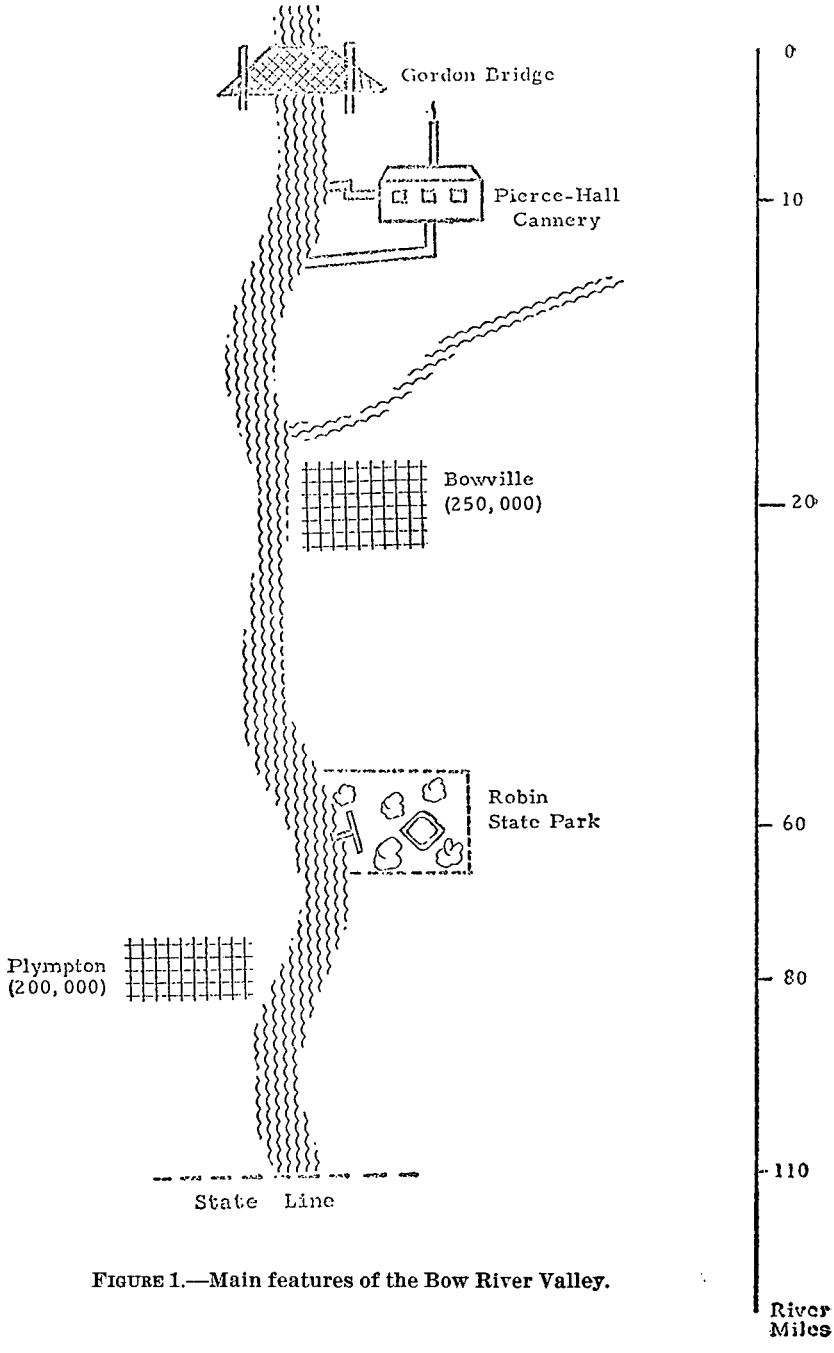


FIGURE 1.—Main features of the Bow River Valley.

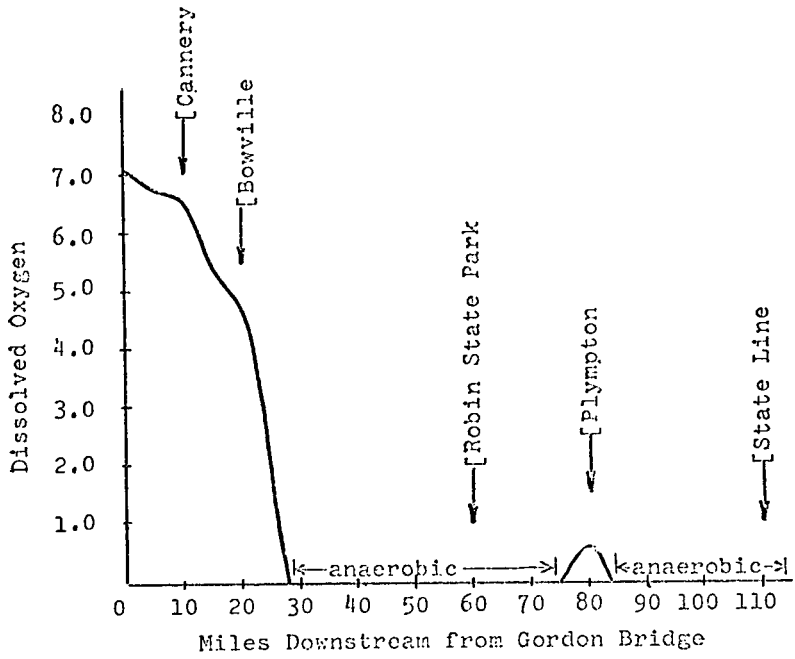


FIGURE 2.—Current water quality in the Bow River, summer drought conditions.

The commission faces two crucial problems. One is to recommend a quality classification for the reach of the river within its jurisdiction.[4] For reasons that we cannot go into, the whole stream from Bowville to the State line has to be assigned to the same quality class. The other problem is to decide on the levels of treatment to be required of the three sources of waste within its jurisdiction. The treatment levels must be high enough to achieve the quality standards for the stream classification that has been adopted, but may be higher. The standards are expressed entirely in terms of dissolved oxygen levels.

Such are the decisions that we must analyze. But before considering how the commission as a body will act, we must see in more detail how the problem looks from the point of view of each of the interested parties.

2.1 *The Pierce-Hall Cannery*

The Pierce-Hall Cannery, located just 10 miles upstream from Bowville, is a relatively large installation compared to other plants of this type around the country: over the course of the 180-day canning season the plant handles approximately 40,000 (24/303) case equivalents per day of a variety of fruits and vegetables. On the average these products sell for $3.50 per equivalent case, and so most years the plant runs a total gross sales of around $25 million.

[4] Quality classifications will be explained more fully below.

In response to mounting public concern for water quality, the plant managers already have identified and incorporated some internal process changes in order to reduce effluent volume and cut back on biological pollutants discharged into the Bow River, and they have installed primary treatment facilities in the form of screening and sedimentation equipment. Even after these changes, however, when the plant is in full operation it discharges a waste stream of approximately 30 million gallons per day (mgd) which carries an ultimate carbonaceous biological oxygen demand of 28,000 pounds per day and an ultimate nitrogenous demand of 19,600 pounds per day.

In order to reduce these waste flows further, Pierce-Hall would have to install additional treatment facilities, and the plant manager has already contacted the consulting firm of Mini-systems, Inc., to obtain preliminary estimates of the cost of attaining different degrees of improvement in effluent quality. The results of that study are shown in table 1.

TABLE 1.—COST OF ADDITIONAL WASTE TREATMENT AT PIERCE-HALL CANNERY

Type of treatment	Percent [1] removed of—		Additional annual cost	Additional net cost to Pierce-Hall per year	Profit after taxes per year
	Carbonaceous BOD	Nitrogenous BOD			
Primary (now in place)	30	30	0	0	$375,000
Primary plus low efficiency secondary	80	40	$13,000	$8,000	367,000
Primary plus high efficiency secondary	90	55	59,000	35,000	340,000
Primary plus high efficiency secondary plus tertiary	95	90	159,000	95,000	280,000

[1] The figures shown are percentages of the gross waste load.

As the table shows, 30 percent of the carbonaceous and 30 percent of the nitrogenous BOD in the Pierce-Hall effluent is removed by sedimentation and screening in the firm's primary treatment plant. To accomplish higher levels of BOD removal, the waste stream would be passed through a tank where biological degradation—which in the absence of treatment would occur in the stream itself—can take place under controlled conditions. The degree of purification can be varied over a wide range by proper design of the plant. High degrees of BOD removal are naturally more expensive than low.

From the wide range of possible choices, the engineering consultants to Pierce-Hall have provided two alternative secondary treatment plant designs. The first, referred to rather loosely as a "low-efficiency secondary" plant, would bring the total carbonaceous removal up to 80 percent of the original load, and would raise the level of nitrogenous removal to 40 percent. The total cost for the low-efficiency unit would be $13,000 per year as shown in the third column of table 1. The second design, referred to (again rather loosely) as a "high-efficiency secondary" unit, would accomplish a greater total carbonaceous removal, 90 percent instead of 80 percent for the low-efficiency unit, and an even better performance in removing nitrogenous material, 55 percent as opposed to 40 percent. The cost, however, goes up considerably as shown in column 3 of the table.

If it is necessary to subject the effluent from a high-efficiency plant to still further treatment, there is yet another set of processes, generally referred to as "tertiary treatment," which may be used. In effect,

tertiary treatment is any process which reduces the waste contained in the effluent of a secondary unit, by such diverse and expensive methods as holding in stabilization ponds and adding special chemicals. One such design has been provided by Pierce-Hall's consultants. It would remove up to 95 percent of the firm's carbonaceous wastes and up to 90 percent of the nitrogenous. As may be seen in table 1, tertiary treatment costs considerably more than "secondary" processes.

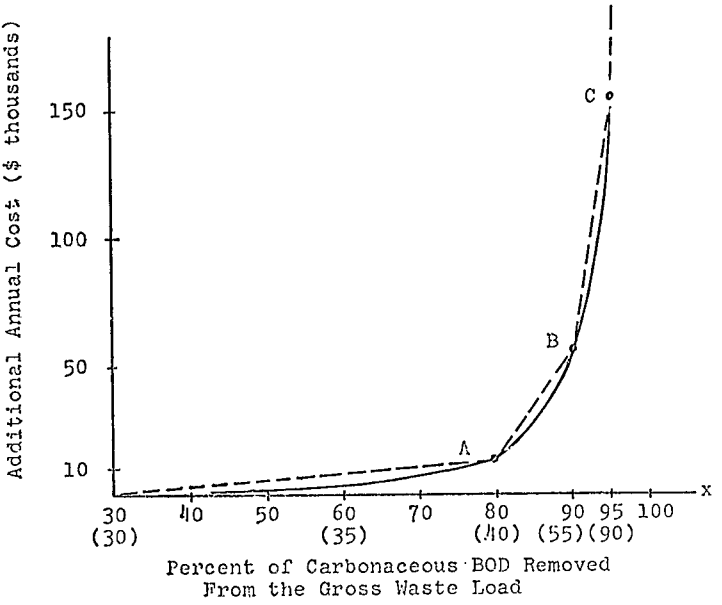


FIGURE 3.—Cost function for additional waste treatment at Pierce-Hall Cannery.[a]

[a] The numbers shown in parentheses indicate the percentage of nitrogenous BOD removed by the treatment plants designs which form the basis of the cost function.

Of course, the plants shown in table 1 are only three of an infinite number of alternative designs. The curvilinear cost function in figure 3 indicates the actual range of alternatives, with additional treatment cost stated as a function of the percentage removal of carbonaceous BOD from the gross waste load. Because the cannery already has primary treatment facilities, the zero point on the cost curve is at the 30 percent removal level. Since the data are so limited, we shall inter-

polate linearly between the data points provided by the consultants. The resulting piecewise linear cost function is shown as a dashed line in the figure. The costs entailed by any treatment plant can be stated as a function of its degree of carbonaceous BOD removal because, to a good approximation, the percentage of nitrogenous material removed is a linear function of the degree of carbonaceous removal. For each of the consultants' designs, the percentage of nitrogenous BOD removed is shown in parentheses. For example, a plant that brings total carbonaceous BOD removal up to 60 percent of the gross waste load will also remove to 35 percent of the nitrogenous BOD. Figure 3 shows that such an installation would cost $6,500 per year.

Due to the influence of the Federal corporation tax, the net impact of additional treatment on the after-tax profits of the Pierce-Hall corporation would be less than the total treatment cost. Pierce-Hall's accountants have estimated that the net cost to the firm would be 6/10 of the total cost, and this latter figure is shown in the table as well.

The firm's net operating revenues, after income taxes, are 1.5 percent of gross sales when primary treatment only is employed. This net profit amounts to approximately $375,000 a year, which is equivalent to 7.5 percent of the stockholders' equity. The firm is not a price leader and does not anticipate that it will be able to raise its prices appreciably even if a large increase in treatment costs is imposed. Neither does it know of any changes in its methods of processing that would enable it to reduce its waste load at the current scale of operations. Therefore any increase in treatment costs would have to come out of net profits. The estimated impact of different levels of treatment on net profits is also shown in table 1. Notice that the effect is appreciable; the highest level of treatment would reduce annual profits by over 25 percent.

On the other hand, the management of Pierce-Hall is not adamantly opposed to improving the quality of the Bow River, even at some cost to themselves. As the major industrial polluter in this reach of the river, they recognize that they have some responsibility to users and inhabitants farther downstream. Besides, many of their employees live in Bowville and make use of Robin Park, so that improvement of the river will enhance the amenities available to the plant's work force. Finally, the firm has some tentative plans for expanding by constructing a more modern plant on a site near Plympton. The efficiency of the branch plant would be greatly increased if the river near Plympton were of good enough quality to be tapped for washing water, which it is not at present. But the management also keeps in mind that it will not be able to raise the capital necessary for expansion if the return on the present equity falls below about 6 percent.

In short, the position of the Pierce-Hall management is a bit complicated. The prospect of improving the quality of the river is both a threat and an opportunity. They would like to see the river improved but do not feel that they can afford to contribute very much toward bringing improvement about.

2.2 *Bowville*

The city of Bowville, 10 miles downstream from the cannery, is the second major source of pollution in the valley. The city receives good quality water. According to the data in figure 2, Bowville escapes the real pollution problems even during severe summer droughts. But the 250,000 inhabitants plus assorted light industries dump a heavy load of wastes into the river. Even after primary treatment, Bowville discharges 89,600 pounds of ultimate carbonaceous BOD and 33,600 pounds of ultimate nitrogenous BOD into the river on an average summer day. The total volume of effluent discharged is about 51 mgd. This load, added to the cannery wastes, renders the river unsuitable for recreational use farther downsteam.

In anticipation of the prospective discussions of pollution management in the valley, the mayor of Bowville requested his public works department to prepare estimates of what it would cost the city to install additional treatment facilities. The results of those preliminary studies are shown in table 2. Like the cannery, Bowville has obtained cost estimates for three different treatment plant designs, each removing a certain percentage of carbonaceous and nitrogenous waste. These estimates are shown in the table.

TABLE 2.—COST OF ADDITIONAL WASTE TREATMENT AT BOWVILLE

Type of treatment	Percent [1] removed of—		Additional annual cost	Additional cost to city per year	Addition to property tax rate (thousands)
	Carbonaceous BOD	Nitrogenous BOD			
Primary (now in place)	30	30	0	0	0
Primary plus low efficiency secondary	80	40	$650,000	$490,000	$1.17
Primary plus high efficiency secondary	90	55	880,000	660,000	1.58
Primary plus high efficiency secondary plus tertiary	95	90	2,520,000	1,890,000	4.52

[1] The figures shown are percentages of the gross waste load.

Bowville would not have to bear the total cost of cleanup unaided. Under the provisions of the Federal Water Pollution Control Act, Bowville could count on a Federal grant in the amount of 50 percent of the construction cost of these facilities, and since capital cost is about half the total cost of waste treatment in this case, the citizens and local industries would have to bear only about 75 percent of the total outlay for these facilities. The adjusted cost is shown in the fourth column of table 2. All these costs are based on the assumption of a 20-year life of facilities and a 5-percent interest rate. Just as with the cannery, both the total costs of different degrees of waste removal and the net costs to the city can be approximated by segmented linear cost curves of the type shown in figure 3.

Table 2 shows also the estimated effect on the tax rate of the different levels of treatment. These data are of particular interest to city officials and taxpayers. It is seen that the adoption of either high- or low-efficiency secondary treatment would have only a moderate effect on the property tax rate. (It is now $63.50 per thousand assessed valuation.) Tertiary treatment is far more expensive and the city comptroller has expressed some alarm that it might be adopted. He points out that tax rates are bound to rise in any event, because of recent increases in teachers' and firemen's salaries, and that they are

already higher than tax rates in Plympton, which competes with Bowville for new industries.[5]

It appears that Bowville would gain only moderately from improvement in the quality of the Bow River. Bowville Waterfront Park is already a fine facility. But cleaning up the river would make the entire valley more attractive to tourists and vacationers and, if the improvement were sufficient to permit the development of water-based recreation at Robin State Park, the park would become far more useful to inhabitants of the city. This latter consideration is important. Bowville Waterfront Park is already so overcrowded that some thought has been given to condemning some adjacent warehouses in order to expand it. This would not be necessary if any substantial proportion of the users could be diverted to the State park.

Besides the *Bow River News* has been running editorials like, "Restore the Bow River!" and, after droughts, "The Shame of Bow River Valley." So there is considerable pressure on Bowville to contribute its share to improving the river, provided that the cost is reasonable.

2.3 *Plympton*

The difference between Plympton and Bowville is the difference between upstream and downstream riparian residents everywhere. Bowville has good water; Plympton has poor. Bowville's wastes degrade the river for most of the distance that concerns us, including the waterfront at Plympton; Plympton's wastes discommode no one, affecting only the quality checkpoint at the southern outlet of the valley. Bowville can improve the quality of the river by subjecting its wastes to a higher level of treatment; Plympton is virtually helpless—it cannot even protect its own waterfront. The stage is set for the classic conflict between upstream and downstream users.

Plympton is slightly smaller than Bowville and generally less affluent. It also has a primary treatment plant. Its effluent volume is 43 mgd, containing after primary treatment 67,000 pounds of ultimate carbonaceous BOD and 25,000 pounds of ultimate nitrogenous BOD.

Although Plympton has, in fact, little effect on the quality of the river in the region that concerns us, it must expect to bear its share of responsibility for cleaning up the river. Indeed, since the inhabitants of Plympton will be the major direct beneficiaries of improved quality in the river, this city is particularly eager to contribute what it can, to help put pressure on the more influential users upstream. The cost data for treatment at Plympton are shown in table 3, and these also can be expressed as a cost function of the type shown in figure 3. If you compare these data with the cost table for Bowville, you will see that the dollars-and-cents costs of each level of treatment are lower for Plympton, because it is a smaller city, but the effect on the tax rate is greater. This is because Plympton is a poorer city and the value of taxable property per capita is lower there. Fiscal problems are generally harder for Plympton than for Bowville.

[5] There are many ways for a city to finance waste treatment facilities other than by the property tax as we assume here. Alternative financing methods differ not only in the impact of a particular facility on the city budget and in the distribution of costs among households and business firms; they also affect the total amount of waste to be handled. For example, certain types of sewer fees offer an incentive for industries and commercial establishments to cut back on total waste by means of internal process changes, while financing by the property tax offers no such incentive.

It is true, nevertheless, that Plympton is more eager to participate in a program of river quality improvement than Bowville is. They would like to develop recreational facilities on their own waterfront, which does not now conform to sanitary standards and is occasionally beset by riverine odors. They are more dependent than Bowville on tourism and are closer to the State park. For all these reasons, the Plympton Chamber of Commerce is one of the leaders in the movement for improving the river.

TABLE 3.—COST OF ADDITIONAL WASTE TREATMENT AT PLYMPTON

Type of treatment	Percent [1] removed of—		Additional annual cost	Additional cost to city per year	Addition to property tax rate (thousands)
	Carbonaceous BOD	Nitrogenous BOD			
Primary (now in place)	30	30	0	0	0
Primary plus low efficiency secondary	80	40	$550,000	$410,000	$1.37
Primary plus high efficiency secondary	90	55	730,000	550,000	1.83
Primary plus high efficiency secondary plus tertiary	95	90	2,110,000	1,580,000	5.27

[1] The figures shown are percentages of the gross waste load.

2.4 *Other Interested Parties*

The cannery and the two cities are the principal defilers of the river and represent the interests of most of the people who are directly concerned. But there are other interests too, which have to be taken into account and which are likely to have considerable influence on any decisions about the river.

There is, in the first place, the Federal Government, which has an expressed interest in protecting and improving the quality of all interstate waters, including the Bow. This national interest is implemented by the Federal Water Pollution Control Administration. We have already mentioned that the Bow Valley Water Pollution Control Commission was established under incentives and grants provided by the FWPCA, and that the FWPCA can be expected to contribute generously to meeting the costs of increased levels of waste treatment undertaken by the two municipalities. The FWPCA also has some enforcement powers. It has induced the State to promulgate a set of water quality standards, which will be discussed below. Now it is interested in inducing the Bow Valley Water Pollution Control Commission to classify the river at the highest reasonable use level.

The FWPCA's interest in the river is twofold. First, it is responsible for protecting the interests of all users farther downstream, so that it feels compelled to, and is empowered to, insist on a reasonable quality of water at the southern outlet from the valley. It should be noted that the FWPCA is the only participant with a direct concern for the quality of the water so far south. Secondly, the FWPCA has a generalized interest in the quality of American streams and in protecting our natural heritage. It shares the concern of the inhabitants of Bowville and Plympton for the quality of their waters and for the waterfront potential of the State park. We can expect, therefore, that the FWPCA will press for higher quality of water throughout the Bow without acute concern for the effect of increased treatment costs on local tax rates.

The State government is also concerned, particularly through the State water and sanitation commission and the State department

of parks and recreation. The State industrial commission may take a hand, however, if the financial health of the cannery should be jeopardized by any proposal. On balance, the influence of the State agencies can be expected to be similar to that of the FWPCA except that the State will attach more importance to the quality of the water at Robin State Park and less to the quality at the State line.

There are, in addition, a variety of conservation and special interest groups, perhaps typified by a branch of the Izaak Walton League. Although they have little voting power, these groups make their influence felt through all the media of public communications, through any hearings or investigative procedures of the Bow Valley Water Pollution Control Commission, through participation in municipal politics, and even through direct representation on the commission.

All the groups mentioned in this subsection share a more keenly felt concern for the quality and usability of the river than for the cost of achieving high quality. For the purposes of our discussion we shall lump them together and consider them to be represented adequately by the FWPCA.

2.5 *Water Quality Standards*

The first task before the Bow Valley Water Pollution Control Commission is to adopt a stream use classification for the Bow River between Bowville and the southern outlet at the State line. The State water and sanitation commission has promulgated a set of standards that prescribe the quality of water to be used for different purposes. Once the commission has adopted a use classification, all its subsequent regulations must conform to it; that is, they must be designed to assure water of quality at least as good as that specified in the State standard for water in the use class that has been adopted.

In actual practice stream standards cover many stream characteristics—dissolved oxygen, floating solids, color and turbidity, coliform bacteria, taste and odor, temperature, pH, radioactivity, and others. For simplicity of exposition and analysis, however, we shall pretend that the State standard specifies only one dimension of stream quality; namely, instream dissolved oxygen.[6]

The State water standards are accordingly assumed to divide streams into five use classes: A, B, C. D, and U. Class A waters are very nearly in their pristine state—almost unaffected by man. Waters classified U are essentially uninhabitable for fish life, unusable for higher forms of recreational activity, and unpleasant to the sight, taste, and smell. Classes B, C, and D identify intermediate conditions of quality, and associated with each is a specified level of stream use. The language of the State standards is as follows:[7]

Class A.—Water designated for use as public water supplies. Character uniformly excellent.

Minimum dissolved oxygen—3.5 mg/l

Class B.—Suitable for bathing and recreational purposes including water contact sports. Acceptable for public water supply with appro-

[6] Other pollutants that would be especially important in the sort of situation we are discussing include coliform bacteria from municipal waste, and nutrients. We assume that the former is taken care of by chlorination (which is relatively cheap), and the latter we do not handle explicitly but mention here because it is a lively issue currently.

[7] Here we use the Massachusetts standards, slightly adjusted, as a model.

priate treatment. Suitable for agricultural, and certain industrial cooling and process uses; excellent fish and wildlife habitat; excellent esthetic value.

Minimum dissolved oxygen—5 mg/l

Class C.—Suitable for recreational boating; habitat for wildlife and common food and game fishes indigenous to the region; certain industrial cooling and process uses; under some conditions acceptable for public water supply with appropriate treatment. Suitable for irrigation of crops used for consumption after cooking. Good esthetic value.

Minimum dissolved oxygen—3.5 mg/l

Class D.—Not objectionable; suitable for power, navigation, and certain industrial cooling and process uses.

Minimum dissolved oxygen—2 mg/l

These standards must be met under the average minimum consecutive 7-day flow to be expected once in 10 years; for the Bow River, this flow is 800 cfs. Under these low-flow conditions, the entire river below Bowville currently is below the quality specified for Class D waters; indeed, as shown in figure 2, in many parts of the valley the stream goes anaerobic and is occasionally rather unpleasant to be near.

It should be noted that the imposition of defined standards of quality, which is almost inevitable in framing administrative regulations,[8] transforms the decision problem in a fundamental way. In the absence of codified classifications, a DO concentration of 5.1 mg/l will be recognized as only imperceptibly safer and more pleasant for swimming than one of 4.9 mg/l. But once the higher concentration qualifies the river for a higher use classification, there is all the difference in the world between them—one permits the river to be developed legally for water contact recreation; the other does not.

A fundamental discontinuity is thereby introduced into decisions that impinge on water quality. This will have important consequences for our analysis, as we shall see later.

2.6 *Waste Discharge and Stream Quality*

We have been using instream dissolved oxygen as an indicator of water quality and "biochemical oxygen demand" or BOD as the measure of the pollution content of the effluent of the two cities and the cannery. Improvement in dissolved oxygen concentrations in the river can be obtained in a number of ways—e.g., artificial aerators or flow augmentation—but this example assumes the removal of oxygen-demanding material at the waste source as the only management method available. The technical link between the amounts of BOD discharged and the stream quality at selected points downstream will be discussed in the Appendix.

In our model we shall use the simplest formulation of the relationship between waste loads and stream quality based on the work of Streeter and Phelps which dates back over 40 years.[9] According to the

[8] Examples range from the definition of grade A fresh eggs to the occupancy and safety regulations in building and zoning codes.

[9] Only the most casual treatment of the technical aspects of water quality is offered here and in the Appendix. A brief introduction to the topic is presented by Kneese and Bower [6], pp. 13–29. For more detailed treatment see Fair, Geyer and Okun [4].

Streeter-Phelps model of stream quality, the effect of discharging BOD into a stream at any point is to reduce the dissolved oxygen in the water at all points downstream by amounts that are directly proportional to the amount of BOD and that depend in a complicated way on the distance from the point of waste discharge and on the hydrology of the stream.

In subsequent discussion of the mathematics of the model, we shall use the symbol d to denote the factor of proportionality between waste discharge and quality response. Of course, the precise value of d depends on where along the river the waste is being discharged and at what point downstream the impact is being measured. We shall use the subscript i to denote the different polluters in the Bow River system, and the subscript j to indicate individual quality control points. Since we know the geographical location of each of the waste sources and control points, we can calculate an appropriate value of d for each pair of i and j, and these proportionality factors or "transfer coefficients" we shall denote as d_{ij}[10]

The following example will illustrate the use of these parameters. At the present time the cannery is discharging 28,000 lbs/day of carbonaceous BOD. The impact of a pound of carbonaceous material dumped at the cannery ($=1$) on dissolved oxygen at Bowville ($j=2$) is $d_{12}=.0000556$. If the carbonaceous load from the cannery were cut by 26,000 lb/day, the response at Bowville would be an improvement in dissolved oxygen of 26,000 x .0000556 or 1.45 mg/1 A similar calculation would apply to any reduction in the cannery's nitrogenous waste discharge.

In addition, the impacts of different waste sources of downstream quality are additive. For example, if both the cannery and Bowville reduce their waste discharges, then the impact of each reduction on water quality at the park may be calculated in the manner shown in the preceding paragraph, using appropriate values of d_{ij}. The overall quality improvement at the park will be the sum of the influences of the two abatement measures.

Of course, since we use a river as our example, as opposed to a lake or tidal estuary, waste discharges have no effect on upstream water quality. An individual polluter does not even affect the quality of the stream at his own waterfront, that is $d_{ii}=0$.

3. Model of the Bow Valley Water Pollution Control Commission

The tedious recital of data in the last section will be recognized as a small-scale replica of the docket of any proceeding concerned with the use and control of public waters. Data and considerations of the sort that we have described are amassed in the form of staff and consultant reports, briefs, submissions, affidavits, transcripts of public hearings, court records, judicial findings, rulings of administrative agencies, and so on. The task of the commission is to digest, assimilate,

---

[10] The units of d_{ij} are rather complicated. We measure waste discharge in pounds of BOD per day (lb./day) and water quality in milligrams per liter (mg./l.) of dissolved oxygen. Thus the units of d_{ij} are milligrams per liter per pound per day or mg./l. per lb./day.

There is yet another complication in that, due to differences in the speed of reactions, separate values of d_{ij} must be calculated for carbonaceous and nitrogenous BOD. This additional detail will be covered in the appendix.

and ultimately evaluate this mass of data, argument, rhetoric, threat and cajolery dealing with a mixture of technical, economic, legal, demographic, political, esthetic, moral, and social considerations. We are now concerned with what can be said about the upshot of this task.

One kind of consideration that we have not yet mentioned is precisely what the commission is trying to achieve. If you ask them, they will point to the preamble of their charter where they are instructed "so to regulate and control the use of the Bow River between the Gordon Bridge and the State line and the discharge of liquid and solid matter of any form whatsoever thereinto as to assure the highest practicable quality of water between and including the aforementioned points and to conform to all applicable State and Federal laws and regulations."

That sets the tone of the problem in a general way, and the concluding phrase sets some limits on what the commission can do, but the only word that conveys any hint of the true complexity of the task is the word "practicable." What does it mean? Engineering calculations, to be discussed more fully below, show that it is simply not within the realm of technical practicability to classify the river as class A according to the State standards given in the previous section. At the other extreme, Federal regulations on the classification of interstate waterways preclude the adoption of class U. Between these two extremes a wide range of choice remains. What is "practicable"?

Besides, the charter provides only very vague guidance about an important implication of the commission's decisions: Who is to bear the financial burden of such improvements as it may ordain? The commission will decide this, in effect, when it imposes requirements for waste treatment on the various polluters within its jurisdiction. All the charter says, in another section, is that the treatment requirements must be "reasonable, fair, and equitable and in conformity with State and Federal law."

The perplexing aspect of the problem before the commission is that it is obliged to impose regulations on potent political and economic bodies whose interests, as we have noted, by no means coincide. Besides, the commission is not distinct from the people it must regulate. They are represented on it, both formally and informally, and the commissioners share the varied interests of the people they are charged with regulating.

In the nature of the case it will not be possible to please everyone, but the commission can be counted on to try.

In the circumstances, a practicable set of regulations must mean one to which all concerned can be persuaded to submit without recourse to the law or to higher political authority or to other drastic action. It means, in our specific context, that the policy must not threaten to bring about the downfall of the municipal administrations if they agree to it, that it must not undermine the profitability of the cannery, and, of course, that it must not be illegal.

In the early stages of consideration, it will not be clear whether these are loose or stringent restrictions on the commission's scope for decision; all interested parties will endeavor to make them appear very stringent. At any rate, the first order of business before the commission is to find some policy that is practicable in this sense.

It is highly unlikely that there will be only one such policy. Then a choice must be made among practicable policies and another ambiguity in the commission's charter obtrudes. What is meant by "highest quality"? In a genuine instance where many characteristics of water are taken into account one must ask whether the quality of water is increased by measures that increase the dissolved oxygen concentration at the expense of increasing the amount of phosphates or other plant nutrients in the water. Such questions cannot arise in the Bow Valley, where the dissolved oxygen concentration is the sole measure of water quality. But one must still ask whether a measure that increases dissolved oxygen at the park but reduces it at Plympton constitutes an improvement in quality. Such questions must be decided, somehow, by the commission, and it is not likely that any answer will command the wholehearted assent of all interested parties.

Now that we have appraised the complexity of the commission's task, we see that it would be folly to aspire to an analytical procedure that would enable us to predict the precise outcome of its deliberations. We can do some useful analysis, but we must be content with less than an unequivocal prediction.

Our position is that coldblooded, objective analysis of a political decision process such as this can achieve three things. First, it can organize the data and use them to delimit the range of likely outcomes of the process, and, furthermore, delimit them somewhat more narrowly than may seem possible at first blush. Second, it can indicate the forces and considerations that tend to move the decision toward one or another portion of the range of likely outcomes, and can indicate the balance of forces required to produce any outcome. Third, it can facilitate greatly the comparison of different possible outcomes by showing who are the losers and who the gainers, and in what respects, and by how much by moving from one outcome to another. Furthermore, we maintain that the results of such analysis are useful and informative both to the commission itself and to outside critics and observers. Our defense of these claims will be actually to perform the calculations for the test case under inspection and to interpret the results.

Our first challenge, then, is to delimit the range of possible outcomes of the elaborate process of deciding on a policy. The limitations on the range of outcomes follow partly from the requirement that any decision must be practicable in the sense that we have described above, and partly from the general economic principle of Pareto admissibility. In general, a decision that affects a number of persons is said to be Pareto admissible, or simply admissible, if there is no alternative, allowable decision that is preferred by one or more persons and that nobody regards as inferior. In other words, admissibility requires that no opportunity is foregone that will improve the welfare of some people without harming anyone. A commission such as ours would reach an inadmissible decision only by mistake, and the widespread representation of interested parties is a virtual guarantee that no such mistake will occur.

In our particular context an admissible decision has to satisfy the following conditions:

1. There must be no alternative practicable decision that provides the same quality of water everywhere and that requires a

lower level of treatment by one of the polluters and no higher treatment by any of them.

2. There must be no practicable alternative that provides a higher quality of water anywhere without requiring a higher level of treatment by at least one of the polluters. In fact, there must be no alternative that increases the quality of the water without necessitating an increase in treatment by some polluter that costs more than he is willing to pay for the improvement.

3. Any alternative with lower treatment costs must reduce the quality of the water so much that at least one of the polluters would prefer the higher level of treatment for himself and the others.

Fortunately, all these conditions, both of practicability and of Pareto admissibility, can be expressed mathematically. We shall see very shortly how this can be done. When it has been done, we can determine whether any proposed decision by the commission is admissible and practicable by a straightforward mathematical calculation. Furthermore, it is possible by purely mathematical computations to sketch out the range of admissible alternative decisions. When we have shown how to do this, we shall have substantiated the first claim that we made for the possibilities of analysis.

In the subsequent discussion, we shall refer to the mathematical formulation of the conditions of admissibility and practicability as a model of the commission. The mathematical formulation is a model in the sense that it enables us to say a great deal about the outcome of the commission's deliberations. But it should be noted that this model does not mimic in any way the actual procedures of decision-making that the commission will follow. The mathematical analysis provides much of the same information that the commission develops by means of a method that is quite different from the one followed by any regulatory body.

One might almost say that the elaborate procedures of assembling engineering and economic reports, conducting public hearings, holding commission meetings, conferring with interested parties, and so on are society's way of formulating and solving the mathematical equations of the model. Such an interpretation has a grain of truth but is basically misleading. Society's procedures in some ways do more and in some ways do less than the mathematical computations. They do more in that the stage of report writing and public hearing assembles and digests the necessary data, while the mathematical analysis contains no provision for data acquisition. The conferences and commission meetings enable the commission to ascertain two or three admissible decisions and to select among them. The mathematical analysis can ascertain many more admissible decisions but does not contain any procedure for choosing among them.

Now it is time to construct our model.

3.1 *Decision Variables*

The first step in formulating the commission's problem mathematically is to express the decisions open to it in numerical form. The first decision is the use classification of the river. It will be recalled from the description of the water quality standards that the effect of adopting any use class is to prescribe the minimum permissible amount of dissolved oxygen in the river. For example, if use

class C is adopted there must be at least 3.5 mg/l of dissolved oxygen in the river at all points. Let us then denote by Q the minimum permissible concentration of dissolved oxygen. Then the commission must, in effect, choose a value of Q, either 7, 5, 3.5, or 2, depending on the use class selected. In order to avoid having to consider the effects of the commission's decisions on users farther downstream, we impose that the dissolved oxygen concentration at the State line shall be 3.5 mg/l. This specification is regarded as outside the purview of the commission and will not be considered further.

The only other decisions that the commission has to make concern the level of treatment to be required of each polluter. These, too, can be expressed numerically. Each polluter has a range of possible treatment plant designs from which to choose; as illustrated in figure 3, the more effective the waste removal the higher the cost. The commission has to prescribe the degree of removal which each discharger must attain, or in more precise language, the percentage of carbonaceous BOD which must be removed from each of the polluter's waste outflows.[11] We express these decisions numerically as follows. First we assign identifying numbers to the three polluters: 1 will designate the Pierce-Hall Cannery, 2 will designate Bowville, and 4 will designate Plympton (the number 3 denotes the State park, and no wastes are generated there). Then we introduce a set of variables, called x_i, where x_i denotes the proportion of gross carbonaceous BOD removed by polluter i from his effluent. Accordingly, the commission must decide on x_1, the proportionate removal to be required of the cannery, x_2, the proportion to be removed by Bowville, and x_4, the proportion to be removed by Plympton. For example, if the commission decides on $x_1 = 0.85$ then the cannery will have to remove 85 percent of carbonaceous BOD contained in its effluent before treatment.

The decision on the x_i simultaneously determines the treatment cost that each polluter has to bear. The relationship is shown clearly in figure 3. If $x_1 = 0.85$ the cannery will have to build the plant that lies halfway between points A and B in that figure, which will cost it $36,000 a year—i.e., $1/2(13{,}000+59{,}000)$. The simplest way to express mathematically the relationship between percent removal and cost is to introduce some auxiliary variables to represent the proportion of BOD removed in accordance with each segment of the broken-line cost curve. These auxiliary variables will be distinguished by appending a second subscript. In the case of the cannery, x_{11} will denote the proportion of BOD removed by methods along the line segment from the origin to point A, x_{12} will indicate how far along the segment AB the cannery is required to go, and x_{13} will show how far along the segment BC. Note that x_{11} cannot exceed 0.5, x_{12} cannot exceed 0.1, and x_{13} cannot exceed 0.05. The total amount of BOD removed is simply the amount removed by primary treatment plus the sum of the amounts indicated by the auxiliary variables. In the case of the cannery

$$x_1 = 0.3 + x_{11} + x_{12} + x_{13} \qquad (1)$$

If $x_1 = 0.85$, we shall have $x_{11} = 0.5$, $x_{12} = 0.05$, and $x_{13} = 0$.

[11] Recall that we are assuming that the technology of treatment is such that the degree of removal of nitrogenous BOD is fixed in relation to the degree of carbonaceous removal. So although the commission may set treatment requirements in terms of carbonaceous BOD only, implicit in the requirements is a certain percentage cut in nitrogenous load.

The cost of any level of treatment, x_1, is determined by the corresponding values of the auxiliary variables and the costs of advancing along the line segments, which are indicated in the figure by the slopes of the segments. Numerically,

$$\text{Cost of achieving } x_1 = 26{,}000x_{11} + 460{,}000x_{12} + 3{,}180{,}000x_{13}.$$

For example,

$$\text{Cost of achieving 0.85 removal} = 13{,}000 + 23{,}000 + 0 = 36{,}000,$$

as we found before by simple interpolation.

The cost of achieving x_2 removal from Bowville's effluent and x_4 from Plympton's can be computed by using analogous concepts and relationships. An important property of the cost curves is that in the case of all three polluters the cost of removing a unit of BOD increases as you move off one of the cost segments to the segment on its right in the diagram.

The commission's decision problem has now been reduced to the selection of four numbers: Q, which determines the use class of the river, and the three x_i ($i=1, 2, 4$) which specify the levels of treatment required of all polluters. The decisions about these numbers are not all independent. Once Q has been chosen the x_i are forced to be high enough so that the level of dissolved oxygen does not fall below Q mg/l at any point in the river. The relationship between the x_i and the concentration of dissolved oxygen at points downstream from polluter i was discussed above in the section on "Waste Discharge and Stream Quality." That analysis led to a set of coefficients, d_{ij}, that give the effect of a unit reduction in the oxygen demanding discharge of polluter i on the dissolved oxygen concentration at any downstream point j. Since the x_i determine the amount of oxygen demanding discharge by polluter i, there is an equation, in which the d_{ij} are the important coefficients, relating the x_i to the dissolved oxygen content of the water at any downstream point j. The x_i have to be chosen so that the dissolved oxygen at all points is at least as great as Q, except that at the State line it must be at least 3.5 mg/l, irrespective of Q. There is one such equation for each of the quality checkpoints in the river, designated by $j=2$ for Bowville, $j=3$ for Robin State Park, $j=4$ for Plympton, and $j=5$ for the State line. A typical equation of this group can be represented in the form: [12]

$$\sum_i d_{ij} L_i (x_i - 0.3) \geq Q - \bar{q}_j.$$

In this equation L_i is the gross BOD load (i.e., before primary treatment) generated by polluter i, d_{ij} is the transport coefficient expressing the effect of his load on point j downstream, and x_i is his proportionate removal of BOD from his effluent.[13] The summation on the left-hand side expresses the increase in dissolved oxygen concentration at point j resulting from the waste treatment in excess of primary employed by all polluters.

[12] The constant 0.3 is subtracted from each x_j since this is the BOD reduction achieved under the basal condition of primary treatment by all polluters.

[13] In the calculations a somewhat more complicated formula was used in order to allow for the differences between carbonaceous and nitrogenous BOD. See the detailed discussion in the Appendix.

On the right-hand side, $\bar{q}_j$ is the dissolved oxygen concentration at point j when only primarily treatment is used by all polluters. These are the numbers plotted in figure 2. The right-hand side is then the increase in dissolved oxygen at point j required by the use classification corresponding to Q, and the entire inequality asserts that the x_i must be chosen large enough to achieve the required improvement at point j. There is, as we said, one such equation for each quality checkpoint.

3.2 *Practicability Constraints*

Once the x_i have been chosen the increases in treatment costs imposed on the polluters follow as we have seen. The formulas expressing the relationship of removal level to cost (as sensed by the individual polluter) are part of the model. In the sequel we shall frequently denote the cost imposed on polluter i by g_i.

The costs imposed on the FWPCA are somewhat special. In this model the FWPCA plays the role of the custodian of the overall national interest. Therefore, the costs of any pollution-abatement plan as perceived by the FWPCA are the full economic costs of the additional treatment, without any allowance for the effects of tax advantages and Federal grants-in-aid. In other words, the FWPCA is taken to react to economic resource costs rather than to budgetary costs.

There are limits on the treatment costs that the commission can realistically impose on the polluters. As was mentioned above, the commission cannot impair Pierce-Hall's earning abilities excessively. This consideration can be expressed in the mathematical model by including a condition such as $g_1 \leq \$50{,}000$, which would prohibit any decision that cost Pierce-Hall more than $50,000 a year.

The attitude of Plympton, described above, suggests a different kind of practicability constraint on the decision. That city feels an acute need for water-based recreational facilities, which it does not now have. Indeed, to them, the entire point of the Bow Valley Water Pollution Control Commission is to provide them with water of suitable quality. Therefore their representatives on the commission cannot agree to any decision that does not provide for at least, say, 3.5 mg/l dissolved oxygen at Plympton, which is adequate for boating and fair-to-middling fishing. They'd like more, but if compelled, could accept that standard for their locality. It is easy to incorporate this consideration into the model. In equation (2) for $j=4$ (denoting Plympton) simply substitute 3.5 for Q if Q is not already that high.

So far we have described a set of mathematical relationships that any practicable decision must satisfy. Conversely, any decision that does satisfy these relationships will be technologically, economically, and politically acceptable. However, the conditions so far described are not sufficient to assure that the decision will be Pareto admissible.

3.3 *Pareto Admissibility*

The conditions for Pareto admissibility can be introduced into the model by a combination of a mathematical trick and some further refinement of concepts. It will be expedient to discuss the conceptual refinement first.

Up to this point our formal model has taken account of the costs of alternative decisions and of technical and political limitations on what

can be done but has scarcely mentioned the benefits that would accrue from various policies for pollution abatement. This is the defect that must now be corrected.

This reticence about the advantages of water quality improvement is a consequence of the fact that we habitually talk and think about technological relationships and about economic costs in numerically quantified terms, but resort to much vaguer forms of expression when we consider the overall advantages and drawbacks of different courses of action. We have already discussed, in narrative form, the attitudes of the citizens of Plympton and Bowville and the managers of Pierce-Hall toward improvements in the quality of the Bow River. But those qualitative descriptions do not throw much light on what improvements the various participants are willing to pay for or on how much each of them is willing to pay for given improvements. Yet we must have that information if we are to form any just appreciation of how the commission is likely to decide among alternative plans, for it will take both the advantages and the costs into account in reaching its decision.

Some such information will transpire in the ordinary course of events. The commission will reach a decision, the water will be improved in a quantitatively measurable degree, measurable costs will be imposed on the participants. Then, after the fact, we should have reason to believe that the advantages secured by that decision were worth the costs to each participant. For otherwise they would have resisted the decision, and Bowville, Plympton, and Pierce-Hall are all too potent to be overridden by the commission or to be compelled to submit to a decision that any one of them regards as unreasonable, unjust, intolerable, or extortionate. Any actual decision must on balance be acceptable to all participants though it may fall far short of what they would desire.[14]

We must now, therefore, suggest a method for estimating the value of improvements in water quality as perceived by each of the participants before the decision has been taken. Our suggestion will apply to this problem the concepts employed in economics for the analysis of consumption decisions.

Any decision that the commission may make results in a package of consequences to each of the participants. One consequence is the treatment costs imposed. The other consequences are improvements in the quality of the river at various points. Whether the decision is a good one or a bad one from the point of view of any participant is a net resultant of this whole package, depending on whether the improvements are great enough and useful enough to him to be worth the cost to him. We should have the calculus that we need if we could analyze the package into its components, estimate the worth of each of the components to the participant, and total up the results. For example, suppose that a decision would improve the water at Bowville by A mg/l and the water at Robin State Park by B mg/l and would increase Bowville's treatment costs by $\$X$ per year. Then if we had

[14] It is entirely conceivable that quality demands will be so high and limits on practicable treatment costs so low that no decision will satisfy all the conditions. In that event either some of the requirements would have to be modified or the commission would have to disregard some of them and face the ensuing law suits and other distasteful consequences. This model will detect the fact that the decision problem has no solution, but gives no indication of what happens then.

reason to believe that Bowville would be willing to pay $\$Y$ for that much improvement of its own waterfront and $\$Z$ for that much improvement at the park, we could regard the net benefit to Bowville as $\$Y+\$Z-\$X$, and might predict Bowville's reaction to the proposal by noting whether this sum is positive or negative, large or small.[15]

In short, we are raising the following possibility. Suppose that we could estimate the amount that Bowville would be willing to pay for an improvement of 1 mg/l at her own waterfront, say $\$v_2$. And suppose we had a similar estimate, say $\$v_4$, for a 1 mg/l improvement at the State park. Then would we be justified in saying that Bowville would be willing to pay $\$Av_2+\$Bv_4$ for an improvement of A mg/l at Bowville and B mg/l at the State park? Are we justified in breaking up the package into its components, attaching a value to each of the components, adding up, and regarding the total as the total of the entire package? If so, and if we can make the individual estimates, we have the elements of the calculus that we require.

This is an old and classic question in economics. The answer is a qualified affirmative. The value of an entire package to any participant (in the sense of what he would be willing to pay to obtain it) is the sum of the values of its components provided that (1) the quantities of the components are not so great that their marginal utilities change significantly, (2) the components are neither substitutes nor complementary goods, and (3) the total value of the package, computed in this way, is not a significant proportion of the budget of the participant. We should consider these provisos in order.

Successive improvements in the quality of the water at any one point may have different importance or utility to any participant. For example, improvements that bring the water at some point up to swimmable quality are more useful and important than improvements beyond that point, though even those, by improving fishing and the pleasurability of swimming, are of value. It is not difficult to allow for this complication by attaching one value to a unit improvement in the quality of water that is below the swimmable standard and a smaller value to improvements in better quality water. It is, in fact, possible to indulge in as much refinement as seems required in adjusting the value of successive improvements to the level of quality already attained. By so doing we can measure the value of the total improvement in water quality at some point by summing the values of successive increments, each of which values reflects the level of quality already achieved.

For simple summation of the values of improvements at different points to be valid, those values should not be interrelated. Values of improvements at different points would be interrelated if, for example, there were two State parks equally accessible to Plympton. For then, if a water-based recreational facility were created at one of them the citizens of Plympton would be less willing to contribute to the cost of developing the other. In that case, the value of improvements to one of the parks would depend upon the extent of development of the other; those improvements would be substitutable goods as far as Plympton is concerned. Another possibility would be for

[15] There is a delicate point here. The amount that Bowville would be willing to pay for a given improvement at her waterfront might very well depend on whether or not the park is improved, and conversely. We shall come back to this possibility later.

the quality of fishing at the park to be affected by the quality of water both at Bowville and at the park. Then if the water were improved at Bowville it would be more valuable than otherwise to improve the water at the park, because the impact on fishing would be greater. This is an instance of complementarity: improvements of water quality at Bowville would enhance the value of each unit of improvement at the park.

These considerations show that substitutability and complementarity invalidate, in opposite ways, the simple addition of the values of improvements at different points along the river. Whether or not such interactions occur in any instance depends on the circumstances, and particularly on the use that would be made by the participants of the improved water at different points. In the case of the Bow River it appears reasonable to neglect such interactions.

Finally, some attention must be paid to the relationship between the total value of the improvement package and the budgetary position of the participants. This is because a participant would not be willing to pay as much for one component of a package if the cost of other components had already strained his budget as he would be willing to pay for it in isolation, without having other costs to bear.[16] This is a serious complication in principle. In our case, expenditures on water quality improvements are only a minor proportion of the total expenditures of any of the participants. We can, as a first approximation, ignore the effect of increasing budgetary drain on willingness to pay for still more improvement, but we shall have to stay alert to the possibilities for error that this approximation may lead to.

On these grounds we feel justified in adding up the values of the components of an improvement package, with due caution, in order to estimate the value of the entire package to any participant. We now confront the problem of estimating the values of individual components, that is, of unit increases in the quality of the water at different points along the river.

It must be conceded that it is not usual to make explicit monetary estimates of the value of water quality in reaching decisions about pollution abatement and related matters, except in the case of some Federal benefit-cost analyses. But such quantifications are made implicitly all the time and are, indeed, indispensable in arriving at rational decisions. At a public hearing on a proposed pollution-abatement program such an exchange as the following would not be at all extraordinary. The Bowville Commissioner of Water and Sanitation is on the stand:

Q. How much would it cost you to reduce your daily waste load by an additional two percent?

A. About $27,000 a year, I think.

Q. If you did that we could assure swimmable water at the park. How would you folks feel about that?

A. I can't answer for the City Council, but I would certainly support that kind of proposal myself.

The Commissioner from Bowville has made a judgment, albeit implicitly, about the dollar-and-cents value to Bowville of improvement

[16] Note the similarity to the difficulties caused by the declining marginal utility of expenditure in the conventional theory of consumer behavior.

in water quality at the park. He feels that sufficient improvement to permit swimming at the park is worth at least $27,000 a year to Bowville. With more questioning a more precise evaluation could be elicited.

Such judgments are made in various ways, of differing degrees of formality. One method is the "user-days" approach. An estimate might be made of the additional use of the park by Bowville residents that would be induced by improving the quality of the water at the park. Then if we could estimate the value of each day's use of the park by a Bowville resident, we should have an estimate of the value of the improvement. There are grounds for doing this. For example, Bowville's park and recreation budget amounts to $6 per capita. The average Bowville resident uses park or recreation facilities about 30 times a year. Thus, the city of Bowville is paying $0.20 per user-day for its citizens' use of public recreation facilities. This figure, with an estimate of how Bowvillers would respond to improved water recreation at the park can provide an estimate of the value to Bowville of such an improvement.

Another method for valuing public improvements is the "alternative cost" procedure. This is applicable when the improvement meets a need that will be satisfied by other means if the improvement is not undertaken. In such a case the value of the improvement is the saving that it affords by rendering the alternative expenditure otiose. For example, we have noted that a sufficient improvement at Robin State Park will make it unnecessary for Bowville to undertake the expense of expanding its own waterfront park. To illustrate the computations, suppose that expanding Bowville's waterfront park would cost $165,000 a year on an annual equivalent basis and that an improvement of 5 mg/l at the park would avoid the need for this expansion. Then such improvement at the park would be worth $33,000 for each mg/l improvement, on the average, up to the swimmable point.

Such estimates of the value of improved water quality have been made from the viewpoints of Bowville and Plympton, by these and other methods, and are recorded in table 4. It will be noted that each city places a value on improvements at its own waterfront and also on improvements at the park. In addition, the value of improvements, wherever they occur, depends on whether the water has already attained a quality of 5 mg/l of dissolved oxygen, which meets the State standard for water contact use.

TABLE 4.—VALUE OF UNIT IMPROVEMENTS IN WATER QUALITY

[Dollars per year per milligram per liter]

Place improved	Current water quality (mg/l)	Participant		
		Bowville	Plympton	FWPCA
Bowville waterfront	<5	100,000	----------	100,000
	≥5	50,000	----------	50,000
Plympton waterfront	<5	----------	75,000	75,000
	≥5	----------	25,000	25,000
Park	<5	33,000	30,000	94,500
	≥5	17,000	10,000	40,500

The table also contains estimates of the value of unit improvements from the point of view of the FWPCA. The FWPCA is regarded, again, as the custodian of the public interest. Its valuation of the improvement of water quality at Bowville and Plympton is simply a reflection of the values placed by the inhabitants of those cities on the improvement of their own water. Some empirical support for this imputation can be found in the Federal Government's habit of making 50–50 matching grants-in-aid to encourage local improvement. The FWPCA's valuation of improvements at the park is the sum of the valuations of Bowville and Plympton with 50 percent added to allow for the social value of use of the park by local residents who do not live in the two cities and by outsiders.

This table of social valuations enables us to insert numerical estimates of the benefits of water quality improvement into our model, to great advantage as will be seen shortly. Such estimates are necessarily rough and tenuous. As it happens the results of our analysis are not very sensitive to the precise social valuations assigned to the improvement of water quality at different points. In actual application if the results should appear to be sensitive to the estimates of social valuation of improvements in water quality, of course the analysis should be reiterated for a range of such estimates so as to form an appreciation of the seriousness of this source of imprecision.

The refinement that we have made, in short, is to replace the vague appreciations of the importance of quality improvements which were contained in our narrative by estimates of the amount that the participants would be willing to pay in order to obtain unit improvements in quality at different places in the region. This increase in the precision of expression is necessary for the completion of the mathematical model and also, in fact, for the attainment of rational decisions, although, to be sure, this particular mode of expression is not normally employed. The habitual weighing of benefits against the cost of obtaining them, which is usual, makes us feel that this mode of expression is not unduly farfetched.

We can now introduce into the model numerical measures of the benefits that each participant will derive as a consequence of any decision. The participant's gross benefit is the total amount he would be willing to pay for the improvement in the state of the river that would result from the decision. His net benefit is the gross benefit less the costs imposed on him by the decision. For example, suppose it is decided to improve the quality of the water at Bowville (where it is already 5 mg/l by 1 mg/l, the quality at the park by 2 mg/l, and, in order to accomplish this, to raise Bowville's treatment cost by $150,000 per year. Then Bowville's gross benefit from this decision is $1 \times \$50{,}000 + 2 \times \$33{,}000 = \$116{,}000$, and her net benefit is *minus* $34,000. The fact that Bowville's net benefit turned out to be negative, that she would be forced to pay more than the improved water quality is worth to her, does not necessarily preclude this plan. There are considerations of responsibility and harmony that escape our crude measure of benefits. But such a measure does tell something important about the attitude that a participant is likely to take to a proposal. This is not one that Bowville could be expected to favor if there were a practicable alternative.

Now comes the mathematical trick. We are going to add up the net benefits that a plan confers on all participants to obtain a total net benefit. Let it be emphasized that this is a meaningless operation, in spite of the fact that economists and philosophers have been trying to perform it ever since the days of Jeremy Bentham. You can no more add up the total amount of net benefits conferred along the length of the river than you can add up the scores at the Miss America Pageant to calculate the total amount of beauty displayed. But if you go through the motions you can obtain useful results. In fact, you can ascertain the decisions that are Pareto admissible.

An algebraic formulation is necessary here. Let X stand for any decision—remember that X is specified by the four numbers Q and x_i. Let the net benefit that participant i derives from the decision X be denoted by $NB^i(X)$. Finally, introduce some arbitrary weights, denoted by w_i, to be assigned to the different participants. Then for any decision X we can compute from the data given above the net benefits to all participants, the $NB^i(X)$, and also the weighted sum of net benefits, $\Sigma w_i NB^i(X)$. The weights, though arbitrary, are of great significance. They measure the relative importance that is assigned to the fates of the various participants in evaluating the aggregate consequence of the decision.

Now we can set ourselves a purely mathematical problem: With respect to any given set of weights, find the decision X, that (a) satisfies all the requirements for practicability, and (b) makes the weighted sum of net benefits, $\Sigma w_i NB^i(X)$, as large as possible. It is easy to see that the decision X that is a solution to this problem must be Pareto admissible. For if it were not there would be some other decision, say Y, for which $NB^i(Y)$ was at least as great as $NB^i(X)$ for all participants and actually greater for at least one of them. But, then, $\Sigma w_i NB^i(Y)$ would be greater than $\Sigma w_i NB^i(X)$ in contradiction to the fact that X solved the problem of making this sum as large as possible.

So, taking any set of weights, our model and straightforward mathematics determine a practicable and admissible decision. A different set of weights will determine a different decision, more favorable to the participants whose relative weights have been increased. By trying out a number of different sets of political weights, we can map out the range of possibilities. We can also see how much the relative weights have to be changed in order to move the upshot from one decision to another.

What the "true" weights should be we have no way of knowing. Ultimately, that is for the commission in its various deliberations and negotiations to decide. But we can tell what the range of likely outcomes is and even how the relative amounts of influence of the different participants are likely to affect the precise outcome.

Let us see how this approach can be applied to the Bow River Valley.

4. Exploration of Possible Decisions

The conclusion of the previous section was that we can identify the practicable and Pareto-admissible decisions by solving the formal mathematical problem of finding the decisions that maximize the weighted sum of net benefits, denoted by $\Sigma w_i NB^i(X)$, while satisfying the constraints and relationships that describe the conditions of politi-

cal practicability, the technological connection between treatment undertaken and quality achieved, the relationship between treatment levels and costs incurred, and so on. In the present case all the relationships are linear or piecewise linear, so that this mathematical problem is a linear programing problem, which is very convenient for computational purposes. It may be necessary in other instances to resort to more complicated mathematical formulations, with consequent increase in the cost of calculation, but such complications would not be justified for our present illustrative purposes.

The linear programing problem that corresponds to our model is presented in full in the Appendix. It has been coded for the SDS 940 Time Sharing System.

The model is applied to the problem in two steps, corresponding to the two decisions to be taken. First we imagine that Q, which determines the use classification of the river, has been chosen at one of its three possible levels: $Q=2$ mg/l for Class D, $Q=3.5$ mg/l for Class C, or $Q=5$ mg/l for Class B.[17] Then we assume a set of relative weights, w_i, specifying the relative emphasis given to the welfare of each of the four main participants. The sets of weights used in the computations are shown in table 5. They range from a set that pays exclusive attention to the overall national interest as reflected by the FWPCA ($w_4=1$, all others equal zero) to a set that places major emphasis on the profitability of Pierce-Hall ($w_i=7$, all others equal 1). Intermediate weighting schemes, such as $w_i=3$, 3, 1, 3, and $w_i=3$, 1, 3, 3, place varying degrees of emphasis on the welfares of the two cities.

TABLE 5.—WEIGHT ALLOCATIONS USED TO EXPLORE POSSIBLE DECISIONS

RELATIVE WEIGHT ASSIGNED TO—			
Pierce-Hall (w_1)	Bowville (w_2)	Plympton (w_3)	FWPCA (w_4)
0	0	0	1
3	1	3	3
3	3	1	3
1	4	4	1
4	1	4	1
4	4	1	1
1	2	6	1
1	1	7	1
1	7	1	1
7	1	1	1

We have calculated the treatment levels and the consequent costs and net benefits that maximize the weighted sum of net benefits for each choice of use classification and each set of weights, by solving the corresponding linear programing problem. Thereby we have traced out the range of likely decisions and have exhibited the effects of variations in the use class and in the political influence weights.

In addition we have repeated some of the calculations with a special constraint that limits the amount of treatment that can be required of Pierce-Hall. We can therefore see the effect of such special consideration on the decision and on the other participants.

And, finally, we have computed the "least cost" solution corresponding to each of the three use classifications. This was done by solving for the plans which will meet each of the quality standards ($Q=2$,

[17] Class A($Q=7$ mg/l or above) was found to be not attainable given the pollution loads and treatment alternatives which we assumed earlier.

$Q=3.5$, and $Q=5$) at the minimum national resource cost, ignoring the influence of any national benefits which may be attached to improved water quality. These solutions serve as a basis for estimating the increase in economic cost which is incurred when the benefit evaluations estimated in table 4 are taken into account.

It turned out that in spite of all our variation of weights and conditions only 12 different decisions were ever recommended by the analysis. They are described in table 6 in terms of the quality standard attained and the treatment levels, x_i, assigned to each of the polluters,

TABLE 6.—DECISIONS RECOMMENDED BY THE ANALYSIS

Decision No.	Use classification	Treatment level in terms of percentage carbonaceous BOD removal		
		Cannery	Bowville	Plympton
1	D	95	58	75
2	C	95	75	68
3	C	95	80	67
4	B	95	90	60
5	B	95	90	60
6	B	94	90	61
7	C	80	90	67
8	D	90	65	76
9	C	90	82	70
10	C	90	86	67
11	C	90	90	64
12	B	90	92	60

The effect of some of the decisions on the quality of the river is shown in figure 4, which should be contrasted with the portrayal of existing conditions in figure 2. Note that if decision 5 is taken, the quality level is maintained above 5 mg/l at all quality control points except that at the State line. On the other hand, decision 7, which is favorable to the cannery but less attractive to both Bowville and Plympton, yields a lower quality level along almost the entire reach of the river.

In part, our finding that only a small number of distinct decisions is worth considering is a consequence of the usual behavior of linear programing models.[18] Such models do not make very fine distinctions and tend to discover a limited set of alternatives each of which is recommended for an appreciable range of circumstances. It is easy to see why this is so. Suppose that on some trial decision the cannery is required to construct the low-efficiency secondary plant shown in table 1. This is the same design that lies at the breakpoint A on the segmented cost curve of figure 3. Furthermore, suppose that the benefits of removing a small amount more of the carbonaceous BOD at Bowville are greater than the costs of doing so. Then, since the costs of removing each additional one percent of waste from the effluent do not vary over the range from 80 to 90 percent removal, and since the benefits of such treatment change only occasionally (where the quality at some point in the river reaches 5 mg/l) the model will recommend that the removal level be raised considerably, perhaps all the way to 90 percent (point B in figure 3), or above. The model will not consider small, delicate changes in the decision but will jump abruptly from one point where incremental benefits or costs change to another such point.

[18] For an instructive discussion of the technical peculiarities of linear models see Baumol and Bushnell [1a].

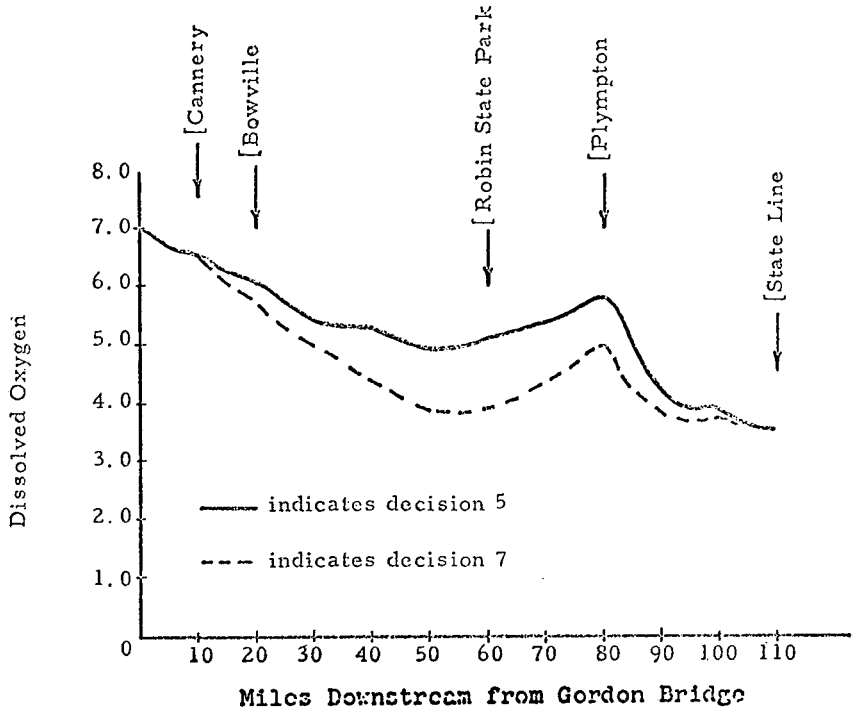


FIGURE 4.—**Water quality in the Bow River, under alternative decisions, summer drought conditions.**

In some circumstances this characteristic of linear programing models leads to unrealistic and even ridiculous results. But in other circumstances, and this model is a case in point, it is realistic for the model not to incorporate a finer degree of discrimination than the institutional system that it analyzes is likely to display. In actuality a commission or other political decision process is likely to focus attention on a few sensible possible decisions, corresponding to actions at which costs or benefits change markedly, and to make its selection from those. In this respect a linear programing model reflects institutional behavior.

The conditions that lead to each of the possible decisions can be found in tables 7 and 8. The tables are read by locating the column associated with a particular assignment of influence weights, and then reading down the column to find the desired use classification. The row in which the desired use classification is found contains the number of the decision called for under those conditions and also the net benefits to each of the participants which result from that decision and the net costs to the two cities and the FWPCA.

TABLE 7.—DECISIONS CORRESPONDING TO SPECIFIED QUALITY STANDARDS AND WEIGHT ALLOCATIONS (NO RESTRICTION ON CANNERY TREATMENT COST)

[Dollar amounts in thousands per year]

Decision	Net cost				Net benefits			Weight allocation [1]					
	Cannery	Bowville	Plympton	FWPCA	Bowville	Plympton	FWPCA	1, 7, 1, 1 4, 4, 1, 1 3, 3, 1, 3	1, 4, 4, 1 0, 0, 0, 1	3, 1, 3, 3	1, 1, 7, 1 1, 2, 6, 1	4, 1, 4, 1	7, 1, 1, 1
1	$95	$271	$367	$1,009	0	0	0	D					
2	95	445	315	1,172	−$125	$172	$54	C					
3	95	495	300	1,219	−161	221	68		D (3.92) C (3.92)				
4	95	651	249	1,359	−281	324	50	B	B	D (5.0) C (5.0) B			
5	95	659	247	1,366	−288	328	46				D (5.05) C (5.05) B (5.05)		
6	90	659	258	1,362	−290	324	47					D (5.0) C (5.0) B	B
7	8	654	304	1,291	−336	218	−17						D (3.95) C (3.95)

[1] Figures in parentheses are dissolved oxygen levels at the park, where they exceed the minimum required by the standard.

The net benefits to the cannery are simply the negative of its net cost since its gross benefits from stream improvement are taken to be nil. The net cost figures are the costs as sensed by the individual participants; that is, allowing for the effects of taxes in the case of the cannery and of grants-in-aid in the case of the two cities. Table 7 shows the decisions when there is no limit on the level of treatment to be required of the cannery. Table 8 shows data for the cases in which the commission was assumed unable to require the cannery to employ tertiary treatment.

It will be noted in these tables that the net benefits to Bowville are invariably negative. This result should not be misinterpreted; in particular it does not mean that Bowville will inevitably be worse off as a result of the commission's activities than it would be if the river were left in its current mediocre state. Many of the advantages of the generally improved environment, in which Bowville's residents will share, escape our crude methods of benefit measurement which are limited, really, to expressing the relative importance to Bowville of improvement to different portions of the river. We have no method of estimating the overall worth to Bowville or anyone else of cleaning up the 100 miles or so of river. In the circumstances all we can measure is the relative worth of different plans, all of which bring the river up to aesthetically satisfying standards (at least Class D).

TABLE 8.—DECISIONS CORRESPONDING TO SPECIFIED QUALITY STANDARDS AND WEIGHT ALLOCATIONS (CANNERY TREATMENT LIMITED TO HIGH EFFICIENCY SECONDARY)

[Dollar amounts in thousands per year]

	Net cost				Net benefits			Weight allocation[1]			
Decision	Cannery	Bowville	Plymp-ton	FWPCA	Bowville	Plymp-ton	FWPCA	1, 7, 1, 1 4, 4, 1, 1 3, 3, 1, 3	1, 4, 4, 1 0, 0, 0, 1	1, 1, 7, 1 1, 2, 6, 1 4, 1, 4, 1 3, 1, 3, 3	7, 1, 1, 1
8	$35	$345	$380	$1,025	−$79	−$21	−$29	D			
9	35	526	325	1,194	−211	156	21	C			
10	35	594	304	1,255	−264	219	30		D (3.97) C (3.97)		
11	35	659	283	1,314	−314	262	23			D (4.42) C (4.42)	
7	8	654	304	1,291	−336	218	−17				D (3.95) C (3.95)
12	35	1,230	246	2,027	−867	329	−623	B	B	B	B

[1] Figures in parentheses are dissolved oxygen levels at the park, where they exceed the minimum required by the standard.

In expressing our results we are therefore compelled to measure net benefits from some arbitrary base of comparison. For this purpose we have chosen the plan that attains Class D standards throughout the valley at the lowest possible economic cost. Therefore, a negative net benefit to Bowville means only that Bowville is worse off under the plan that gave rise to it than it would be under the "least cost" plan for achieving Class D standards. Bowville, as an upstream user with high quality water to begin with, is always the loser when this minimum level of improvement is exceeded.

To illustrate the interpretation of the tables, suppose that the quality standard is set at 2 mg/l (Class D) and that Bowville's interests are favored by employing one of the weighting schemes (3, 3, 1, 3), (4, 4, 1, 1), or (1, 7, 1, 1). Then table 7 shows that decision 1 is ap-

propriate. Notice that this decision corresponds not only to these three different weighting schemes but also to the least economic cost plan. Therefore all net costs and net benefits are recorded as zero.

If, for the same quality standard, the FWPCA's viewpoint were to rule (weighting scheme 0, 0, 0, 1), decision 3 would be called for. The decision would be the same if Class C usage were adopted because it provides a water quality that exceeds Class C's minimum (for example, 3.92 mg/l of dissolved oxygen at the park) even though such high quality was not imposed in advance. The voluntarily high quality is a consequence of the high valuation imputed to the FWPCA for improvements in water quality at the park. On the other hand, if Plympton's interests are favored, as with weights (3, 1, 3, 3), even if the standard is set at 2mg/l a quality of 5 at the park is attained, for much the same reason.

5. Conclusions

At the very outset we raised the issue of whether a complicated problem of governmental decision in which the interests of influential groups were in conflict could be expressed as a formal model and could be analyzed fruitfully in those terms. We claimed that formal, objective analysis is possible, and that it can make three significant contributions to understanding the problem. First it can provide a framework for organizing the political, social, economic, and technical considerations involved in the problem and can use these data, coherently organized, to predict a range of outcomes within which the ultimate decision would be likely to fall. We did not claim that formal analysis could predict a unique outcome of the governmental process but, and this was our second major claim, we asserted that the analysis would reveal the forces and considerations that tend to move the decision toward one portion or another of the range of likely outcomes. Our third claim was that formal analysis can facilitate the comparison of alternative decisions by indicating vividly who gains and who loses, and by how much, in moving from one of them to another.

Now we can assess those claims in the light of the test just performed. Admittedly the test was conducted in a world that we ourselves have made. But this especially constructed world does contain many of the essential ingredients and perplexities of the real world. Therefore it seems reasonable to feel that our experiment has some value in indicating the characteristics and usefulness of the type of analysis here advocated.

5.1 *Appraisal of the Test*

The most significant and obvious result of the test is that it was carried through successfully. There is a widespread tradition to the effect that human, moral, and political affairs are too subtle to be subjugated to the discipline of numbers.[19] On a more pragmatic level it might be feared that political and psychological data are too elusive to be quantified, and indeed we were not able to find an observable measure of political influence and had some difficulty in quantifying the preferences of the participants in our test case. Nevertheless

[19] Rousseau provides a typical example. At the end of a passage in which he indulged in some quantitative modes of expression, he hastened to retract with, "If, to save words, I borrow for a moment the terms of geometry. I am nonetheless aware that moral quantities do not allow of geometrical accuracy." [10, p. 52].

we were able to begin with a governmental problem in all its vagueness and ambiguity, and to refine and reorganize the highly diverse data, ranging from the velocity and other hydrologic characteristics of the river to the objectives and preferences of the riparian residents, so that they could be incorporated in a tractable analytic model. The model itself rests on two guiding principles. The first is that no decision is acceptable if there exists an alternative that is more attractive to one of the interested persons or groups and no less attractive to any of the others. This is the principle of Pareto admissibility. The second is that the Pareto-admissible decision which will be taken in any particular case depends on the relative political influence accorded to the interested individuals and groups. The model operates by assuming a specific set of relative influence weights and determining the corresponding Pareto-admissible decision. By varying the assumed influence weights over a wide range, the spectrum of Pareto-admissible decisions can be determined, and was determined.

It turned out that there were surprisingly few choices in the range of likely outcomes, even with a wide variety of assumptions about the relative political influence of the different groups concerned. Without performing the analysis we should not have been able to foresee that the range of Pareto-admissible decisions would be so restricted. The calculation of the range of likely outcomes substantiated our first claim.

Our second claim held that the analysis would reveal the political circumstances that favor adoption of one or another of the admissible decisions. Tables 7 and 8 confirm this contention. Notice that the columns are ordered so that the predominance of political influence moves from upstream (the cannery and Bowville) to downstream (Plympton) as you move from left to right. Notice also that as you move across the table from left to right the decision that is predicted moves down the list of alternatives, from decisions most favorable to Bowville to those that favor Plympton. This is no profound discovery; it is only commonsense, or perhaps tautology, that as the political influence of any participant grows he will be able to wangle decisions more to his liking at the expense of the others. But the analysis has quantified and sharpened this vague and obvious perception, and has shown concretely how the response to a change in influence is likely to be implemented and how it will effect the welfares of the individual participants.

In this regard, in fact, the analysis went a bit beyond our expectations. It showed that some decisions that are admissible according to the formal definition of the concept require such extreme and implausible distributions of political influence that they can hardly be expected to arise. For example, decision No. 7 was taken only when the cannery was assigned seven times as much political weight as any of the other participants. It does not seem reasonable socially or sustainable politically for a special interest group to have such a predominant influence on a political decision. Therefore a clear implication of the analysis is that decision No. 7 will not be taken.

Our third claim was that analysis using a formal model would facilitate comparison of the social and economic consequences of different possible decisions. A comparison of decisions Nos. 4 and 12 provides an apt illustration. Decision No. 12 is taken only when there is an upper limit to the level of treatment that can be required of

the cannery and when quality standard B is to be attained. Decision No. 4 achieves that same quality level, and a shift from decision No. 12 to decision No. 4 would save Bowville almost $10 for every dollar that it cost the cannery while increasing costs to Plympton only slightly. It is clearly very costly to the other participants to impose an upper limit on the canner's treatment costs.

Further study of the tables will disclose other illuminating trade-offs among different policies that achieve the same water quality standard.

In addition to confirming our claims, the analysis calls our attention to the value of some social devices that we assumed to be forbidden to the Bow Valley Water Pollution Control Commission. The most important of these is the authority to tax and to offer grants-in-aid. If these expedients were available, the wasteful decision No. 12 would never be attractive: the commission could then tax Bowville and use the proceeds to reduce the financial burden on the cannery. In this way the technically efficient decision No. 4 could be implemented without imposing undue financial hardship, to everybody's benefit. The value of such authority, though it might have been suspected, could not have been established without invoking a model that brings out the economic consequences of the technical peculiarities of the river that make the cannery the strategic site for treating waste.

This finding illustrates that our concept of Pareto admissibility is relative. It depends heavily on the range of decisions that are assumed to be technically and legally practicable. Change those assumptions, and the Pareto-admissible set changes correspondingly. Thus the model can be used to estimate the value of changing either the technical possibilities available (by research) or the legal possibilities permissible (by legislation), by comparing the Pareto-admissible sets that correspond to alternative assumptions. Among the legal possibilities that merit consideration are the power to tax and make grants, the authority to impose effluent charges, and the authority to operate regional treatment facilities. All of these are issues now being debated earnestly in the several water pollution control commissions that are currently operative.

5.2 *Applicability of the Model*

Like any model, this one depends on numerous assumptions and can be applied only to circumstances that correspond fairly well to the assumptions. The fundamental assumption in this model is that a water pollution control commission, like any other government agency, is responsive to the wishes of its constituency. This seems highly reasonable. By being responsive an agency reduces its exposure to complaints, litigation and animosity, builds its reputation for efficiency and fairmindedness, accumulates political support and influence, and fosters cooperative attitudes among the people with whom it must deal. In short, it gains the consent of the governed, which is an essential prerequisite of effective government. In keeping with this assumption we have portrayed the Bow Valley Water Pollution Control Commission as consulting continually with its constituents and as endeavoring to formulate a plan that will be as agreeable to them, individually, as possible. Any government agency or commission that acts according to this simple principle will arrive

at one or another of the decisions that we have called Pareto admissible.

The kind of analysis that follows from this assumption stands in contrast to the approach used in much of the literature, which rests on the postulate that Government agencies endeavor to achieve some overarching goals which typically are called the general welfare or the national interest. Such an agency, if such there be, would be unresponsive to the sectional or special interests of its constituents. There is much empirical evidence that agencies in fact behave predominantly as we have assumed. See, for example, the work of Matthew Holden [5].

Furthermore, the institutional arrangements that are employed by water pollution control agencies in order to be responsive to their constituents are frequently similar to those used by the Bow Valley Water Pollution Control Commission. This is no coincidence. We have modeled our commission in the image of some established pollution control commissions. The Delaware River Basin Commission (DRBC) is made up of the Governors of the four States through which the Delaware flows, plus a representative of the Federal Government, just as the Bow Valley Water Pollution Control Commission is made up of representatives from the municipalities and industries located along the river. The Ohio River Valley Water Sanitation Commission (ORSANCO) has a similar constitution. It is comprised of representatives from all the States through which that river flows along with a representative of the Federal Government.

The formal composition of the commission is not essential, however. What is essential is an anxious concern for the interests and responses of the people who are affected by the commission's decisions. Direct representation is one expedient for assuring this concern. A more fundamental expedient is the maintenance of intimate contact with the constituency by both informal consultations and formal conferences and public hearings before any important decision is made. Both the DRBC and the ORSANCO follow such procedures, as does the Bow Valley Water Pollution Control Commission. Indeed, the Federal legislation that provides the charter for all the newer commissions requires elaborate consultations. The requirement to conduct public hearings at which all affected "interstate agencies, States, municipalities and industries involved" can present their cases is stressed repeatedly in the Federal laws. Operating practice, so far as we have been able to discern it, sincerely implements these requirements. In these respects the Bow Valley Water Pollution Control Commission appears to be a fair replica of the pollution control agencies now extant or envisaged.

A subsidiary assumption that we made was that the constituency of the Bow Valley Water Pollution Control Commission consists entirely of local residents except for a generalized public interest represented by the FWPCA. This characteristic is by no means universal, even in the field of water pollution control. In that area Federal legislation contemplates at least three different kinds of agency for the detailed administration of water pollution control. The Federal Water Pollution Control Act authorizes the establishment of interstate compacts for the control of pollution in rivers and estuaries that flow through several states. ORSANCO and the DRBC are leading examples of such compacts. The Water Quality Act of 1965

gave primary responsibility for the regulation of interstate waters to "State water pollution control agencies" which are simply departments of the State governments, either State health departments or other appropriate State agencies. The Clean Water Restoration Act of 1966 encouraged the establishment of basin planning agencies with jurisdictions similar to that of the Bow Valley Water Pollution Control Commission. These three different types of agencies have therefore different constituencies, and the question arises as to how far their behavior will resemble the behavior that we have hypothesized for the Bow Valley Water Pollution Control Commission.

The local basin planning agencies are required by legislation to have "representation of appropriate State, interstate, local or (when appropriate) international interests in the basin." The Bow Valley Water Pollution Control Commission, accordingly, has been concocted to conform to the requirements of Federal law. Real basin planning agencies may be expected to behave much as the Bow Valley Water Pollution Control Commission did.

The State water pollution control agencies, which are entrusted with the establishment and enforcement of water quality criteria, add a new dimension to the problem. They are concerned with the interests of their entire States and not merely with those of industries and municipalities in the affected valley. They must take into account the effect of regulations in one river valley on the welfare and interest of residents elsewhere in the State. For example, if they favor industries located in one valley by permitting comparatively low levels of treatment there, they will be sure to hear about it from the representatives of competing localities. Furthermore, they may wish to have a variety of water qualities in the State, permitting heavy industrial use of some rivers while reserving others for recreational and esthetic purposes. Thus a State agency confronting the problem of the Bow River Valley might have a wider range of considerations in mind than we have incorporated in our model.

This would complicate the problem but would not alter it fundamentally. A State agency, like a basin planning agency, has a constituency to which it is responsive. The difference is that even when it deals with the regulation of a single river basin, it may be affecting the interest of constituents who are not local residents. A model of such a State agency considering regulations for a particular river would therefore have to include participants from outside the local basin, and to that extent would be more complicated than the model that we have presented. In fact, it would be a great deal more complicated because the interests of participants who do not reside or operate in the basin are likely to be much harder to quantify than the interests of local residents.

When a State water pollution control agency is making decisions about a particular river basin, it confronts two different kinds of issue, which we may call local and statewide. The local issues are those that affect only the residents of the basin concerned; the statewide issues are those that affect also the interests of residents of the State who do not live in that particular river basin. An example of a statewide issue would be the determination of the water use classification of a river or a reach of a river. An example of a local issue is the allocation of the total permissible discharge of pollutants among the towns and other

polluters located in the basin. The model that we have presented represents adequately the pressures that even a State agency would feel in conjunction with a local issue, for the only interests that would be felt are those of the immediately affected firms and municipalities. But the considerations that would be involved in reaching a decision about a statewide issue not only involve more participants but are of a more subtle sort and perhaps more difficult to quantify. For this reason the application of the model to State agencies remains a subject for further investigation.

A commission created by interstate compact is structurally intermediate between a very large local basin agency and a State pollution control agency. It resembles a large local authority in that its jurisdiction is confined to a single river or estuary and its tributaries. It is like a statewide agency in that the commissioners who determine its policy represent many constituents who do not reside in the basin that is controlled and do not directly use its water. In practice the commissioners (apart from the Federal representataive) on both ORSANCO and the DRBC, the two leading examples, seem to have been preoccupied with predominantly local pressures and concerns, and to have behaved much like the Bow Valley Water Pollution Control Commission writ large. The analytic apparatus here proposed seems to be highly applicable to the major decisions that they have taken.

Some specialized assumptions were made purely to keep the illustrative computations simple and tractable. The most important of these was the narrow range of discretion permitted the Bow Valley Water Pollution Control Commission. Many actual commissions have a scarcely wider range of authority but, as we noted above, additional powers and privileges can permit superior decisions (in the Pareto-admissible sense) and some authorities have them. Additional policy instruments can be incorporated into the model at the cost of some increase in complexity.

In the interest of simplicity, also, we vastly simplified the hydrologic and technological aspects of the problem. In particular we paid no attention to the unpredictability and variability of stream flows, or to any of the other aspects of uncertainty that are a vexatious part of all real decision problems. There are decision models, of course, that incorporate these complications and they are invariably far more elaborate than ours. To introduce uncertainly in its manifold aspects we should have to modify our model along well-established lines. To have done so in the present model would have been to obscure its main intent, which was to test the feasibility of comprehending both political and technical considerations in a model of a governmental decision process. For practical application, of course, the model would have to be extended so as to allow for uncertainty.

We also evaded dynamical considerations by assuming that neither the towns nor the cannery were expected to grow, and that no new industries were likely to be established along the river. The possibility of growth presents serious analytic difficulties of its own. Any satisfactory method for dealing with them could most likely be built into our model, but we have not allowed our attention to be distracted by this separate, very important, and very difficult analytic problem.

Finally we must admit to having ignored two important features of the political decision process. One is the influence of log-rolling,

pressure politics, and side payments of all sorts. There are pressures, threats, and inducements that Bowville can use to persuade the cannery to acquiesce in a decision that would otherwise be unacceptable, and vice versa. Even Bowville and Plympton, though their interests in the treatment of the river are almost diametrically opposed, can bargain a bit out in the corridor: if Plympton will moderate her demands for waste treatment, Bowville will be more agreeable in sharing the burden of maintaining the county roads. Such bargaining is an essential part of practical politics. It can shift the outcome of the decision process from one Pareto-admissible decision to another. Thus it can influence where in the range of likely outcomes the ultimate decision will come to fall. It reinforces the finding that our prediction is a range rather than a point. But such side bargaining does not upset our prediction.

Our other simplification is, in principle, more fundamental. We assumed that each of the participants had a firm and immutable evaluation of the consequences of every decision for his own welfare. This assumption was contained in our estimation of the values attached by the participants to increments in water quality at different points in the river, and in our treatment of those evaluations as unchanged throughout the decision process. In fact, a significant feature of any group decision process is the attempt by each participant to persuade the others to alter their psychological evaluations so that they are more in line with his. The representatives of Plympton and the Izaak Walton League will emphasize the benefits to everyone of high quality water throughout the river and, in fact, will urge it as a moral imperative. The representatives of the cannery will point out that the prosperity of the entire basin depends on the economical use of the waste absorption potential of the river. To the extent that this rhetoric is not in vain it will succeed in causing some of the participants to change their subjective valuations of the importance of improvements in water quality. It will make our table 2 invalid.

How serious is this difficulty? This is a significant and open question, fundamental to the understanding of the political process. Perusal of the histories of ORSANCO and the DRBC suggests, however, that when specific decisions are being debated most of the rhetoric is ineffectual, that people want at the end much what they wanted at the beginning, and that the operative aspect of the bargaining process is a reconciliation of the pressures that the different interest groups have been able to mobilize.[20]

5.3 *Comparison with Current Practice*

The prevalent method for analyzing public policy decisions of the sort dealt with here is benefit-cost analysis. Superficially, benefit-cost analysis applies to the government sector the calculus of profit and loss that is used in business decisions. Its popularity is at least partly a response to the businesslike ethic that prevails in our culture. But its contrast with the approach here advocated is, in fact, more profound than the question of whether the Government should follow businesslike practices. It is really a reflection of the most ancient cleavage in the tradition of political philosophy.

[20] For an instructive account of the issues faced by ORSANCO and their resolution see Cleary [2].

One great school of political philosophy views a government as the leader of its people, responsible for defining the goals of its citizens and formulating their social standards, preferably under the wise guidance of a philosopher-king or benevolent despot. The other great school sees the government as the corporate embodiment of its people, serving their communal interests and carrying out their wishes, preferably as expressed in direct (nowadays, participatory) democracy.

Traditionally these two views have been advanced as norms, as expressions of what governments ought to be. But they deserve also to be taken seriously as expressions of how governments actually behave. Examples of both governmental leadership and responsiveness are easy to find. ORSANCO, which provides the best documented experience in the water pollution area thanks to Cleary, was established in response to public demand (that is, the demand of a few leading private citizens who mobilized widespread support) and went on to exercise a good deal of leadership and initiative of its own. It did so, however, less by pursuing its own goals irregardless than by undertaking a series of educational activities that increased its constituents' awareness of the importance of abating the pollution in the Ohio.

We cannot pursue here the rather ill-defined issue of leadership versus responsiveness. In practice, Government agencies appear to mix the two in varying proportions, with responsiveness preponderating except for transitory episodes, usually at the highest levels of government.[21]

Benefit-cost analysis is, however, an expression of the leadership role of government. A benefit-cost formula is a tool for evaluating the desirability of different undertakings by the Government's standards. This has long been recognized, though not so frequently articulated. For example, one of the chronic condundrums in benefit-cost analysis is the choice of the rate of discount to be used in evaluating deferring benefits and costs. This rate, selected by agency officials or expert consultants, represents the official evaluation of the relative importance of consequences that emerge at different dates. Similarly the Flood Control Act of 1936 instructs agencies to compute benefit-cost ratios by adding up the benefits and costs of a project "to whomsoever they may accrue." The intent is to maintain neutrality, but the effect is to impose a judgment about the relative social importance of effects upon upstream users and downstream users, rich and poor, farmers and urbanites, and so on. All governmental undertakings redistribute income in some manner, as we have seen in the case of the Bow Valley.[22] Any evaluative formula must incorporate some appraisal of redistributive consequences, either implicitly or explicitly. A third example of the governmental evaluations built into benefit-cost analyses is provided by the problem of aggregating benefits, and costs, of different kinds. When the beneficial results are priced on economic markets, as is the case with irrigation water and hydroelectric power, the market prices are used. Otherwise prices representing social evaluations have to be found. Outdoor recreation benefits, for example, are frequently

[21] Truman [11] presents a full-dress analysis of governmental responsiveness.

[22] This problem is discussed more completely by Marglin [9, pp. 67 ff.]. Marglin recommends that different values be assigned to different consequences, dependent on the beneficiary. For a full treatment of the theoretical significance of income redistribution see Little [7]. See also the papers by Weisbrod, Bonnen, and Freeman, in this volume.

valued at $1.50 per user day but some authorities insist that use by comparatively deprived urban dwellers should be assigned higher "merit" values.[23] Clearly some values have to be used, and any values represent the judgment of the agency that adopts them.

Finally, some consequences of Government undertakings, which are deemed excessively difficult to evaluate, are simply omitted from benefit-cost calculations. This, too, represents an implicit governmental judgment, and one that has drawn much criticism.[24] In sum, there is no way out of it: a benefit-cost formula incorporates many judgments, implicit and explicit, of the relative importance of the numerous diverse consequences of the undertakings being evaluated. These judgments must be those of the agency doing the evaluation or its superiors in the governmental hierarchy.

In fundamental contrast to the benefit-cost approach, the analysis used here invokes no other evaluations than those of the people affected. It is explicitly noncommital with respect to the relative importance or influence of the different participants; that is why it does not lead to an unambiguous prediction. It assembles the data from which those people and the agency concerned derive their decisions, but it does not presume that anyone has a formula for global social evaluation. Therefore it does not purport to recommend what should be done, but only to describe how actors in a political process interact to produce a decision.

This analysis cannot be regarded as an alternative, even in principle, to the decisionmaking methods actually used. But it can be of assistance in understanding and even facilitating those methods. Its advantage lies in its ability quickly and cheaply to sketch out the range of alternative decisions that is worth considering. In our test case, with the data and a moderately fast computer at hand, between 2 and 3 minutes were required to determine the corresponding admissible decision. By varying the assumed data artfully, the main outlines of the entire range of admissible decisions were mapped out with about 60 repetitions of this quick computation.

Of course, the computations that would be required in a practical instance are of an entirely different order of magnitude from those encountered in this simple test. Contemplate any actual river basin. It would contain a half-dozen or more cities and towns, several dozen factories or other points of waste discharge, a main stream, and a number of tributaries with complicated hydrology. The alternative technologies of waste treatment, of manufacturing process variation, and of waste water recirculation would be rich and varied. Water quality standards would take account of numerous characteristics of the water instead of only one. The simple waste transport and decay equations that we have used would be regarded as inadequate approximations. When all these complications are taken into account, the calculations would not only be much larger than those that we have encountered but would be beyond the capacity of any computer now extant or envisaged.

---

[23] Mack and Myers [7] is a careful analysis of the problem of evaluating recreational benefits from a social point of view. There is much additional literature, some cited by them.

[24] These criticisms are reviewed in Dorfman [3] and elsewhere in the volume in which it appears. See also the paper by Margolis in this volume, and the paper by Knetsch in vol. 3 of this collection.

These complexities, however, do not render mathematical analysis inapplicable. They do necessitate a good deal of simplification of the full richness of reality. Simplifying assumptions would have to be made about waste treatment technology, though probably not as severe as the simplifications that we have indulged in. A water quality index of one or two dimensions would have to be introduced in place of the multidimensional specifications set forth in water quality standards documents. The hydrology would have to be simplified. The number of points of pollution would have to be reduced by consolidating groups of nearby installations into a single synthetic polluter. All these and other simplifications would have to be carried to the point where the calculation became manageable.

The result would be an inevitable loss in accuracy. But this necessity does not invalidate the method, for the relevant standard of accuracy is not some unattainable ideal but the level of accuracy that is attainable by alternative procedures. The truth is that the economic-hydrologic-biologic-political ecology of a live river basin in the full majesty of its intricacy far transcends the capacity of any method of analysis or decisionmaking available to man. All methods of decisionmaking require severe simplifications, as perusal of the dockets of any water control authority will establish. And, there is reason to believe, the simplifications required for mathematical analysis are less disabling than the simplifications that are conventional in more informal procedures. Only hard experience can determine how practical and helpful mathematical analyses will be in actual instances, but the fact that they must invoke some serious simplifications is not ipso facto decisive.

REFERENCES

[1] Baumol, W. J. and R. C. Bushnell, "Error Produced by Linearization in Mathematical Programming," *Econometrica*, 35 (July–October 1967), 447–471.

[2] Cleary, Edward J. *The ORSANCO Story—Water Quality Management in the Ohio Valley under an Interstate Compact* (Baltimore: Johns Hopkins Press, 1967)

[3] Dorfman, R., "Introduction" in R. Dorfman, ed., *Measuring Benefits of Government Investments* (Washington: Brookings, 1965), pp. 1–11.

[4] Fair, G. M., J. C. Geyer, and D. A. Okun, *Water and Wastewater Engineering*, 2 (New York: John Wiley and Sons, 1968).

[5] Holden, Matthew, Jr., "Political Control as a Bargaining Process: An Essay on Regulatory Decision-Making," pub. no. 9, Cornell University Water Resources Center (October 1966).

[6] Kneese, A. V. and B. T. Bower, *Managing Water Quality: Economics, Technology, Institutions* (Baltimore: Johns Hopkins Press, 1968).

[7] Little, I. M. D., *A Critique of Welfare Economics*, 2d, ed. (London: Oxford University Press, 1957).

[8] Mack, R. P. and S. Myers, "Outdoor Recreation," in *Measuring Benefits of Government Investments*, R. Dorfman, ed. (Washington: Brookings, 1965), pp. 71–101.

[9] Marglin, S. A. "Objectives of Water Resource Development: A General Statement," *Design of Water-Resource Systems*, Arthur Maass et al. (Cambridge: Harvard University Press, 1962), chapter 2.

[10] Rousseau, J. J., *The Social Contract and Discourses* (Everyman's ed.), (London: J. M. Dent; New York: E. P. Dutton).

[11] Truman, David B., *The Governmental Process: Political Interests and Public Opinion* (New York: Alfred A. Knopf, 1951).

TECHNICAL APPENDIX

In the text of this paper, we presented a model of political decisions and its application to the Bow River example. Insofar as possible, matters of technical detail were avoided. There are, however, two technical aspects of the model that warrant discussion. One is the link between abatement action taken by a particular polluter and the quality response in the river. We shall review briefly the processes involved in organic pollution of flowing waters, and the mathematical expression of these processes which we used to calculate the transfer coefficients, d_{ij}. The other aspect is the linear programing formulation that we used to calculate Pareto-admissible solutions to the river quality management problem.

A–1. *The Streeter-Phelps Model of DO Concentration*

In a stream or estuary in its natural state, unaffected by organic wastes, the amount of oxygen dissolved in the water will be near saturation. The saturation value varies with temperature, but under summer conditions a typical stream might contain around 8.5 mg/l of dissolved oxygen (DO). When organic wastes are introduced, a process of biochemical decomposition takes place which utilizes the oxygen in the water and reduces the concentration of dissolved oxygen. The higher the rate of waste discharge, the greater the impact. In our example, we use DO as the indicator of instream water quality. By the same token, we measure the quantity of waste discharged by each polluter in terms of the quantity of oxygen which will be used up in the process of decomposition; that is, in pounds of biochemical oxygen demand or BOD.[1]

Of course, whenever the oxygen concentration in the stream is below saturation, a process of "re-oxygenation" or "re-aeration" takes place whereby additional oxygen is absorbed into the water from the atmosphere. The interaction between biochemical degradation and re-aeration determines how far dissolved oxygen falls below the saturation value.[2] So long as the oxygen level in the water stays above zero, aerobic decomposition continues. If the oxygen level falls to zero, anaerobic decomposition takes over.

[1] A full discussion of these concepts can be found in Fair, Geyer, and Okun [4, chapter 33].

[2] The oxygen released by plant life during photosynthesis is ignored in this simplified treatment.

Aerobic decomposition is normally thought of as taking place in two stages. In the first stage, carbonaceous matter is oxidized. In the second stage, which begins only after the first stage process is well underway, nitrogenous material is oxidized as well. The impact of a particular effluent on water quality, therefore, depends not only on the total BOD discharged but on the composition of the waste, and the discharges of the three participants in the Bow River example were stated in terms of a mix of carbonaceous and nitrogenous BOD.

Looking first at carbonaceous BOD only, it is possible to represent the impact of a waste source on stream dissolved oxygen by means of a simple linear differential equation. In order to show this, let q be the dissolved oxygen level and q^s the saturation value, and let y be the DO deficit; that is, $y=q^s-q$. If L is the rate of BOD discharge and the receiving stream has a flow rate of F, then L/F is the BOD concentration in the receiving stream at the point of waste discharge. Consider a specific volume of water containing an initial BOD concentration of L/F, flowing down the river. The waste load will decompose gradually. The concentration remaining t minutes after the injection of the load will be denoted by L_t/F. The rate at which this decomposition takes place at any instant is proportional to the BOD load remaining, L_t; that is,

$$\frac{d}{dt}\left(\frac{L_t}{F}\right)=-K_1\frac{L_t}{F} \tag{A-1}$$

where K_1 is an empirically determined constant, the value of which is a function of the water temperature and the particular characteristics of the waste. In integrated form,

$$\frac{L_t}{F}=\frac{L}{F}e^{-K_1t}, \tag{A-2}$$

i.e., the waste concentration in the stream decays exponentially over time. By definition, every gram of BOD decomposed absorbs one gram of dissolved oxygen. Therefore, the dissolved oxygen concentration is constantly reduced by decomposition at the same rate that the BOD concentration is falling, or

$$\left[\frac{dq}{dt}\right]_{deox}=-K_1\frac{L_t}{F}=-\frac{L}{F}K_1e^{-K_1t}. \tag{A-3}$$

At the same time the oxygen concentration is being replenished by aeration. This process goes on at a rate proportional to the current dissolved oxygen deficit, so that

$$\left[\frac{dq}{dt}\right]_{reox}=K_2(q^s-q) \tag{A-4}$$

Adding the effects of these concurrent processes together,

$$\frac{dq}{dt}=\left[\frac{dq}{dt}\right]_{deox}+\left[\frac{dq}{dt}\right]_{reox},$$

we obtain the differential equation for dissolved oxygen concentration:

$$\frac{dq}{dt}=-\frac{L}{F}K_1e^{-K_1t}+K_2(q^s-q). \tag{A-5}$$

The integral of this equation is

$$q=q^s-\frac{LK_1}{F(K_2-K_1)}\left(e^{-K_1t}-e^{-K_2t}\right)-y_0\,e^{-K_2t}. \tag{A-6}$$

In words, equation (A-6) says that the quality (q) at any point in time after the introduction of a load L is the saturation value (q^s) less the combined influence of deoxygenation and re-oxygenation attributable to the load L and, finally, less the residual influence of any initial DO deficit (y_0). If the stream is flowing with a velocity V, then the quality response q occurs at a distance $m=Vt$ downstream from the point where L is entering, and equation (A-6) may be stated in spatial terms as

$$q=q^s-\frac{LK_1}{F(K_2-K_1)}\left(e^{-K_1\frac{m}{V}}-e^{-K_2\frac{m}{V}}\right)-y_0e^{-K_2\frac{m}{V}}. \tag{A-7}$$

For purposes of our model, we are concerned with the impact of polluter i located at point m_i on the river on some quality control point j located a distance (m_j-m_i) downstream. As can be seen in equation (A-7), if polluter i reduces his load, L_i, by one unit, the DO concentration at m_j will increase by

$$d_{ij}=\frac{\partial q_j}{\partial L_i}=-\frac{K_1}{F(K_2-K_1)}\left(e^{-K_1\frac{m_j-m_i}{V}}-e^{-K_2\frac{m_j-m_i}{V}}\right). \tag{A-8}$$

Equations (A-7) and (A-8) express the system response to a single source of carbonaceous BOD. As noted above, municipal and industrial effluents contain nitrogenous as well as carbonaceous matter, but the nitrification process begins only after some time lag. By virtue of the fact that the impacts of different waste sources on stream dissolved oxygen are additive, the effect of the nitrogenous component of the waste can be approximated by imagining it to be inserted by a separate dummy waste source some distance downstream. Carbonaceous and nitrogenous decomposition proceed according to different constants K_1, but the dissolved oxygen response to the removal of one unit of nitrogenous BOD is governed by equation (A-8). We shall denote the transfer coefficients for carbonaceous and nitrogenous BOD removed as d_{ij}^C and d_{ij}^N respectively.

The impact of different waste sources—e.g., the cannery, Bowville, and Plympton—also are additive, so that the increase in DO concentration at any point is the sum of terms of the form (A-8), two for each upstream waste source.

According to the assumptions we made about the hydrology of the Bow River, the constants used in our computations were: $K_1=.30$ for nitrogenous BOD, $K_1=.18$ for nitrogenous BOD, and $K_2=.40$. The velocity of the river under drought conditions was assumed to be 0.5 feet per second, and the nitrogenous reaction was assumed to start 20 miles downstream from the point of waste discharge.

A–2. *The Linear Programing Formulation*

In order to explore the range of Pareto-admissible choices, we combined the functional relationships and data introduced above into a linear programing problem. Recall that the decision to be made, X, includes the choice of a minimum permissible concentration of dissolved oxygen along the river, Q, and the degree of waste removal to be undertaken by each of the polluters x_i. Associated with any practicable decision is a net benefit to each of the participants, $NB^i(X)$. The linear programing algorithm is used to solve for the decision, X, which maximizes the weighted sum of net benefits,

$$\sum_i w_i NB^i(X),$$

subject to all the requirements of practicability. The resulting value of X indicates the Pareto-admissible decision which is consistent with any given set of influence weights.

The constraint set consists of three different groups of equations. First, there is a set of constraints that defines the relationship between treatment undertaken and quality improvement in the river. A second set of constraints defines the costs to the various interested parties, and a third set combines these cost relations with benefit data to yield the function.

We formulated the model as follows. Recall that the index i denotes the interested parties, and j the quality control points. The individual parties and geographical locations denoted by different values of i and j are summarized in table A–1.

TABLE A–1.—INDEXES DENOTING INDIVIDUAL PARTIES AND QUALITY CONTROL POINTS

Name	Index for— Interested party	Index for— Quality control point	River miles from Gordon Bridge
Cannery	i=1		10
Bowville	i=2	j=2	20
Robin Park		j=3	60
Plympton	i=4	j=4	80
State line		j=5	110
FWPCA	i=6		

The cannery, cities, park, and the State line are shown in figure 1. The FWPCA is an interested party but, of course, is not identified with any particular location in the valley.

Each polluter produces a waste stream containing, before treatment, amounts of carbonaceous and nitrogenous BOD equal to L_i^C and L_i^N respectively. For each discharger, the cost to cut back on pollution discharge is expressed as a piecewise linear function of the proportion of carbonaceous BOD removed, x_i. At the outset, all polluters have primary treatment facilities which remove 30 percent of the gross carbonaceous load, i.e., ex ante $x_i = .30$.

We denote each of the segments of the cost curve by the subscript k where the higher the value of k the higher the marginal cost of removal (i.e., numbering from left to the right in fig. 3). All of our treatment cost functions are divided into three segments, i.e.,

$k=1, 2, 3$. The breakpoint at the right-hand end of each segment, k, is labeled x_{ik}^*. Next we introduce the variables x_{ik} to denote the amount of carbonaceous BOD removed by methods along segment k of the cost function; these variables range from zero to the length of that segment of the function, i.e.,

$$0 \leq x_{ik} \leq x_{ik}^* - x_{i,k-1}^*, \qquad k=1, 2, 3. \tag{A-9}$$

As noted in the text above, the total proportion of carbonaceous waste removed is that taken out by primary treatment plus the sum of the amounts indicated by the x_{ik}, that is,

$$x_i = .30 + \sum_k x_{ik}.$$

Thus the total number of pounds of carbonaceous BOD removed by polluter i by additional treatment, over and above those taken out by existing primary facilities, is

$$\sum_k L_i^C x_{ik}.$$

Along any segment x_{ik} the proportion of nitrogenous BOD removed is related to the degree of carbonaceous removal by a simple factor of proportionality, a_{ik}. So the total nitrogenous removed, over and above that by primary treatment facilities, is (again in pounds)

$$\sum_k a_{ik} L_i^N x_{ik}.$$

The impact at point j as the result of the removal of a pound of carbonaceous or nitrogenous BOD at point i is denoted as d_{ij}^C and d_{ij}^N respectively, and these values are calculated using equation (A-8) above. Because the impacts of the carbonaceous and nitrogenous wastes are additive, the overall influence of a particular polluter's actions on a control point j downstream may be stated as

$$\sum_k (d_{ij}^C L_i^C + a_{ik} d_{ij}^N L_i^N) x_{ik}.$$

This sum yields the increase in DO concentration at point j resulting from the decrease in BOD discharge by polluter i that is designed by x_{ik}.

Next, we use the variable q_j to denote the quality (in mg/l of dissolved oxygen) attained at point j. This variable, too, we break down into two auxiliary variables, q_j^1 and q_j^2. For, just as treatment cost is a piecewise linear function of x_i, benefits are a piecewise function of q_j. Therefore we break q_j at the point q_j^* so that

$$q_j^1 + q_j^2 = q_j,$$

but

$$q_j^1 \leq q_j^*, \qquad j=2,3,4,5. \tag{A-10}$$

In our case, q_j^* is 5.0 mg/l (see table 4). As in the text, let $\bar{q}_j$ be the ex ante water quality at point j when only primary treatment is in use by all polluters.

In order to obtain the water quality q_j at point j, the x_{ik} pertaining to users upstream from j must satisfy:

$$-\sum_i\sum_k(d_{ij}{}^C L_i{}^C + a_{ik}d_{ij}{}^N L_i{}^N)x_{ik} + q_j{}^1 + q_j{}^2 = \bar{q}_j, \qquad j=2, 3, 4, 5. \tag{A-11}$$

Equation (A–11) states that the ex post DO concentration at point j, which we denote as $q_j{}^1+q_j{}^2$, is equal to the ex ante concentration ($\bar{q}_j$) plus the sum of the increases in DO that result from the waste treatment by all polluters upstream. Equations (A–9) and (A–10) simply define the upper bounds on certain auxiliary variables.

In addition, any practicable solution must attain a minimum water quality standard, Q. This may be expressed by another set of constraints of the form[3]

$$q_j{}^1 + q_j{}^2 \geq Q, \qquad j=2, 3, 4, 5. \tag{A-12}$$

Next, we must define the costs associated with each solution. For each segment of each cost curve there is a (constant) marginal cost of treatment which may be stated as a function of the proportion of carbonaceous waste removed. This marginal cost, as sensed by the individual polluter, we denote as c_{ik}. Then we denote the total additional cost incurred by polluter i (over and above primary treatment) as g_i where

$$\sum_k c_{ik}x_{ik} - g_i = 0, \qquad i=1,2,4. \tag{A-13}$$

An example of this calculation is given in section 3.1 above. We assume, on the other hand, that the FWPCA is concerned about the national cost of abatement measures, and the marginal national income cost of additional treatment by polluter i we call n_{ik}. Then the total cost as seen by the FWPCA, g_6, is defined by

$$\sum_i\sum_k n_{ik}x_{ik} - g_6 = 0. \tag{A-14}$$

Finally, the marginal value to participant i of an increment in water quality at point j is indicated by v_{ij}. As noted in table 4, the value of v_{ij} differs as between increments in water quality in the ranges above and below 5.0 mg/l (our q_j^* above). If we let v^1 denote the marginal value of quality increments below q^* and v^2 indicate the same for increments in the range above q^*, then the benefit to participant i of improvement at quality control point j is

$$v_{ij}{}^1 q_j{}^1 + v_{ij}{}^2 q_j{}^2.$$

[3] As noted in the text above, in all our calculations, q_5, the quality at the State line, was set at 3.5 mg/l. At all other quality points, the minimum quality level was set—in alternative solutions—at $Q=2$, $Q=3.5$, and $Q=5$.

Of course, the total benefit to participant i is the sum of the benefits attained by virtue of improved water quality at each of the points, j, along the river. And when costs are netted out from this total benefit the result is the net benefit, NB_i, which he receives, i.e.,

$$NB_i = \sum_j (v_{ij}{}^1 q_j{}^1 + v_{ij}{}^2 q_j{}^2) - g_i.$$

It is permissible, and indeed nearly inevitable where there are upstream and downstream users, for some of the NB_i to be negative, as Bowville's net benefits turned out to be. There is, however, a limit imposed by political practicality on the amount of sacrifice in net benefit terms that can be required of any polluter. This limit for polluter i is denoted by D_i, and we require that in all cases $NB_i \geq -D_i$.

The values of D_i, $i=1, 2, 4, 6$, were set equal to \$1 million in all computations where the cannery costs were unlimited. These values of D_i yield decisions 1 through 7 in table 7. To restrict the costs which could be imposed on the cannery, we set $D_1 = \$35{,}300$, leaving all other values of D_i unchanged. The revised value of D_1 yields decisions 8 through 12 in table 8.

To incorporate this condition in the model it is convenient to introduce new variables, XB_i, defined by $XB_i = NB_i + D_i$, to denote the excess of net benefits over the minimum permissible amount. Then

$$\sum_j (v_{ij}{}^1 q_j{}^1 + v_{ij}{}^2 q_j{}^2) - g_i - XB_i = D_i, \qquad i=1, 2, 4, 6. \quad \text{(A–15)}$$

It is obvious that the decision that maximizes $\Sigma_i w_i XB_i$ also maximizes $\Sigma_i w_i NB_i$. Therefore, the variables XB_i were used in place of the NB_i in the computations.

The entire programing problem is then to choose the 26 variables $Q, x_{ik}, q^1{}_j, q^2{}_j, g_i, g_6, XB_i$, and XB_6 for $i=1, 2, 4$, $j=2, 3, 4, 5$, $k=1, 2, 3$ so as to maximize the sum of the excess benefits for a specific set of weights $\Sigma_i w_i XB_i$, subject to the 29 equality and inequality conditions (A–9) to (A–15) and to the requirement that none of the choice variables can be negative. The solution to this standard linear programing problem will always be a Pareto-admissible decision that satisfies all the technical and legal requirements that have been imposed.

Section B

The Efficiency of Government Expenditures: The Pricing of Public Outputs and the Evaluation of Their Worth

EFFICIENCY GOALS, MARKET FAILURE, AND THE SUBSTITUTION OF PUBLIC FOR PRIVATE ACTION

BY JOHN V. KRUTILLA*

John V. Krutilla is Senior Research Associate at Resources for the Future, Inc.

In this paper, Dr. Krutilla examines the question of whether public investments which are undertaken to improve the allocative efficiency of the market necessarily attain that objective. He points out that while the failure of the market, to reflect the social worth of certain outputs and inputs may make public intervention *necessary*, "it need not be *sufficient* to improve allocative efficiency." The "sufficiency" of public investment to improve market allocation may be hindered by several factors. Among those noted by Dr. Krutilla are biases in the application of public investment criteria; constraints within the government on replacing traditional policies with alternatives more relevant to emerging problems; pricing policies for publicly provided goods which are inappropriate to the demand for the good or the state of the economy. These factors can lead to over-investment by the public sector to public investment in areas not relevant to current concerns or objectives and to private sector responses to these inappropriate government actions which will further distort the allocation of resources. Dr. Krutilla concludes that the effectiveness of public investment in counteracting market allocation failures requires a concerted effort to improve both decisionmaking procedures and the pricing policies underlying government investment activities.

I. RECENT HISTORICAL BACKGROUND

As Professor Baumol suggests in his incisive monograph Welfare Economics and the Theory of the State,[1] governments have operated in the economic sphere over a substantial historical period even in societies which are organized economically around the private ownership of property. This intervention in the economic sphere has taken many forms ranging from regulation, both macro and micro, to direct public investments. It is the latter type to which I wish to address my observations. I also wish to distinguish between governmental expenditures in the pursuit of various welfare program objectives, or considerations of equity, and expenditures which are of the nature of investments to improve allocative efficiency when confronted with market failure. In the latter case, to which I shall devote most of my attention, the decision criteria can be expected to differ from the former since the objectives differ.

The effort devoted to developing decision rules for efficient investment by the Government has a respectably long history. The initial

*Some portions of this paper are drawn from the article "Is Public Intervention in Water Resources Development Conducive to Economic Efficiency," published in the January 1966 issue of *Natural Resources Journal*. The author wishes to thank Dr. Albert Utton, Managing Editor of that journal for his permission to reproduce these portions.

[1] William J. Baumol, *Welfare Economics and the Theory of the State* (Cambridge: The Harvard University Press, 1952).

major contribution to this field, to my knowledge, was in the work of the personnel of water resource development agencies, the Bureau of the Budget and the former Bureau of Agricultural Economics during the period of the 1930's and 1940's. Among academics, notable contributions were made by Professor Ciriacy-Wantrup at the University of California, and by the agricultural economists, principally, at the land grant colleges. The framework of decision rules evolved to some extent pragmatically and existed in rather complete form for application under some circumstances in the so-called Green Book.[2] However, in the middle 1950's the field was discovered by other economists as well and the flow of literature was notably increased.

There were several developments which probably accounted for the increased interest in the field. First, the Treasury-Federal Reserve accord in 1951 liberated the latter from its obligation, assumed during World War II, to support the price of Government securities, thus returning to the Federal Reserve System its traditional function of seeing that interest rates performed more nearly their economic role in the economy. A second was the change in national administrations in 1953 which resulted in there being exhibited in Washington a marked preference for having more of the Nation's economic activities carried out in the private sector. Finally, there was the change in administrations in 1961 which ushered into Government, first in the Department of Defense and later throughout the Federal Establishment, an interest in performance budgeting and systems analysis that had its early origins in the work of the personnel at the RAND Corporation.

The change in administrations in 1953 with the consequent tendency to withdraw direct governmental action in the economic sphere induced examination of two sorts of economic questions more specifically than before. On the one hand, attention was given to reviewing or renewing the examination of the economic bases for governmental intervention in the economic sector. On the other hand, the interesting problem of suboptimization or "second besting" was addressed, with considerable emphasis on efficiency criteria in the public sector under conditions of budget or capital constraints.

Several related, albeit independent, efforts were undertaken during the mid-50's directed toward an attempt to develop the economic rationale for public intervention. One line of development at the most general level was the work of Samuelson,[3] which was an extension of some earlier work of Bowen,[4] and which was pursued further by Musgrave.[5] Another line had its origin in applications in the water resources development field, and particularly here were addressed the questions of investment criteria under budget constraints.[6]

Turning in particular to the justification of public intervention by reason of market failure, the argument may be summarized somewhat

---

[2] Subcommittee on Benefits and Costs, Report to the Federal Interagency River Basin Committee, *Proposed Practices for Economic Analysis of River Projects* (1950).

[3] Paul A. Samuelson, "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, Vol XXXVI, No. 3 (August 1954) pp. 387–89.

[4] Howard R. Bowen, *Toward Social Economy* (Rinehart, 1948). (New York : McGraw-Hill Book Co., 1959).

[5] Richard A Musgrave, *The Theory of Public Finance, A Study in Public Economy* (New York : McGraw-Hill Book Co., 1959).

[6] Otto Eckstein, *Water Resources Development: The Economics of Project Evaluation* (Cambridge: The Harvard University Press, 1958). Also Peter O. Steiner, "Choosing Among Alternative Public Investments in the Water Resources Field," *American Economic Review*, Vol. 49, No. 5 (December 1959) pp. 893–916.

as follows. In an economy characterized by competitive conditions, production and exchange at free market prices result in the highest valued production of goods and services (for any given distribution of income) of which the economy and its resources are capable. However, owing to some technical conditions incompatible with perfect competition in particular areas of the economy, free market results will fail accurately to reflect the social worth of inputs and outputs through the intermediary of market prices. Inefficiencies in the allocation of resources (or, equivalently, a relative reduction in the aggregate value of goods and services) result.

Where inputs to the production process are not readily divisible, the least cost scale of output may be large in relation to the market demand, and production may occur under conditions of falling average unit costs. Now, a necessary condition for efficiency is the pricing at marginal costs, but under the conditions mentioned above, marginal costs are below average costs and the efficient price and scale of production will not recover total costs. This poses a problem for efficient use of resources under purely private ownership and management without extra-market incentives.

On the consumption side, indivisibilities associated with outputs mean that the good or service in question cannot be discretely packaged and offered separately to each consumer subject to payment of the sales price.[7] Accordingly, without possibility of excluding any consumer for reason of nonpayment, the conditions essential for the formation of a market to service a demand are absent and we have what is referred to as a collective consumption (or public) good which cannot be provided without extramarket inducements.

Similarly, where the production processes of two or more fiscally and managerially independent producers are directly interdependent (that is, have a physical, rather than a market, intervening mechanism) so that the quality or quantity of one producer's output is affected by the production decisions of other fiscally independent producers,[8] the resulting benefits or costs escape incorporation into factor, or product, market prices. These phenomena, external to the decision calculus of producers, will thus appear as divergences between market values and real costs and benefits, and result in resource misallocations and the relative reduction in the potential value of the Nation's output. Accordingly, public intervention could be demonstrated to be a necessary condition for the attainment of efficiency goals.*

In this connection, however, perhaps not enough explicit attention was paid the fact that while public intervention was necessary, it need not be *sufficient* for improvement in allocative efficiency. For public action to be also a sufficient condition for improvement in efficiency, appropriate criteria must be developed and, assuming in the final analysis that there is a feasible way to do this, applied with sufficient fidelity to ensure that the efficiency objective of public intervention are reasonably approximated. This, of course, has been the ostensible purpose of public expenditure criteria whether referred to as benefit-cost,

[7] National defense or air pollution control are examples where if provided for any individual in the referrent group it is provided for all, and not subject to exclusion.

[8] For example, the adverse effect on a commercial farming enterprise afflicted with toxic flue gases emanating from a metal smelting operation, or perhaps the adverse effect on a commerical fishery from the agricultural application of persistent perticides.

*Further discussion of this issue is found in the papers in Part I of this volume.

cost effectiveness, or planning-programing-budgeting system analyses, particularly in the area of public investment programs.

II. The Problem of Bias in the Application of Investment Criteria

When a problem exists—in this case the problem of the allocative efficiency of the economy—which is generally recognized by the public, action is taken and the instrumentality for dealing with the problem is created. A new organization is formed, whether a new agency created or an existing agency assigned responsibility and provided resources to expand its scope of action; and the problem is attacked with considerable zeal and idealism. Generally, while there are possibly many false starts and errors made during an experimentation phase, the right combination is likely to be put together for effective operation to meet the objectives for which it was established.

If the objective is to provide continuing service for a situation which exists as a permanent feature of the scheme of things, one of the hazards of performance erosion is avoided. However, if the problem is one which occurs as a phase in the development of society and is dealt with effectively by the agency in question, after a time it becomes increasingly more difficult for the agency to justify its continued existence or scale of activity. At this point, or along the route, subtle changes may take place; namely, the organization begins to substitute for the general public objectives (which it has accomplished) more restricted agency objectives. In short, perhaps not even consciously, it may begin to substitute considerations of its own welfare for consideration of the public welfare. The means becomes confused with the end and the criteria for behavior become subverted by subjective considerations not wholly consistent with its original mission.

The attempts by agencies with functions of waning importance to preserve their status is but one of many similar institutional sources of bias in the application of investment criteria. Any agency, be its duties of waning importance or very significant, will be anxious to preserve or expand its area of responsibility and activity. If such agencies must perform in response to externally imposed standards, they begin to show pronounced biases in their application as judged by, say, *independent* professional opinion. For example, we might expect investment criteria to be applied with insufficient fidelity to warrant confidence that efficiency objectives are being achieved. This can be illustrated by the case of the interest rate in evaluating long-lived investments by the water resources agencies.

As a backdrop, it is perhaps well to consider the circumstances given rise to by the Treasury-Federal Reserve accord. Following World War II, the existence of inflationary pressure made it abundantly clear that the Federal Reserve System should be liberated from its obligation to support the price of government securities, which, in turn, resulted in a substantial rise in the interest rate over time. To the extent that these rates reflect the marginal return to capital in the sectors from which it is withdrawn in funding public investments, the rise in the rate of interest needs to be reflected fully in the benefit-cost calculations evaluating prospective investments. That is to say, if economic efficiency is to be achieved by governmental intervention, the marginal efficiency of investment must be equated between the private sector

surrendering funds and the public sector utilizing them. If this equation is not achieved, then while public action may remain a necessary condition for efficiency, it need not be sufficient. In that event, the justification for public intervention must be judged by the relative degree of inefficiency under private as compared with public development.

What is the quantitative significance of the preceding observations? Although interest rates rose considerably following the Treasury-Federal Reserve Accord, the rates used in the federal water resource project evaluations, for example, were very slow to respond. Initially, in 1952, the Bureau of the Budget requested of agencies that projects submitted for Congressional authorization employ a standard rate computed on the basis of the average rate of interest payable by the Treasury on outstanding long-term government bonds.[9] Budget Bureau Circular A–47, containing the standard, had a rather ambiguous existence, and not until 1962 and Senate Document 97 [10] was the standard made mandatory on all agencies participating in the water resources development field.

It should be observed that the interest rate referred to above was the coupon rate on bonds of long maturities rather than current yields. Accordingly, the rate continued to be determined in part by the level of rates prevailing before the Treasury-Federal Reserve accord, and thus reflected in part the persistent influence of the artificially depressed interest rate. Table I gives the trend in actual yields of long-term Government securities and the computed rates for project evaluation purposes based on the average of coupon rates.

TABLE I.—COMPARISON OF YIELDS ON GOVERNMENT LONG-TERM SECURITIES AND COMPUTED COUPON RATES

[In percent]

Year	Current yields	Computed rates	Year	Current yields	Computed rates
1952	2.68	2.5	1961	3.90	2.625
1953	2.93	2.5	1962	3.95	2.625
1954	2.70	2.5	1963	4	2.875
1955	2.94	2.5	1964	4.15	3
1956	3.08	2.5	1965	4.21	3.125
1957	3.47	2.5	1966	4.66	3.125
1958	3.43	2.5	1967	4.85	3.125
1959	4.08	2.5	1968	5.25	3.25
1960	4.02	2.5			

Such differences, while nominal in the early period, had become pronounced by 1968. In fact, projects with benefit-cost ratios of 1.3 to 1 and 1.4 to 1 when the nominal rate was used need not have actual benefits equal to their costs; that is, benefits would be overstated by 30 to 40 percent using the nominal rate for discounting.

[9] Hearings before the House Committee on the Interior and Insular Affairs: *Discussion of Budget Bureau Circular A–47 and the Related Power Partnership Principle,* 84th Cong., 1st 54 (1955).

[10] *Policies, Standards and Procedures in the Formulation, Evaluation and Review of Plans for Use and Development of Water and Related Land Resources,* Senate Document No. 97, 87th Cong., 2nd Sess. (1962). While the formula for computing interest or discount rates for application to water resources project evaluations appears in the Congressional promulgation of Senate Document No. 97, the staff work on the substance of the document was done by an inter-agency committee comprised of representatives from the water resources development agencies. To indicate the agency bias in the results, compare interest rate of 2.625 per cent generated under the S. D. 97 formula during 1961–62 with the recommendation for the same period of an interim rate (until a fully adequate investigation could be conducted) of between 4 and 5 per cent by a panel of independent professional economists in Maynard M. Hufschmidt, John V. Krutilla, and Julius Margolis, *Standards and Criteria for Formulating and Evaluating Federal Water Resources,* Report of Panel of Consultants to the Bureau of the Budget (Washington, D.C., June 30, 1961).

The above results relate only to the differences between the computed rate and the yield on long-term Government bonds. Students of the problem have uniformly recommended rates which exceed the yield on Government bonds, unless risk is accounted for explicitly in the calculation of benefits and costs.[11] Testimony recently given before the Subcommittee on Economy in Government of the Joint Economic Committee,[12] and the subcommittee's report,[13] establishes that the relevant concept for interest or discount rates in economic analysis of public investment decisions is the opportunity cost of capital, and that this may lie in the range between 7.5 to 12.5 percent under circumstances in which the risk premium is reflected in the discount rate.

As a partial indication of the inefficiencies resulting from inappropriate rates used in analysis of water resources investment, Fox and Herfindahl's work should be cited.[14] Applying interest rates alternatively of 4, 6, and 8 percent to the analysis of projects authorized by Congress in 1962 for construction by the Corps of Engineers, they found respectively that 9, 64, and 80 percent had benefit-cost ratios falling below unity. From these observations one can infer that artificially depressed rates have been used to "justify" projects which may be either inherently uneconomic or at least undertaken prematurely. And, of course, even with respect to projects which would show benefits in excess of costs at the higher interest rates, the use of depressed rates results in inefficiencies in the design of works; that is, excessive capital intensity and/or scale of facility.*

III. The Vulnerability of Venerability

This brings us to another consideration not related to public intervention per se, but rather to the efficiency with which the mechanism of public intervention operates. Typically, a governmental agency is established initially to carry out specific functions deemed desirable in the public interest. In the course of its activities the agency defines its role consistent with the functions it was established to discharge, develops the means or capabilities appropriate to these particular responsibilities, and in the process evolves a pattern of analysis and action unique to its role as it is perceived. The

[11] Otto Eckstein, *Water Resources Development: The Economics of Project Evaluation, op. cit.* Jack Hirshleifer, James C. DeHaven, and Jerome W. Milliman, *Water Supply, Economics, Technology and Policy* (Chicago: University of Chicago Press, 1960). John V. Krutilla and Otto Eckstein, *Multiple Purpose River Development: Studies in Applied Economic Analysis* (Baltimore: The Johns Hopkins Press, 1958). Roland N. McKean, *Efficiency in Government Through Systems Analysis* (New York: John Wiley & Sons, Inc., 1958). Robert H. Haveman, *Water Resources Investment and the Public Interest* (Nashville: Vanderbilt University Press, 1965). William J. Baumol, in his paper in this volume, presents a rationale for a 9-percent rate as a measure of the opportunity cost of public investment funds. See also William J. Baumol, "On the Social Rate of Discount," *American Economic Review,* Vol. LVIII, No. 4 (September 1968), where Professor Baumol presents a rationale for a 10 percent rate as a measure of the opportunity cost of public investment funds.

[12] *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis.* Hearings before the Subcommittee on Economy in Government of the Joint Economic Committee, U.S. Congress, 90th Cong., 2nd sess., July 30–31 and August 1, 1968 (Washington: GPO, 1968).

[13] *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis,* a Report of the Subcommittee on Economy in Government of the Joint Economic Committee, U.S. Congress (Washington: Government Printing Office, 1968).

[14] Irving K. Fox and Orris C. Herfindahl, "Attainment of Efficiency in Satisfying Demands for Water Resources," *American Economic Review,* vol. 54, No. 3 (May 1964), p. 198.

*Further discussion of this issue is found in the papers by Hirshleifer & Shapiro & Baumol in this volume.

more venerable the agency the more traditional the perception of its role and the means for performing its functions become, not only internally but also as a reflection of its public image. But with the passage of time, growth of the community, and changes in peripheral institutions with which it deals, the functions which the agency has been created to perform may recede in importance in the face of emerging new circumstances outside the competence of its original character. The legacy of traditional policy governing its behavior not only internal to the agency, but also imposed from without to a certain extent, restricts its capabilities to deal with the problems coming into dominance. The means of dealing with new situations is often restricted to the reservoir of expertise developed over the years in its traditional role. An example of this phenomena is discussed below.

In the Eastern United States, emergence of water quality management as a matter of dominant concern is increasingly recognized. Yet, despite the qualitatively different character of the problem, the first major planning effort undertaken in this context, the Potomac River plan, bears a remarkable resemblance to the conventional Corps of Engineers plans. In this case we discover the traditional emphasis on physical structures, storage reservoirs, but for purposes of augmenting low flows (of relatively rare occurrence but considerable severity) to dilute pollution concentrations. A very large part of the problem to be dealt with is concentrated in the Potomac estuary associated with the discharge from the Washington metropolitan area's sewage treatment plant. Assuring specified flows for an increasingly higher proportion of the time requires disproportional amounts of storage capacity which come at steeply rising costs because of conflicting uses for the technically superior reservoir sites. In this range of costs careful evaluation of the plan's latter increments is required by reference to alternative technical means or combinations of measures to achieve specified goals. Nevertheless, no detailed consideration was given alternatives such as the more uniform distribution of the treatment plant's effluent throughout the estuary to mitigate the concentration of biochemical oxygen demand, nor was consideration given to effluent oxygenation or other means of mechanical reaeration or chemical treatment some of which are relatively low in capital costs and suitable for dealing with occurrence of the rare event.[15]

We may speculate as to the reasons for this. Typically, challenges are met with the aid of expertise within the responsible agency, and it is not surprising that such expertise has been developed in response to meeting qualitatively different problems. Secondly, the Congress has not been very receptive to providing the Corps of Engineers with resources to mount a program of research and investigation relevant to the new range of problems which would support addition of staff with competence qualitatively consistent with the new circumstances. There is periodic static observed on Capitol Hill directed at suspected "duplication and waste" in the area of water resources research in the Federal Establishment. It does not require an overly perceptive agency

[15] As an alternative, see the approach in Robert K. Davis, *The Range of Choice in Water Management: A Study of Dissolved Oxygen in the Potomac Estuary* (Baltimore: The Johns Hopkins Press, 1968).

head to conclude that prospects remain slim for obtaining the support to staff expanded facilities which are deemed by the Congress to be esoteric to the agency's traditional functions. Finally, irrespective of the problem posed by the foregoing considerations, there is little disposition by any agency to consider seriously nonconventional alternatives falling outside the scope of its own capabilities when the institutional mechanism for introducing such measures is either ill-defined or nonexistent. There is a wholly understandable tendency for the responsible agency to feel constrained to rely only on the measures subject to its control. Given the permanence of reservoir structures, and agencies with a highly developed capability for their construction, the effect is to diminish seriously the prospects for introduction of known alternative technology which at certain times and under specified conditions will be superior to the measures relied on.

IV. The Specious Validity of the "Evaluation-Reimbursement" Dichotomy

A third area in which the sufficiency of public intervention for the improvement of economic efficiency may be questioned involves the practice of separating the problem of economic evaluation of a public investment opportunity from considerations of financial reimbursement.[16] The theoretical basis for this problem can be attributed to the French engineer Dupuit [17] in the early nineteenth century and resurrected and further developed by Harold Hotelling [18] in his classic work on marginal cost pricing. The burden of the marginal cost pricing thesis is that, within the context of the general economy, the efficiency of investment is independent of repayment for individual chunks of productive capacity. In a somewhat different form, but cut from the same cloth, is the Hicks-Kaldor thesis [19] of welfare economics to the effect that judgments regarding economic efficiency can be made independently of the income redistributive effects resulting from the implementation of the efficiency change. That is, if beneficiaries from a change in the state of affairs can compensate out of their gains those who are affected adversely, whether they do or not is not critical to the change being an improvement in productive efficiency. The income redistribution which attends the investment is asserted to be an ethical question without implications for economic efficiency.

It is necessary to distinguish more finely among three situations in order to assess the relevance of the above positions and their implications for economic efficiency. The first relates to the technical conditions of production, namely, economies of large scale. The second relates to the likely duration of the problem to which such technical conditions give rise. The third relates to the location in the economic process where the pricing policy for the output is to have its effect; that is, in the distribution of intermediate products used in further production or in the distribution of final consumption goods.

---

[16] This discussion abstracts from the "pure" public good, e.g. national defense, a good which if provided at all is available to all members of the referent group and not subject to exclusion for non-payment of a price.

[17] Reprinted with comments by Mario di Bernardi and Luigi Einaudi, "De l'Utilite et de sa Mesure," *La Riforma Sociale*, Turin, 1932.

[18] Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, Vol. 6, No. 3 (July 1938).

[19] John R. Hicks, "Foundations of Welfare Economics," *Economic Journal*, Vol. 49 (1939).

Now to recap briefly the theoretical basis for the marginal cost pricing thesis. Some costs, such as sunk capital, are fixed at the time the investment in plant and equipment is made and do not vary with the level of operation. Others vary as a function of the level of use. Insofar as the *operating phase* of any facility is concerned, so long as the increment to value per unit of output (price) exceeds the increment to cost (marginal cost), production should be expanded because more is added to the value of the national product than to costs occasioned by its production. Conversely, if the marginal cost exceeds the market clearing price, production should be curtailed; the resources thereby being released and redirected to uses where their opportunity returns are greater.*

Now, there are circumstances in which large-scale economies in production lead to a capacity which is large in relation to the market to be served and production takes place under falling average costs. This implies that marginal costs are below average costs and thus marginal cost pricing will not recover full costs (fixed as well as variable). It is under these conditions that marginal cost pricing creates a problem unless the efficient level of production is subsidized from public sources.

The point which needs to be made in this connection, however, is that the justification for marginal cost pricing under the circumstances related above, is grounded in static analysis and is relevant under static conditions—or what is practically more significant, that the condition of productive capacity which is large in relation to the market *is one of some considerable duration.* Of course, at the time of Hotelling's article there was excess capacity throughout the economy and the entire economy was afflicted with general stagnation. There was widespread justification for pricing policies which were indifferent to the recovery of full costs. Consequently, while full cost pricing under these conditions cannot be defended, the relevance of a pricing policy not recovering full costs can be questioned when the circumstances change materially. In a large, highly industrialized economy, experiencing vigorous economic expansion, it is not so clear that a policy of below average cost pricing is conducive to efficient resource use. For example, while there may be a temporary condition of excess capacity upon the construction of a facility that would justify, initially, a price below average cost, when use begins to approach capacity a rarely emphasized extension of the marginal cost pricing rule requires that prices be raised to ration the scarce capacity. A point will be reached ultimately (assuming continued expansion in demand) at which beneficiary willingness to pay for the service is sufficient to justify expansion of facilities.

Consider the case of the inland waterways or an express highway. Initially, with excess capacity a user fee or toll sufficient to recover only variable costs is required. However, as use of the facility expands to the point at which congestion begins to appear, each additional unit of use inflicts costs associated with congestion on all other users of the facility. In this sense, a social cost (spillover effect) occurs. This cost will rise progressively until the congestion costs approach the cost of enlarging the capacity. If the toll or user fees are raised correspond-

*Further discussion of this matter is found in the papers by Vickrey, Schultze, and Milliman in this volume.

ingly, the user fee set to equal the marginal social cost will, at some point, be just equal to the average cost of an increment of capacity.* In short, marginal cost pricing when costs are defined to include congestion spillovers will tend to recover costs of facilities of optimal scale and schedule of expansion.[20]

The third situation which needs to be discussed, and which is more generally related to the Hicks-Kaldor proposition than strictly to the marginal cost pricing issue, has to do with the stage in the economic process on which the pricing policy will have its effect. Under conditions in the United States—a large, highly industrialized economy characterized by growth in productive capacity and markets—the only case for divorcing reimbursement from economic evaluation which makes sense to me involves productive facilities which provide *consumer goods* under a set of special circumstances. Welfare propositions would suggest that consumer goods can be provided without regard to reimbursement provided that the production and distribution is regarded primarily as an income redistributive measure. While there may be some implications for allocative efficiency even here, efficiency in this case may be a subordinate consideration.[21]

Some would argue that the income redistributive goal should be advanced for public investments intended to increase the production of *intermediate goods and services.*[22] However, there seem to be enough problems with this case in practice to raise doubts regarding its efficacy as an income redistributive measure, as well as its consequences for efficient resource allocation.[23] Income redistribution can be looked upon technically as relaxing the budget constraint of specific disadvantaged individuals or groups. Or, in some case, as making provision allowing items, catering to what are referred to as "merit wants," to enter into the consumption patterns of individuals whose incomes are inadequate for this purpose. Subsidization of producer goods and services via reimbursement policies, on the other hand, has the effect of redistribution several stages removed from the point of intended impact with the consequent diffusion of redistributive effects among many individuals and groups not qualified on welfare or redistributive grounds. It resembles too much *a scattergun approach* in which only a small proportion of the shot-charge ultimately hits the intended target.**

---

[20] As a practical matter, of course, it is neither realistic nor desirable to have a continuously varying price of unpredictable magnitude (although peak load pricing of definite and known characteristics is quite another thing). The degree of uncertainty attending unstable prices for public works projects' output would affect adversely the investment environment for related industries. A practical alternative developed by Electricite de France is a stable price or rate equal to the average cost of the incremental capacity, incorporating where appropriate peak load pricing. See M. Boiteux, "Marginal Cost Pricing" in *Marginal Cost Pricing in Practice,* J. R. Nelson, ed. (Prentice-Hall, 1964). This point is made also by Herbert Mohring, "Urban Highway Investments," in *Measuring Benefits of Government Investments,* Robert Dorfman, ed. (Washington, D.C.: The Brookings Institution, 1965).

[21] Unless marginal cost pricing is practiced in the case of consumer goods, the marginal conditions necessary for efficiency in distribution (and consumption) are not met. However, if income redistribution is at issue, pricing *under the prevailing distribution of income* loses its normative significance.

[22] Arthur Maass, "Benefit-Cost Analysis: Its Relevance to Public Investment Decisions," *Quarterly Journal of Economics,* Vol. LXXX (May 1966).

[23] Robert H. Haveman, "Benefit-Cost Analysis: Its Relevance to Public Investment Decisions—Comment," *Quarterly Journal of Economics,* Vol. LXXXI (November 1967).

*Further discussion of this matter is found in the paper by Milliman in this volume.

**Further discussion of this issue is found in the papers by Weisbrod, Bönnen, Freeman, and Schmid in this volume.

Furthermore, on efficiency grounds, the practice of eliminating or reducing payout requirements in connection with facilities intended to provide intermediate goods or services has serious economic consequences. In the first instance, if producer goods or services from public investments are provided without costs, users of these factor services consider them as zero priced and combine their productive factors in proportions which reflect the "free" input. A greater use of the facility is made than real costs and potential returns would warrant. That is, not only those for whom the marginal value of the service is at least equal to an efficient user fee take advantage of the subsidized service, *but also other firms which would not find it profitable to do so if they were required to meet the marginal cost of such services.* Thus excess demand [24] will lead to premature congestion and pressure for expansion of facilities to alleviate the ostensible capacity shortage. When reimbursement considerations are separated from the investment decision, there thus develops a systematic bias in favor of facilities to be of greater capacity than economically warranted and/or to be redeveloped or expanded prematurely.

In addition to the inefficiency a systematic bias in favor of overexpansion introduces, there appear inefficiencies throughout the system in productive enterprises which utilize the underpriced publicly provided factor services.[25] That is, if materials and services in process are obtained by related production units at below their opportunity costs, there occur distortions in the marginal conditions of production by the relative overutilization of materials and services priced below cost and the relative over-allocation of all complementary factors devoted to the production of the end products in which the subsidized services are used. As an illustration of the inefficient use of complementary facilities consider the pricing of public power in the Pacific Northwest at rates below real incremental costs.[26] Low rates are potentially attractive to electro-process industries because of the importance of the power item in mill costs. Aluminum smelting is a case in point. The raw materials used in making aluminum are bauxite which is beneficiated to produce aluminum oxide on the Gulf Coast for imported bauxites and in Arkansas for domestic sources of the ore. The markets for the aluminum pigs and ingots are predominantly in the industrial heartland of the United States, the North Central States, and the East.

If the regional differentials in the rates for power are sufficient to compensate for increased material assembly costs, and distribution costs of mill end-products to the fabrication centers, the total of assembly, production, and distribution costs may dictate location of smelter capacity in the Pacific Northwest. But, if the favorable difference in total assembly, production, and distribution costs for the Northwest location is occasioned only by the subsidy element in the power rates, we see the wholly inefficient hauling of bulk commodities from mid-

[24] Excess demand here means the demand for the service which would exist beyond that which would be taken at an economically efficient price where the marginal value product of its use would fall in a range between the efficient price and zero.

[25] A similar argument has been elaborated for a somewhat different circumstance by Lionel McKenzie, "Ideal Output and the Interdependence of Firms," *The Economic Journal*, Vol. 41 (1951).

[26] The rates at which power is made available to industrial customers is below real costs because of the subsidies to capital invested in hydroelectric facilities and the manipulation of annual debt service charges by various techniques in the accounting procedures to reduce the annual requirements of revenue from sale of power.

America to the Pacific Northwest, and returned to North Central and Eastern America as finished items. Here the combination of transportation as a complementary service represents an excessive use of transportation as result of a non-efficient price of public power. When situations of this sort can occur, it is specious to maintain that pricing or repayment policy is irrelevant to efficiency considerations.*

There is an additional problem which arises as a consequence of the separation of project justification from reimbursement considerations. Typically, both theory and practice in project design abstracts from the reimbursement question and seeks to address only the question of whether or not the project is justified on the basis of efficiency criteria irrespective of pricing policy governing its output during the operating phase of the facility.[27] A project is to be included in the efficient program if its total benefits exceed its total cost. The project scale is determined by equating incremental benefits and costs, i.e., where the net benefit is maximized. However, if the pricing policy governing the output is not taken into consideration explicity, there is a dilemma which almost everyone involved in proposing investment criteria has failed to face up to directly.[28] *The problem arises in connection with determining the scale of a facility when it is anticipated that no user charge will be levied.* We have the dictum that the scale of a facility be fixed at the point at which incremental benefits and costs are equal. However, unless user charges are levied equivalent to cover the cost of providing the marginal units of output, there will develop excess demand represented by those users who would not find incentive to use the services of the facility if charges appropriate to the design criteria were imposed, but who will make use of the facilities because the service is provided at a lesser or zero price.

If such use by any beneficiary at capacity output adversely affects the utility received by any other, the design criteria result in a project which is inappropriately scaled relative to realizing the benefits estimated for purposes of its design. Rationing by means other than peak-load pricing will not distinguish between those for whom use of the facility has a value or benefit equal to or greater than the marginal social cost and those for whom the value is less. In short, there is no way in which to equate the marginal conditions for efficient use of the facility by rationing with any means other than price. At any rate, it is quite clear that the design criteria that have been generally proposed relate to an irrelevant case when reimbursement policies are at variance with design criteria. If the public agencies responsible for the service over-design the capacity of their facilities to avoid predictable congestion, it represent a pragmatic adjustment involving capacity provided at a cost not justified on economic criteria.[29] *Accordingly, in*

*Further discussion of this issue is found in the papers by Schultze in this volume, and Knetsch in volume 3 of this collection.

[27] An example of this is provided in Arthur Maass, Maynard Hufschmidt, Robert Dorfman, Harold Thomas, Stephen Marglin, and Gordon Fair, *Design of Water Resource Systems* (Cambridge: The Harvard University Press, 1962), pp. 38–40.

[28] The personnel of Electricité de France represent a notable exception in this regard. See Nelson, *Marginal Cost Pricing in Practice, op cit.*

[29] If the question is one of attempting to determine whether public intervention is more or less efficient than potential market failure in the provision of the goods and services in question, then the characterization of the problem as above is appropriate. If, on the other hand, we take as given public investment and operation of, say, the inland waterways, and concede that it is in the nature of things that no user charge is a practical alternative to present zero-pricing policies, we can acknowledge that expanding capacity (over-capacity) may relieve the congestion introduced by the excess demand, with attendant savings in social costs equal to or greater than the cost of capacity expansion. But, this is a second

*the absence of a pricing policy consistent with efficiency criteria, either an adulterated quality of service will be provided or a continuously existing over-capacity will prevail.** In the case of some types of facilities providing ultimate consumption goods, e.g., recreational, the providing of uneconomic capacity for the infra-marginal users may be dismissed perhaps with the "merit wants" rationale. This, however, would be obviously specious in connection with waterway users and other producers using intermediate goods provided without cost by the public sector.

Conclusion

These observations on the institutional machinery of public intervention at the present time merit serious consideration when the question is posed with respect to the sufficiency of public intervention for the improvement of allocational efficiency. Perhaps the time has arrived when the greatest dividend to study and experimentation would come in response to a concerted effort to review the character and capabilities of existing agencies, to identify the barriers to improved performance, and to address the problem of organizing a capability consonant with the requirements of the present and future. In such an evaluation and assessment of alternative institutional arrangements, the centrality of reimbursement policy for improvement in efficiency would need to be faced squarely with a recognition of the powerful influence which cost bearing by beneficiaries would have on the discriminating use of the entire range of cost reducing technological alternatives. Not a little could be learned from the pioneering work in water quality management by the cooperative water quality management associations in the Ruhr, or the integration of investment planning, design criteria, and pricing policy of Electricité de France. Doubtless a similar spirit of innovation in the public works sector in the United States would add a considerable measure of assurance that public intervention would be sufficient as well as necessary to improvement of efficiency in the allocation of economic resources.

---

best solution and is not the same as claiming greater efficiency for public intervention than muddling through with a situation characterized by market failure. Second-besting is muddling through as well.

*Further discussion of this issue is found in the paper by Milliman in this volume.

BENEFICIARY CHARGES AND EFFICIENT PUBLIC EXPENDITURE DECISIONS

BY JEROME W. MILLIMAN*

Jerome W. Milliman is Professor of Business Administration and Director of the Institute for Applied Urban Economics at the School of Business of Indiana University.

In the judgment of most observers of public expenditure policy, the failure of governments to impose charges for the benefits of publicly produced outputs is a major impediment in developing efficient government expenditures. Elsewhere in this collection of papers, for example, both Charles Schultze and John Krutilla argue that failure to use cost-sharing and user charge arrangements when they are feasible has led to inefficiency in public expenditure decisions. In this paper, Professor Milliman presents an analysis of the economics of financing publicly produced outputs.

In reviewing the literature on beneficiary charges, Professor Milliman finds three rationale for employing such charges—equity, revenue production, and economic efficiency. After analyzing each of them, he concludes that the "arguments advanced for user charges are quite varied, fragmented, and sometimes inconsistent." In appraising the argument that user charges are necessary to encourage short-run efficiency (to ration the output produced by a government project), Professor Milliman sees the need for further work on the theory and practice of such charges before definitive recommendations can be offered. However, with respect to the role of beneficiary charges in promoting the proper scale and level of public investment, he states that "Unless substantial payment for benefits is required from beneficiaries, or from the jurisdictions in which they reside, the forces to 'discipline' public investment decisions will be very weak. Clearly, when the discipline of the market is absent, there are serious problems of how to obtain responsible public investment decisions. We lack measures to reward good decisions and to penalize poor ones."

I. Introduction

The economic analysis of public expenditure decisions has advanced a great deal in the last two decades. Public finance, which was largely concerned with problems of economic stabilization and with the "finance" problems of how to levy the taxes and sell the bonds, has turned to the long-neglected topics of what public goods and services should be produced, how much should be invested, how should public goods and services be produced, and how should they be distributed. A large body of literature now exists on the evaluation and "systems" approaches to public investment decisions. The techniques of benefit-cost analysis, program budgeting, systems analysis, cost-effectiveness analysis, and operations research are becoming accepted as standard tools for aiding public expenditure decisions.

* I am grateful for comments on an earlier draft received from James Buchanan, William Baumol, David Davies, Robert Haveman, Orris Herfindahl, John Krutilla, David Martin, David Maxwell, Edgar Olsen, Richard Pfister and Louis Shere. At several places I have failed to heed their advice. At other points, my critics were far from unanimous.

With these developments, it would seem that analysis of production and distribution of public goods and services would have come of age. Yet, as one studies the theory and practice of the production and distribution of public goods and services large gaps appear in our knowledge of demand and supply relationships. On the supply side, it is important to note that we have trouble measuring outputs and therefore have very few empirical production functions expressing the technical relationships between factor inputs and product outputs. We have even fewer cost functions with which to determine the optimal combination of factors to produce given outputs. As a result, we have often relied on expenditure functions per capita to describe relations between costs and outputs. Changes in expenditures are clearly ambiguous numbers in the absence of independent measures of outputs. Moreover, we often tend to count expenditures on inputs as measures of output.[1]

It is also evident that the demand for public services is often an ambiguous concept and that very scanty information exists on how well political processes reflect preferences of consumers and voters. As Margolis suggests, it is difficult to estimate demand functions for public goods because:

> The consumers of the goods are not the purchasers; the purchasers are a mix of elected and appointed officials who pay with tax revenues; the taxpayers may not be users of the services and the decision-makers may be neither taxpayers nor users. Observations on prices or quantity are rare; costly surveys are often necessary to tell us who uses the services; and the handful of studies on who pays for the services are highly oversimplified. Not only are there several steps between the consumer and the payer, but often the consumer may not be part of the political constituency which is doing the paying.[2]

These knowledge gaps on both the demand and supply aspects of the production and distribution of public goods and services are clearly related to the question of beneficiary charges. On the one hand, if we do not have very solid knowledge of outputs, costs and demands, the problem of how to determine optimal prices, charges and taxes for these goods and services is clearly in a precarious state. On the other hand, such prices and charges might be helpful as a form of voting by consumers which may supply (in many cases) valuable information about consumer demands. For example, if a public good is supplied at a zero price, we can expect that excess demand will develop so that "cries of alarm" may suggest that more production is needed. Yet, we may have little information as to the value of this service and its claim on resources versus alternative uses of the same resources.

One upshot of these gaps in our knowledge is that the theory and practice of beneficiary charges has received very little attention. Despite the apparent wealth of literature and the existence of case studies in the general field of public investment evaluation, there often appears to be a dichotomy between the assumptions used in the evaluation

---

[1] Some examples of this tendency are cited by Hirsch and Musgrave. See Werner Z. Hirsch, "The Supply of Urban Public Services" and Richard A. Musgrave, "Discussion of Part III" in Perloff and Wingo (eds.), *Issues in Urban Economics*, Johns Hopkins Press, Baltimore, Md., 1968. See also the papers in Volume III of this collection.

[2] Julius Margolis, "The Demand for Urban Public Services," in Perloff and Wingo, *op. cit.*, p. 536. See also the paper by Margolis in this volume.

of the initial investment and the decisions about the charges for use of the public service by consumers. The decision whether or not to employ beneficiary charges is clearly one that will affect the use of the facility and hence the amount of investment needed. More importantly, beneficiary charges (positive or zero) should be an important element in the design and investment evaluation process itself. Why do we worry a great deal about the choice of a discount rate in public investment evaluation, yet pay very little attention to assumptions about the future beneficiary charges? We give little explicit theoretical consideration to their influence on consumer behavior.

Krutilla is one of the few writers who has worried about the difficulties of determining the proper scale of facilities when no charge is to be levied on the beneficiaries:

> We have on the one hand dicta that the scale of a facility should be extended to the point at which incremental benefit equals incremental cost. However, unless user charges are levied to cover the cost of providing the marginal unit of output, there will develop excess demand represented by those users who would not find incentive to use the services of the facility if charges appropriate to the design criteria were imposed, but who will make use of the facility at zero price. If such use by any beneficiary at capacity output adversely affects the utility of any other, the design criteria result in a project inappropriately sized relative to realizing the benefits estimated for purposes of design. At any rate, *it is clear that design criteria as presently advanced relate to the correct design for an irrelevant situation where reimbursement policies are at variance with design criteria.*[3]

The purpose of this paper is to analyze in critical fashion the view that failure to employ beneficiary charges for the financing and rationing of public goods and services has led or can lead to inefficiency in public expenditure decisions and to wastes in production and consumption of these goods and services. The paper sets forth the analytical framework for the employment of beneficiary charges. Three types of rationale have been proposed. Possible inconsistencies between these three purposes are discussed. Conditions under which beneficiary charges are desirable and feasible are outlined. Criteria for charges and the various bases for determination of charges are described.

In general, it is found that literature on beneficiary charges is surprisingly sparse and fragmented. Some of the literature tends to confuse the needs for public revenues and financing with the need to promote efficient resource allocation. There is apparently no comprehensive or definitive statement of the relationship between beneficiary charges for public goods and services and efficient resource allocation. Moreover, beneficiary charges (positive or zero) clearly are related to the question of income transfers. In order to make rational decisions about the extent and direction of income transfers related to public production and consumption one should have a clear notion of the relationship of beneficiary charges to recovery of costs and the welfare gains and losses associated with the distributional effects upon consumers and taxpayers.

---

[3] John V. Krutilla, "Is Public Intervention in Water Resources Development Conducive to Efficiency?" *Natural Resources Journal*, Vol. 6 (January 1966), p. 72. Italics added here. See also the paper by Krutilla in this volume.

My attempts to find important case studies and empirical measures of the possible inefficiency generated by the failure to employ beneficiary charges have been largely unsuccessful. By the same token, empirical studies are generally lacking on the distributional transfers involved when beneficiary charges are not employed. It appears that the important relations between beneficiary charges and efficient resource allocation and desirable distributions of income are so little understood (and appreciated) that few people have attempted to produce evidence in this matter.*

II. Asserted Need for Beneficiary Charges

Recently a growing number of writers have advocated greater employment of beneficiary charges. A few examples of this emerging line of thinking show that questions of efficiency, of equity, and of financing are intermingled. I shall point out below that some of these objectives may be inconsistent with each other.

Fox and Herfindahl in reference to efficiency of federal investment in water resources state:

> The desirability of two modifications is almost self-evident, and immediate attention is merited. One is to place greater reliance upon charges and prices in the allocation process. It is extremely doubtful that the income redistributive consequences of existing subsidy provisions achieve any clear social objective. If the direct beneficiaries were required to pay for the services they receive, political support for projects would more accurately reflect their social value. *Probably no other single measure would contribute more to the attainment of efficiency in satisfying demands for water services and in decisions such as location that are presently distorted by subsidized prices.* Such a change may be difficult to bring about for flood control, navigation and irrigation because existing subsidy policies have been so firmly established, but particular attention should be directed to water quality which promises to demand such large investments in the future. As has been so well demonstrated in the Ruhr, penalty prices adjusted to the amount and kind of effluent can be used as an effective device for stimulating economically efficient behavior where pollution is a serious problem.[4]

A longer statement by Professor Fox advocating increased use of pricing and cost-sharing policies for water resources is presented below (in condensed form). The Fox arguments on water resource expenditures would seem to have some relevance to many other kinds of public production. Note that Fox intermingles arguments for efficiency in making new investments in capacity with arguments for making efficient use of existing facilities already constructed. As I will suggest below, beneficiary charges for the former may not be consistent with beneficiary charges for the latter argument even though both arguments are concerned with efficiency in public expenditure decisions:

---

[4] Irving K. Fox and Orris C. Herfindahl, "Attainment of Efficiency in Satisfying Demands for Water Resources," *American Economic Review* (May 1964), vol. LIV, No. 3, pp. 205–206. Italics added here.

*Further discussion of this issue is found in the papers by Freeman and Bonnen in this volume.

There is mounting evidence that present pricing and cost-sharing policies suffer from major limitations which merit explicit examination.

First, as previously noted, existing policies deprive the decision maker of valuable information for decision-making. This is done in two ways. Since we do not price water services to reflect value, planners have little data to go on in estimating whether a given level of service is worth the costs. Also, water services are used in accord with the price tags they bear. If water is priced at zero it is used as though it were worthless. This lack of a price signal on many water services results in misuse of water and misallocation of resources.

Second, existing cost-sharing policies for water services tend to cause those affected not to consider alternative ways of investing available funds which might achieve an equal or better result.

Third, the system of costsharing, which is often justified on the basis of being a way of redistributing income from the more prosperous to those in need, or from the wealthier parts of the country to the less wealthy, serves this purpose inefficiently.

Fourth, and an especially significant point, is that the cost-sharing system gravely handicaps the effort to have those affected by the decisions properly represented in the decisionmaking process.

Fifth, existing financial policies discourage experimentation with new policies and institutions. A state, or regional organization cannot compete with the cost-sharing policies of Federal agencies because only through Federal agencies is there access to the favorable terms of support offered by the Federal Government.[5]

A new plea for user charges in transportation was recently made by President Johnson in his 1970 budget proposals:

> The President again recommended new or increased taxes on truckers, waterway users and airway users, including airline passengers. If enacted by Congress, the taxes would raise an estimated $402 million in fiscal 1970, which begins July 1. But similar taxes have been proposed before and have failed to get anywhere in Congress.[6]

The increased application of beneficiary charges to public production by State and local governments has received endorsement by a number of writers in recent years. Stockfisch has shown how city governments can raise substantial revenues by imposing fees and service charges on many functions they perform in addition to covering costs of enterprise services, by applying "rational" pricing principles. He estimated that the city of Los Angeles, by a failure to charge appropriate fees and prices in 1957, forfeited revenue sources that could have taken care of 43 percent of its actual expenditures.[7] Note that Stockfisch is as much concerned with the question of how to raise revenues as with the question of efficiency in the production of public goods and services.

---

[5] Irving K. Fox, *Essays on United States Water Resources Policy* (Mimeo) February 1968. University of Wisconsin, Madison, Wisconsin.

[6] *Wall Street Journal*, January 16, 1969.

[7] J. A. Stockfisch. "Fees and Service Charges as a Source of City Revenues: A Case Study of Los Angeles," *National Tax Journal* (June 1960) Vol. XIII, No. 2.

Davies has estimated that between 1946 and 1963 local governmental user charge revenues increased approximately 588 percent, second only to sales taxes as the most rapidly growing source of urban public revenues. Davies concludes, "On the basis of its history during the most recent two decades, prospects for market pricing in the urban public sector appear to be quite bright." [8]

Vickrey has appealed for greater use of specific taxes, fees and prices of urban services to increase the efficiency with which the services are used and also to improve the spatial patterns along which metropolitan areas grow. Vickrey proposes extending the concept of marginal cost pricing as far as possible in the realm of municipal services.* However, Vickrey also sees the need for more revenues for urban governments and argues that the adverse effects on resource allocation of general taxes may be considerable so that:

> . . . if any specific charges are to be made, they should in nearly all cases be *designed in part to contribute to the public treasury over and above the amount that would flow in on the basis of charges strictly reflecting marginal costs.*[9]

Note that Vickrey has suggested that marginal cost pricing be used to promote efficiency, but at the same time he urges that charges be set *higher* than marginal costs to bring in general revenues to the public treasury. Clearly, these two purposes or rationales for beneficiary charges may be inconsistent.

Netzer estimates that user-charge types of revenues finance about one-sixth of public services in metropolitan areas. If State and Federal highway-user taxes are included the proportion increases to one-fourth. Netzer believes that increased reliance on user charges can overcome part of the problem of political fragmentation in metropolitan areas because payment would be based upon use and not the domicile of the taxpayer. The result might be that there would be an increased ability of some public enterprises in metropolitan areas to increase output and to upgrade service, but Netzer notes that there has been little empirical study of the effects of user charge financing on the output of services.[10]

In addition to suggesting the need for greater and more sophisticated employment of user charges by urban governments, Netzer makes an incisive observation about the traditional rationale advanced for the application of user charges:

> One of the most serious obstacles to effective utilization of pricing devices in financing public services can be traced to the traditional justification for user charges (and analogous taxes) on the basis of benefits received. The benefit principle, however, is concerned with equity, not allocation: How can we spread the costs of public goods among individuals? But the principle is inappropriate for allocation branch decisions on the financing

---

[8] David Davies, "Financing Urban Functions and Services," *Law and Contemporary Problems*, Winter, 1965, p. 160.

[9] William W. Vickrey, "General and Specific Financing of Urban Services" in Schaller (ed.), *Public Expenditure Decisions in the Urban Community*, Johns Hopkins Press, Baltimore, Md., 1963. Italics added here.

[10] Dick Netzer, "Federal, State and Local Finance in a Metropolitan Context," in Perloff and Wingo (eds.), *Issues in Urban Economics*, Johns Hopkins Press, Baltimore, Md. (1968), p. 454.

*Further discussion of this issue is found in the paper by Vicky in this volume.

and provision of services with a substantial private character. Efficiency in allocation requires that prices (or other types of charges) and the level of services provided be determined on the basis of the marginal *costs* of the services. *More often than not, there is little correspondence between benefit-determined charges and cost-determined charges.*" [11]

In spite of these statements lending support to the view that increased use of beneficiary charges is desirable, a strong flavor of skepticism can be found in the literature. Brownlee, in a review of the role of user charges versus the use of general revenue sources, states:

> Although I favor using price as a rationing device whenever a reasonable opportunity exists, I believe that the appropriate area for the application of market pricing to the determination of how much various goods and services government should produce is a relatively small one. Opportunities undoubtedly exist for financing such items as fire and police protection partially on the basis of service rendered. Nevertheless, such cases would be relatively insignificant in terms of the overall patterns of public expenditure. I see few *major* services that ought to be financed exclusively from sales revenues that are not already being financed in this way. However, the criteria currently used for establishing prices and for determining how much to produce are not necessarily the best ones. Furthermore, charges at less than cost might well be established for some services that now are provided free.[12]

We will return to some of these views on the desirability of beneficiary charges later on. It is clear now that we need to sort out the various questions relating to the rationale for beneficiary charges and the criteria for their determination before we can evaluate some of the claims and issues. However, we can note that beneficiary charges cannot be "all things to all men" in the sense that efficiency considerations for long run investment decisions may not be the same as efficiency in the short run, treated as best use of existing facilities. And, in turn, questions of financing and revenue production may require a different approach to beneficiary charges than efficiency considerations. We will have to decide what it is that we want beneficiary charges to accomplish before we can go very far in dealing with possible obstacles to inefficiency in public expenditure decisions.

III. What Are Beneficiary Charges?

At first blush it might appear that a beneficiary charge is a rather straightforward term applied to a charge upon the recipient of a publicly provided good or service. And in many ways this is a good place to start. For a fairly restrictive definition we can note that the Bureau of the Census defines fees, service or user charges as "amounts received from the public for the performance of specific services benefiting the person charged and from sales of commodities." [13] The Census Bureau

---

[11] Netzer, *op. cit.*, p. 457. Italics added here. Netzer's position that the benefit principle is concerned with equity, not allocation, would not be acceptable to Wicksell and the Italian writers in public finance. Netzer's position is also in conflict with Musgrave's position below. This observation was suggested to me by David Davies.

[12] O. H. Brownlee, "User Prices vs. Taxes," in *Public Finances, Needs, Sources and Utilization*, National Bureau of Economic Research, 1961, p. 421.

[13] U.S. Bureau of the Census, *Compendium of City Government Finance in 1963*, U.S. Government Printing Office, Washington, D.C., 1964, p. 96.

views such charges as "voluntary" as opposed to taxes which are "compulsory." It also points out that the service benefits the person charged rather than the community at large.

By contrast, we can note that the whole theory of benefit taxation is based upon the notion that the taxpayers are beneficiaries and it is proper to require them to pay. As Musgrave has stressed, the benefit approach to taxation is not appropriate to problems of stabilization and distribution but it is quite relevant to problems of the allocation branch of government.[14] In theory, even taxes based upon the ability-to-pay principle, but levied upon the beneficiary group itself, might be called beneficiary group itself, might be called beneficiary charges in the broadest sense. However, taxation of one group to provide for benefits to another group would not be ordinarily classified as a beneficiary charge in the usual sense. Yet, we might even stretch this point even a bit farther. For example, assume that one group is taxed to provide benefits to another group (e.g. free milk). If the taxpayers are presumed to receive some collective benefit from this income transfer or if their utility functions are judged to include the welfare needs of the poor, then we might term the tax a beneficiary charge, at least to some extent.

The purpose of this discussion is not to create a tent under which we can include all service charges and tax payments as beneficiary charges. Instead, we want to stress that there is no hard and fast definition and that the term "beneficiary charge" can mean many things to many people. As a result, a beneficiary charge can refer to a whole host of charges ranging from the sale of postal services, rent charged for public housing units, inspection fees, sale of public utility services, local property taxes on properties served by sewers, property taxes to finance given school districts, charges levied upon irrigation districts in Bureau of Reclamation projects, gasoline taxes upon motorists, and perhaps even to Federal income taxes levied to finance such collective goods as national defense.

To be sure, a property tax levied upon a district to finance a service to that district may provide only a very rough correspondence between the beneficiaries of the service and the charge (not all property owners in a school district have children). Gasoline taxes place the costs of highways upon drivers as opposed to non-drivers. However, gasoline taxes in one region may be used to pay for highways in another region. In such cases, one might fairly say that the tax is no longer a beneficiary charge. Musgrave would tend to restrict the term benefit taxes to cases in which the benefits and costs are seen by individuals. Correspondence of benefits and costs across groups (group equity) or across regions (regional equity) would be less acceptable to him.[15] By contrast, Buchanan is willing to classify gasoline and motor vehicle taxes as user prices when the funds are earmarked for highway construction even though some vehicles produce more ton-miles than others and even though different roads involve different amounts of cost and different amounts of use.[16]

[14] Richard A. Musgrave, *The Theory of Public Finance*, McGraw-Hill Book Co., New York, 1960, Chapter 4. James Buchanan has criticized me for accepting Musgrave's separation of the budget into three parts which he considers methodologically unacceptable. I think a good case can be made for the view that distributional and efficiency considerations can not be separated.

[15] Musgrave, *op. cit.*, p. 177.

[16] James M. Buchanan, *The Public Finances*, Richard D. Irwin, Homewood, Ill., 1965, pp. 557–558.

IV. The Rationale for Beneficiary Charges

Three major strands of literature in economics deal with the question of beneficiary charges: The public finance literature, the public utility literature, and the welfare economics literature on marginal-cost pricing. To my knowledge, no one has pulled these three strands together to present a unified theory of beneficiary charges. As we have seen, the arguments advanced for such charges are quite varied, fragmented, and sometimes inconsistent.

A. The Public Finance Literature

Although there are a few exceptions, the public finance literature on beneficiary charges deals primarily with the question of the benefit approach to taxation. The question of how to levy fees or user charges is largely ignored. Some of the best known public finance texts do not even treat the subject. In general, the stress is on the equity and the efficiency gains of having taxpayers pay for benefits received. However, these equity and efficiency gains are still considered within the context of how to finance expenditures and raise revenues. Governments must raise revenues to finance expenditures and benefit taxes are a "fair" way to do it. Problems of optimal resource allocation and efficiency are introduced only indirectly and are usually more concerned with the longer run questions of investment in public facilities and not the efficient use of facilities.

The equity aspects of benefit taxes are viewed generally as a matter of simple justice that users of a public service should pay for its costs when the benefits do not spill over on to other people. For example, the benefits of irrigation water service provided by the Bureau of Reclamation accrue to the irrigators and it would be "unfair" to require the public at large to pay for this service. By contrast, the provision of free hospital service to the poor might be viewed as a case where the indirect benefits to the community at large justify subsidizing the service. The usual textbook discussions do not elaborate on *which community should subsidize the service*—the Federal or National Government, the State or the local community.[17] The answer would probably be that it should be the community which receives the indirect benefits.

The case for subsidy on the basis of indirect benefits received probably is best applied to localities as opposed to the Nation. When benefits are provided to one group or region and the costs are borne by other areas or groups (who do not receive direct or indirect benefits) income transfers are involved and such redistribution must be justified on grounds *other than* the benefit principle of taxation. Yet, it is very seldom argued in theory that local publics, for example, ought to share more fully in Federal programs (such as flood control or urban renewal) because most of the indirect, as well as direct, benefits reside locally. There seems to be the feeling that benefits which are "widespread" are automatically national in character.[18] We shall make the same point below in reference to collective consumption goods when the "public" character of such goods is often attributed to the Nation

[17] An exception is James Buchanan. See *The Demand and Supply of Public Goods*, Rand McNally, Chicago, 1968.

[18] As Edgar Olsen has sugegsted, the benefits of flood control and urban renewal are viewed as widespread because many communities have projects. The relevant question from the viewpoint of the beneficiary charges is the extent of benefits from any one project.

as a whole when, in fact, collective consumption may be highly restricted to a given locality or user class, e.g., how widespread is the collective consumption of the services of a lighthouse or those of a local police force?

In the traditional public finance literature, the efficiency aspects of benefit taxes (charges) are largely limited to the longer run questions of the "proper" investment in capacity or the optimal scale of service.* Thus questions of how best to ration service from existing capacity are seldom discussed. The investment in public facilities is considered justified if consumer preferences in the case of direct tolls, or taxpayers preferences in the case of benefit taxes, return funds to pay for the "cost" of the facility. The literature on cost-recovery and cost-reimbursement for public investment takes its rationale from these questions of equity and "long-run" efficiency discussed here. If the cost is not to be recovered from the beneficiaries either an error has been made in the original investment or if the beneficiaries are not required to pay there would be a transfer of income from other taxpayers to the beneficiaries of the service. If "profits" are earned, then the investment should be expanded.

Moreover, so the argument goes, if we do not require cost-recovery from the beneficiaries we will not *really* know whether the benefits from the service exceed the costs. This argument is often made both in an *ex ante* sense and in an *ex post* sense. The requirement of cost-recovery from the beneficiaries as a precondition makes "sure" that the beneficiaries will carefully consider the worth of the benefits in relation to the costs during the plan formulation period. *Ex post* calculations can point out past errors and serve as a basis for future improvements in decisions for new investments. All of this seems reasonable except that we shall point out that emphasis on cost-recovery may be inconsistent with the best shortrun use of existing facilities. In addition it often places excessive emphasis on the recovery of historical or "sunk" costs, which may not be efficient for new investment decisions.

The economic costs relevant at any one time are the opportunity costs of resources and the alternatives sacrificed now and in the future rather than in the past. Although cost-recovery from beneficiaries has many obvious virtues, recovery of historical or sunk costs will be inefficient because historical costs are not likely to coincide with opportunity costs. We are always faced with the possibility of change in the future and the possibility of not making correct projections when constructing long-lived facilities. To insist on cost-recovery of historical costs in the light of greatly changed conditions may clearly result in charging too little or to much. Moreover, cost-recovery tends to foster the idea that once facilities are "paid for" they should be free. More on this later.

B. THE PUBLIC UTILITY LITERATURE

The second major strand of literature dealing with beneficiary charges is found in the writings on public utility economics. The principles of public utility rate regulation are an uneasy blend of legal principles and economic principles tied together with a framework

*Further discussion of this issue is found in the papers by Schultze in this volume, and Knetsch and Nelson in vol. 3 of this collection.

of cost-accounting. The legal principles tend to be concerned with the financial requirements of the utility and the notions of equity and fairness to stockholders and consumers. In general, economic theory has had to adjust to legal and financial constraints.

This is not the occasion to launch into a full-scale examination of the principles of public utility rate regulation (in general these principles have been applied to publicly owned utilities as well as to private ones). However, it will be helpful to sketch out the basic features because any study of the possible inefficiency generated from the failure to employ beneficiary charges must also come to grips with the possible inefficiencies from the use of improperly applied or incorrect user charges. I suspect that a thorough study of the matter would show that major efficiency gains could be achieved by a drastic overhaul of the system of rate regulation of public utilities (Federal, State, and local).[19]

In large part, the level of public utility rates is determined by the following relationship:[20] $R=E+(V-D)r$

where R is the total revenue required to cover total costs; E is the full operation and maintenance expense; V represents the "fair value" when the facilities were new;[21] D is the depreciation allowed in the value of the facilities; and r is the "fair rate of return" to be allowed on the current "fair value" (V–D).

This method of rate determination has the following schematic properties:

Total revenue (TR) = Total Costs (TC)

Average revenue (AR) = Average Costs (AC)

In short, public utility rates are based in large part upon two principles which are questionable from the standpoint of efficiency in the use of resources: (1) rates tend to be based upon recovery of historical or original costs and (2) rates tend to be determined by the average cost of service as opposed to the marginal cost of supply. This sort of price policy probably represents a nice balance between such multiple objectives as equity in the distribution of wealth between consumers and stockholders, fairness among classes of service, and the provision of the financial needs of the utility companies. Yet, it is questionable that it deals adequately either with efficiency in the use of existing services or with the development of optimal criteria for new investment. On the one hand, it is quite likely that the scope of public utility operation is expanded and applied to an increasing range of public or quasi-public services. On the other hand, the opportunities to change or overhaul present practices and accepted principles of regulation are not bright because of the long weight of court and legal precedents supporting the present structure and because marginal-cost pricing, itself, has some practical and theoretical difficulties.

---

[19] Similar positions are taken by William J. Baumol, *et al.*, "The Role of Cost in the Minimum Pricing of Railroad Services," *Journal of Business*, vol. 35 (October 1962), pp. 357–366; Jack Hirshleifer, *et al., Water Supply: Economics, Technology and Policy*, University of Chicago Press, Chicago, Ill., 1960, ch. V; and William S. Vickrey, "Some Implications of Marginal Cost Pricing for Public Utilities," *American Economic Review*, vol. 45 (May 1955), 605–620. See also the paper by Vickery in this volume.

[20] This formulation is used by Eli W. Clemens, *Economics and Public Utilities*, Appleton-Century Crafts, Inc., New York, N.Y., 1950, p. 127.

[21] "Fair value" can mean values other than original cost. A better alternative from an economic standpoint would be replacement cost. In actual practice James C. Bonbright says, "* * * *all* commissions administer, with more or less efficiency or laxity, versions of, or minor deviations from, an actual cost or sunk cost standard of reasonable utility rates." See: *Principles of Public Utility Rates*, Columbia University Press, New York, 1961, p. 283.

C. THE MARGINAL COST PRICING LITERATURE

The third major body of economic literature bearing upon the theory and use of beneficiary charges comes from the writings of "welfare" economists who have developed the theory of marginal-cost pricing. Although some of the principles of marginal-cost pricing are now being absorbed or adopted in writings in public finance and in public utility economics, as yet a full-scale synethesis has not been achieved. For example, a public utility textbook will generally have a chapter on marginal-cost pricing but then very little is done to relate the implications of marginal-cost pricing to the theory and practice of rate regulation. Neither public finance nor the public utility literature has succeeded in rationalizing the needs for reimbursement of financial costs with the efficiency rationale of marginal-cost pricing which may generate "surpluses" or deficits. The theory of marginal-cost pricing stresses that investment and operating decisions on social investments should be made *independently* of reimbursement policies for individual lumps or units of productive capacity.

It seems clear that marginal cost pricing is still a very controversial topic on which there is not a complete concensus among economists. At one extreme are those who have accepted it as a major pillar of public policy;[22] at the other extreme are those economists who have rejected it as invalid or wholly impracticable.[23] In between these polar points of view are many economists who have advocated changes or modifications in the principle itself or qualified the claims that its strict application will produce an optimal allocation of all resources.[24]

In its most simple form the cardinal rule of marginal-cost pricing is that the demand price should be made equal to marginal cost, with marginal cost defined as the incremental costs of production (more technically as the derivative of the total cost function with respect to output). Since resources are drawn away from alternative uses, marginal costs should reflect accurately the social opportunities foregone. The equality of price and marginal cost insures that consumers equate marginal benefits from this use of resources with the real alternatives foregone elsewhere. In a world of pure competition the market mechanism would operate to insure this equality.

For some types of public goods, production is characterized by long-run decreasing costs or increasing returns to scale. Marginal cost will lie below average cost in increasing return cases.[25] If demand price is made equal to average cost (full-cost recovery) the price will exceed marginal cost and it can be demonstrated that this result is inefficient, because the value consumers place upon extra output exceeds the cost of alternative production that could be sacrificed else-

---

[22] Abba P. Lerner, *The Economics of Control,* The MacMillan Company, 1944.

[23] Little concludes that, "The general case against marginal-cost pricing is clearly overwhelming." See *A Critique of Welfare Economics,* Oxford University Press, London, 1950, p. 194.

[24] For an excellent review of early marginal-cost pricing literature see: Nancy Ruggles, "The Welfare Basis of the Marginal-Cost Pricing Principle," *Review of Economic Studies,* Vol. 17 (1949–50), pp. 24–46; and "Recent Developments in the Theory of Marginal-Cost Pricing," *Review of Economic Studies,* Vol. 17 (1949–50), pp. 107–126.

[25] Several of my critics have objected to this statement asserting that cases where long-run marginal costs diverge from long-run average costs are not frequent or important. Because our knowledge of cost functions for a wide range of public goods and services is so skimpy, it is apparent that neither position is presently supported by much empirical study.

where to produce this extra output. Therefore, it is desirable to expand production up to the point where price equals marginal cost. However, this solution will generate a deficit and the goals of cost-recovery come into conflict with efficiency. It is thus fair to ask why should these costs be borne by the general taxpayers when the benefits are seen by the users of the facility.* By contrast, equating price and marginal cost when the average costs of production are rising will require restricting output to the point where price equals marginal cost even though total costs could be recovered at larger outputs and lower prices. In this case marginal-cost pricing would generate surpluses.

At this point it is correct to observe that these simple sounding arguments, even though I believe they are valid, may hide a whole set of complicating circumstances and situations. As Oort suggests below, answers can be made for most of the theoretical and practical objections which can be raised against the marginal cost pricing principle. In particular, a whole set of devices have been proposed to deal with the troublesome problems of surpluses and deficits.[26] Also, economists always point out that the case for marginal-cost pricing is based upon *three* important conditions:[27] (1) that there are no important spillover effects in production and consumption which are not reflected by P=MC, (2) that the current distribution of income be acceptable or certainly not biasing the particular measurement of prices and costs, and (3) that deviations from P=MC elsewhere in the economy do not require compensating adjustments in this sector (the Theory of the Second Best).[28]

The important point here is not to go into the theoretical complexities of the marginal-cost pricing principle (many of which can be resolved by economists) but, instead, to note that the notions involved, even in their simplest form are not easily grasped by persons not well-versed in economic theory. How does one explain to a "man of affairs" or to a public servant that a deficit is desirable sometimes even though placing demand price equal to average cost may cover total costs. By the same token, it is doubtful if public utility commissions and consumers can be "sold" on the idea of producing a surplus by making price equal to marginal cost when average cost pricing in the same situation would yield lower prices and increase consumption of the service.

Moreover, if it would seem difficult to convince a public official or "man of affairs" that marginal-cost pricing is superior to average cost pricing, how much more persuasion and understanding would be required to convince people who did not believe in any price policy—ones who might believe that beneficiary charges are neither "fair" nor efficient. However, I suspect that the arguments for equity and fairness, as a basis for urging that beneficiary charges be levied, would

---

[26] See, Hirshleifer, *et al., op. cit.*, Chapter V.

[27] Most of these qualifications apply equally well to any system of beneficiary charges, particularly to policies based upon average cost pricing.

[28] William Baumol has just sent me a manuscript co-authored with David Bradford titled, "Marginal-Cost Pricing and the General Welfare Revisited." The paper contains a valuable survey of earlier work. It reaches the disturbing conclusion that systematic deviations throughout economy from marginal cost pricing will be required for an optimal allocation of resources because all taxes (except Pigouvian poll taxes) to cover deficits will unavoidably make some prices depart from marginal costs!

*Further discussion of this issue is found in the paper by Vickrey in this volume.

have more "sex appeal" among public officials, the electorate and consumers than the efficiency arguments. Yet, to me, the efficiency arguments for beneficiary charges although subtle and often complex are the ones which are most persuasive and valid.

To illustrate this point I want to quote from a summary of searching critique by Oort of the marginal cost pricing principle that is designed for economists. I believe that Oort's position would be supported by many economists, yet I suspect that few non-economists would be convinced:

> The marginal-cost pricing principle certainly does not solve all problems of welfare economics. In fact, the greater part of this essay has been devoted to an examination of some areas of policy in which marginal-cost pricing does *not* provide a complete solution. Nor is the principle under all conditions and in all cases a completely valid rule even on its own merits, i.e., as a rule to ensure an optimum allocation of resources at the margin. But in spite of these reservations, the principle is a very important tool of welfare theory.
>
> We have shown that many of the objections to the principle are either invalid, or irrelevant for policy, or require certain more or less important modifications rather than a rejection of the principle. Invalid are all those objections to the rule which are based on alleged or actual indeterminacies of the marginal cost function; in most cases, the correct interpretation of the principle, which requires in the first place the equilibrium of demand and supply, will lead to a perfectly determinate solution in terms of optimum prices and outputs even where marginal cost in indeterminate. In particular, this holds for the case of joint production in which separate marginal cost functions generally speaking do not even exist, but in which the marginal cost pricing principle nonetheless applies without any modifications. Irrelevant are those objections which apply to any and all rules of policy; this holds for the entire class of problems raised by the factor "uncertainty." Of the many modifications to which the marginal cost pricing principle must be subjected before it can be applied in practice, none appears to be so drastic as to actually imply the rejection of the rule as such. Most are in the nature of a compromise between the principle itself and some other economic objective such as the minimization of the administrative costs incurred by putting the proposed price and output policy into effect.
>
> In concluding this discussion of marginal cost pricing we may once more quote Vickrey who has summarized the issue very aptly in the following passage: "But whatever arguments may be advanced for departing in various degrees from a strict marginal-cost pricing policy, no sound pricing policy can be developed without using marginal cost as one of the principal determinants without using marginal cost as one of the principal determinants." [29]

---

[29] C. J. Oort, *Decreasing Costs as a Problem in Welfare Economics,* Drakkerij, Holland, N.V. Amsterdam, 1958. See Appendix on "The Marginal-Cost Pricing Principle." The Vickrey statement appears in his 1955 article in the *American Economic Review* cited above.

V. Cost-Recovery, Marginal-Cost Pricing and Efficiency

Before going on to the next section where I shall attempt to pinpoint the conditions under which beneficiary charges are feasible and desirable, it seems important that we examine more carefully the questions of whether cost-recovery promotes economic efficiency and also whether the marginal-cost pricing doctrine requires cost-recovery. I am not sure that economists have resolved some of the issues involved. My purpose here is to sort out some of the relevant questions on the matter of efficiency, cost-recovery and marginal-cost pricing (long run and short run).

It will be helpful to go back to some of the views expressed by Krutilla above. Krutilla argues that the marginal-cost pricing doctrine makes the investment decision "independent of the reimbursement policy for individual chunks of productive capacity." [30] He goes on to argue that initially there may be excess capacity in facilities which justifies a marginal-cost pricing policy of not recovering full costs. However, as use of capacity increases prices should be raised "to ration scarce capacity until a point is reached at which the revenue demonstrates a beneficiary willingness to pay for service sufficient to justify an expansion of facilities." [31] Krutilla goes on to say that if user-fees are raised to reflect congestion costs they will sooner or later equal the average cost of capacity: "In short, marginal cost pricing under these conditions will result in exactly recovering costs of facilities of optimal scale and schedule of expansion." [32] All of this seems to suggest that marginal-cost pricing, correctly applied, will provide for full cost-recovery and that cost-recovery is a desirable efficiency objective.

To make these relationships appear even stronger Krutilla next points out the difficulties that may arise when criteria for project design and policies for cost-reimbursement are not consistent. The design criteria require that the scale of the investment be extended up to the point where marginal benefits equal marginal costs. Unless user charges are imposed to cover the costs of the marginal unit, Krutilla argues, excess demand will develop and use of the facility will be inconsistent with the design criteria. Therefore, reimbursement policies should be made consistent with design criteria.

I am not at all sure that the policies of cost-recovery and marginal-cost pricing are nearly so compatible as Krutilla suggests when we apply the opportunity-cost concepts in a world of uncertainty and changing supply and demand relationships, especially as may be the case for long-lived facilities. There are some very difficult questions raised when we ask what are the relevant social costs in terms of alternatives sacrificed the "day before" a project is built and the "day after."

Even in a world of perfect foresight and no unexpected changes in future supplies and demands, the day after a project is built most of the resources involved may be "sunk." The alternatives sacrificed from

[30] Krutilla, *op. cit.*, p. 66. This argument is elaborated in the paper by Krutilla in this volume.

[31] Krutilla, *op. cit.*, p. 67. See also the paper by Krutilla in this volume.

[32] Krutilla, *op. cit.*, p. 68. Some of my critics think I have been unfair to Krutilla here, others agree with me. As I suggest below Krutilla is correct only if he means recovery of new facility costs in an *ex ante* sense—and not recovery of sunk costs. See the paper by Krutilla in this volume for a further elaboration of this point.

then on are usually only current operation and maintenance costs which may be very low. In such cases, prices should be set to ration supply and demand, and this may very well mean that deficits are the "rule of the day." The common answer in the literature is that eventually prices will rise to ration demand as demand increases in the future, and that the later surpluses generated will balance the early deficits so that full costs will be recovered.

I do not find this line of reasoning convincing when this argument is placed in a context of uncertainty and unanticipated change. The original investment costs the "day after" the project is constructed are historical costs—no more and no less—and they will not necessarily reflect changing supply and demand conditions and alternative social costs in the use of resources from that day on. Prices that correctly ration the use of capacity and deal with congestion may or may not return historical costs. Moreover, whether or not the revenues generated justify an expansion of facilities should not be determined by whether or not the historical costs are covered (even though such may be required for legal purposes), *but whether the revenues are sufficient to cover the costs of expansion or replacement costs at the time in the future when they are contemplated.* Clearly, future replacement or expansion costs may bear little relation to investment costs at an earlier period.

The Krutilla argument pointing toward consistency and compatibility of marginal-cost pricing and cost-recovery would seem to hold only when the investment calculations were correctly made and when the future conditions correctly forecasted. Not only can actual demands exceed or fall short of those originally forecasted, but also the costs of future replacement or expansion may be greatly different from those that governed the cost of the original investment. What I am saying is that the costs that should be "recovered" are the opportunity costs sacrificed at any time.[33] And it is highly unlikely that opportunity costs in the use, replacement or expansion of long-lived facilities will be identical to those incurred the day a project is constructed. Why, then, is recovery of original costs efficient?

If my argument has validity it would mean that the cost-recovery of historical costs should be viewed with suspicion when questions of efficiency in a world of change and uncertainty are relevant. This argument would imply that repayment contracts and price policy be made "flexible" in light of changing conditions. In other words, what would promote efficient resource use in a future world is only vaguely perceived at the time of project construction. The same logic would also apply to project operation. Sharp changes in operating rules and product mix of a multipurpose project may well be efficient and productive for the same reasons.[34] Fixed repayment contracts, fixed beneficiary charges and fixed operating rules may be necessary for legal and "finance" reasons but they may not be at all conducive to achieving

---

[33] Note the difficulties implied by this argument for privately-owned public utilities. Investors must be paid or the market for securities will be affected. Notice also that the argument also has implications for all private firms—not just those with public utility status.

[34] A similar argument was advanced earlier by Hufschmidt, *et al., Report of the Panel of Consultant to the Bureau of the Budget on Standard and Criteria for Formulating and Evaluating Federal Water Resources Developments,* Bureau of the Budget, Washington, D.C., June 30, 1961, p. 59.

efficient resource use. Although these points seem relatively straightforward from the standpoint of the efficiency of resource use, they have not, to my knowledge, been carefully analyzed in the literature of public expenditure decisions.[35]

Notice, however, that the argument has now been shifted from traditional emphasis on cost-recovery and repayment to the "rule" that public projects should be operated to maximize their net social product. The "*cost recovery*" *that takes place is the meeting of opportunity costs*. The covering of opportunity costs will recover "historical" costs only in a world where future opportunity costs coincide with historical opportunity costs. I think we can agree with Krutilla that evaluation policies and design criteria should be made consistent, as far as possible, with assumptions about future price policies. But no amount of *ex ante* rationalizing can deal adequately with problems of uncertainty and change regarding future opportunity costs.

In spite of all of this, it might be important to note that the idea of cost-sharing (as opposed to cost-recovery) appears to have a great deal of merit and could perhaps promote efficiency in terms of the *behavior* of the parties at issue. Fox suggested earlier that (1) cost-sharing provides valuable information on the demand for the public service to policy makers, and (2) cost-sharing tends to keep the claims of perspective beneficiaries in check. Note that Fox is apparently talking about cost-sharing and not cost-recovery. The arguments that we posed earlier for user charges based upon marginal costs would imply some degree of cost-sharing even though "original" costs may or may not be recovered The fact that user charges would equate supply and demand would or could provide information to decisionmakers, as well as providing "discipline" against excessive claims of beneficiaries. It seems to me also that the correct argument is *not* zero prices versus cost-recovery (or cost reimbursement) but, instead, is that prices will be used to ration service and to equate marginal social costs and marginal social benefits. Costs thus will be "shared" and perhaps "recovered" but they will probably not be original costs.

Most of this discussion is couched in terms that may be convincing to economists but, again, how does it stack up in the "real" world of affairs? If the original (financial) costs of construction are not recovered from the beneficiaries they still have to be "paid for" by someone. If it is not the beneficiaries of the project, who will return the costs of bonds with interest? In the business world, when future is incorrectly anticipated, the losses of a private firm are borne by the suppliers of equity capital and occasionally by the suppliers of debt capital. In the public world, the state governments and local governments have the power to tax, and the federal government has the power to print money as well as the power of taxation. The facts of the matter are that taxes will probably be employed to recover the costs in this case. The efficiency questions should be concerned with the possible *adverse effects of these taxes upon resource allocation versus the ad-*

[35] One of my critics argued that more firms should be nationalized in the interest of efficiency to avoid the legal difficulties of deficits arising from the dictum not to recover historical costs. The deficits could presumably be paid for out of the general treasury. However, as Baumol suggests, taxes *themselves* may introduce significant deviations from marginal costs.

verse effects of levying beneficiary charges which might return historical costs but still be inefficient.[36]

It is difficult to see a clear-cut answer to the issues posed here. As far as I know they have not been given careful treatment in the economic literature. The legal questions of how to make repayment contracts and project operating rules more "flexible" in light of changing economic conditions over the life of a project are also a source of major difficulty. Clearly, these problems of cost-sharing and efficiency are most likely to arise in the construction of large "chunks" of investment capacity with long-lives. When public investments are relatively divisible and small in relation to the size of the market, the problems of future uncertainty and the large divergencies between average and marginal costs will not exist. However, one point made here still stands: the future opportunity costs to be covered may be more or less than historical costs. How can it be that recovery of historical or original investment costs is consistent with economic efficiency either in the short run or the long run?

VI. When Are Beneficiary Charges Feasible and Desirable?

Three observations seem important by way of summary of the discussion to this point. First, we have seen that beneficiary charges can mean many things to many people. They can span the whole range of fiscal devices from the most general forms of benefit taxes to the charging of specific fees, prices or user-charges. Second, beneficiary charges can be based upon many greatly different principles such as: charge what the market will bear, maximize revenues, charge what is "fair," charge on the basis of the ability to pay (this also has many variations), charge upon the basis of average cost, base charges upon marginal costs (short-run or long-run), and charge to recover "opportunity costs."

And finally, we have noted that three major purposes or rationale have been used by the advocates of beneficiary charges: (1) *Equity*—it is "fair" to charge beneficiaries and not force the general public or non-users to bear the burden; (2) *Revenue production*—most levels of government are strapped for funds and beneficiary charges could be used to supplement other revenue sources; and (3) *Efficiency*—beneficiary charges can promote efficiency in the use and production of public goods and services.

The efficiency arguments for beneficiary charges should differentiate or separate (but seldom do) *two* notions of efficiency that tend to be lumped together. We have stressed that the principles which will promote best use of existing capacity (ration wisely) may not always provide us with answers of how much to invest in capacity in the long-run. Questions of short-run efficiency must be distinguished from questions of long-run efficiency. Prices based upon marginal costs may or may not recover historical costs. Moreover, cost recovery that is correctly based upon meeting opportunity costs will imply that the

[36] My critics are of many voices on the questions raised here. Some say that cost-recovery of historical costs involves no economic questions and that my worries here are incorrect. Others see that failure to make beneficiaries share original costs will have dire consequences. Still others see beneficiary charges as second-best choices over tax levies. Finally, some readers worry about the meaning of some of these arguments for utility firms in the private sector.

original costs are obsolete as a guide to efficiency in a world of changing supply and demand conditions.

The purpose of this section is to draw together some of these different claims and notions and to sort out the conditions under which beneficiary charges will be more suitable for some purposes than for others. Beneficiary charges should be tailored to what it is that we intend to accomplish. Perhaps this paper will serve as a springboard or introduction to a more definitive study of the role of beneficiary charges in the efficient allocation and distribution of resources.

EQUITY AND REVENUE PRODUCTION

There is not much more to say at this point about the purposes of equity and revenue production as a basis for beneficiary charges. The "fairness" of a beneficiary charge is directly related to how beneficiaries are perceived and whether we do in fact want beneficiaries to pay. Beneficiaries can be both direct and indirect. And in the latter case, the "indirectness" can be so broad or widespread as to include the general well-being of all and perhaps could be used to justify the most general forms of taxation. Presumably, if some general form of negative income taxation were instituted or if some general scheme of lump-sum income payments for persons below a "poverty line" were adopted, there would be less pressure to redistribute public goods and services to the poor. If such redistributive schemes were in force the case for more general use of beneficiary charges on direct beneficiaries would be greatly strengthened. In other words, there would be more clear separation between the distribution and allocation functions of the budget regarding the production and distribution of public goods and services.*

Yet, as Margolis has suggested, many important questions can be raised about whether public agencies and congressional committees making public expenditure decisions do reflect adequately the wishes of voters, and whether the voters are, in turn, the taxpayers and the beneficiaries. It is not obvious, for example, that decision rules now employed with respect to projects approved in the Rivers and Harbors Committee will be affected very much by types of rationale for beneficiary charges discussed here. No amount of "rational" argument will carry the day if Congress and other levels of government want to subsidize goods and services to favored clientele and pressure groups without explicit efficiency or redistributive considerations. It is simply a fact of life that beneficiaries always prefer that someone else pays. Groups with political muscle will be more successful in shifting burdens than groups without political power regardless of merit considerations on grounds of efficiency, fairness or a more desirable distribution of incomes.

As for revenue production and beneficiary charges, it is clear that different governments and different public agencies vary a great deal in their powers and abilities to raise revenues by taxation, by receipt of grants and by the use of charges and fees. The more limited are other revenue sources, the more attractive will be beneficiary charges. It is quite possible, for example, that municipally owned public utili-

*Further discussion of this issue is found in the paper by Schmid in this volume.

ties could be run at a "profit." Colberg cited a study (now quite old) which suggested that cities in Florida in 1950 transferred 22 per cent of total revenues from municipal utility operations to general funds.[37]

As Vickrey has suggested, the inefficiency generated by alternative sources of revenue production (e.g. property taxes) may be more adverse than levying user charges at levels above marginal cost, even though efficiency criteria for user charges would point toward $P=MC$. Clearly, the proposals for greater revenue-sharing by the Federal Government to State and local governments would affect the needs for revenue production by these governments. Greater revenue-sharing could mean that the emphasis on the efficiency aspects of beneficiary charges would be more likely to take priority over their use as a source of general revenues for State and local governments.

PUBLIC, PRIVATE, AND MERIT GOODS

As for efficiency considerations, there exists a great deal of confusion about the conditions under which beneficiary charges are desirable and feasible. It is helpful at this point to introduce the concepts of "public" goods, "merit" goods, and "private" goods.[38] It is technically possible in the case of "private goods" to exclude persons who are not willing to pay for the goods or service. In private goods, the benefit is received largely by the individual person or household. As a result, private production is technically possible and usually desirable. However, there are a number of "private" goods and services that are supplied by governments, particularly public utility types of services.

By contrast, a "public" good, in its pure form, is a good which is equally available to all. First, there are no feasible ways of excluding any consumer from enjoying the good. Second, the consumption of one consumer does not interfere with the amount available to all others. Classic illustrations of public goods are national defense or radio transmission which covers a whole area.

"Merit" goods have been defined in several ways and it is to be admitted that this category is not clear cut. Perhaps it is easiest to say that merit goods are private goods that have been endowed with the public interest. In the case of merit goods the individual receives more of a public service than he would have purchased on his own. Margolis states that:

> The initial attitude toward merit goods was to see them as imposed on the population by a group of moralists, or the intellectual elite, or a pressure group with power, but with a recognition that the imposition might be a legitimate activity in a democratic society.[39]

However, it now seems less restrictive to consider a merit good a private good that has some public good characteristics.[40] That is, part of the benefit is "seen" by the individual consumer and part of the benefit is "seen" by persons external to the individual or by the public

---

[37] Marshall R. Colberg, "Utility Profits: A Substitute for Property Taxes," *National Tax Journal*, vol. III. No. 4 (December 1955), p. 382.

[38] See Musgrave, *op. cit.*, ch. I or Margolis, *op. cit.*, pp. 538–547.

[39] Margolis, *op. cit.*, p. 541.

[40] Several readers have argued that I ought to abandon the concept of "merit" goods. Considering both consumption and production externalities, almost all goods are part public and part private. Public and private goods are polar cases along a continuum.

in general. Although it is possible to levy user charges, total production could be subsidized up to the extent that collective benefits were perceived. If the "subsidy" were supplied by a tax on the general public, it could be perhaps thought of as the most general form of a "benefit" tax. This would not necessarily mean that the good would have to be supplied free of charge to the individual consumer. A proper user charge to the consumer would equate marginal private benefits to marginal private cost, the extra consumption and production justified by collective benefits being subsidized by general taxes.

In some cases the external or collective benefits are satisfied by the first amounts consumed so *at the margin* the benefits received may be largely individual in nature. As a result, we must be careful to distinguish between "all or none" decisions and those involving a little more or a little less consumption. For example, it can plausibly be argued that an urban water supply confers a collective benefit to the community in the form of public health, in addition to the benefit received by individual households. On this basis, it could be argued that some form of public subsidy might be justified in support of urban water investment and consumption. Yet, all of the public health requirements might be satisfied by a per capita consumption of 60 gallons per person per day so that a consumption rate of 120 gallons per person per day should reflect only the equation of individual marginal benefits and marginal costs. However, the generation of collective benefits at the margin for merit goods could justify prices at less than marginal costs for individual consumers.

SHORT-RUN EFFICIENCY

The efficiency problem in public production is two-fold: (1) what is the optimal level of investment in capacity (long-run efficiency), and (2) what is the best use of existing facilities (short-run efficiency). Clearly, the long-run efficiency question is applicable to all forms of public production—public, merit, or private goods. Yet, the short-run efficiency question is not relevant with regard to pure public goods nor to the collective aspects (marginal ones) of merit goods. The short-run efficiency question refers to private goods where they may be a need to ration service if my consumption interferes with your consumption and to cases where the relevant marginal costs are positive. For public goods, then, not only is it technically impossible (or difficult) to exclude or to ration service, *it is also unnecessary and undesirable to do so.* The marginal costs of an extra consumer is zero (or nearly so) and there is no excess demand, and therefore, there is no need to ration. Even if it were possible to ration use, the correct price would be zero.[41]

Sometimes this point is not well-understood by students of public policy who want more cost-sharing of public expenditures employed to achieve efficiency in public expenditure. As we shall see below, there

[41] Note, that we must distinguish between the marginal cost of adding an additional TV set *once* the station is on the air and the marginal cost of changing the amount of production that is equally available to all. In the latter sense, each consumer is rationed to the quantity of the service that is actually produced. Here neither marginal utilities nor marginal costs are zero. Once the production decision has been made the marginal costs of additional consumers are zero even though the amount and kind of service for each individual is fixed.

are indeed many problems of trying to get the proper level of investment for public goods, but this problem should not be confused with the role or need for user charges in the short run.

For example, P.L. 89–72 (the Federal Water Project Recreation Act, July 9, 1965) was hailed by many as a laudable advance in providing for cost-sharing by beneficiaries of recreation and fish and wildlife investments in Federal water projects. The provisions of the act with respect to cost-sharing specify that non-Federal agencies must bear 50 percent of the separable costs allocated to recreation, fish and wildlife, and all of the operation, maintenance, and replacement costs thereafter. The non-Federal share of the separable costs can be borne in either (or both) of two ways: (1) payment, or provision of land, or facilities for the project; or (2) *repayment with interest within 50 years provided that the source of payemnt be limited to entrance fees and user-charges.*

Notice that repayments were restricted entrance fees and user-charges. We have just seen that employment of such charges is feasible only for goods of substantial "private" character where the benefit is largely individual and where it is possible to apply exclusion. Moreover, we pointed out that there is need for rationing only when my consumption interferes with your consumption and when the marginal costs of extra consumption are positive. Apart from the legitimate question of whether the State and local agencies for recreation and fish and wildlife do have the necessary legal powers to incur debt and to collect fees, it is clear that many types of recreation services and fish and wildlife services are "public" in character. It would, in many cases, not be feasible to collect entrance fees and user-charges nor would it be desirable to do so as long as congestion did not develop and cause my consumption to interfere with yours.[42] I am not suggesting that cost-sharing *per se* is undesirable in this case.

It is clear that the discipline of cost-sharing by local agencies would keep in bounds all sorts of exaggerated claims of benefits for recreation and fish and wildlife. Moreover, having to bear only 50 percent of the separable costs would not seem to be excessive when many of the collective benefits probably reside or accrue to people within the State or region. All of these points are certainly desirable in trying to achieve *long-run* efficiency in the proper scale of facilities.

However, the restriction of repayment revenues to revenues from entrance fees and user charges could well cause short-run efficiency problems to arise. Although such fees might be justified on efficiency grounds to ration service in some instances they clearly would not be applicable across the board. Imagine trying to collect user-charges when such recreation benefits may be widespread downstream on a long river. On the basis of the argument here, it would seem desirable to modify P.L. 89–72 to remove the restriction that repayments for cost-sharing be *limited* to entrance fees and user-charges. We must not confuse the rationale for cost-sharing to promote long-run efficiency with the proper role of user-charges in promoting efficiency in the short-run.

However, we should stress that an argument could and should be made for user-charges on recreation and fish and wildlife facilities

[42] Robert Haveman suggests that if a local agency recognizes that user charges are not conducive to short-run efficiency, it could choose to make the non-Federal share in the form of an outright payment.

without any reference to cost-sharing of original costs—but instead, as we argued above, to ration space and to make the users see the congestion or pollution costs they impose upon others. A colorful statement of this point of view is provided by Gaffney:

> As I recall the summer roar of Seahorse motors on peaceful lakes, the oil slicks, the loud speaker radios, the boat toilet problem, and the beer can outrage, I lose enchantment with the notion that the marginal boater is such an innocuous fellow who sheds no external costs. The marginal swimmer maybe, but how much space is he allowed? The age of genteel canoeing is dead. Man has so magnified his powers to invade his neighbor's privacy, and placed such terrible engines of nuisance in the hands of so many barbarians and adolescents who are enjoined from releasing their aggressions in useful labor that the marginal curve of psychic pollution rises vertiginously. So long as we refuse either to civilize or employ our young, the only salvation is to tax their more destructive pursuits, and certainly not to pretend they aren't bothering anyone.[43]

Before turning to the efficacy of beneficiary charges for promoting long-run efficiency in investment in public facilities, it will be helpful to emphasize three points on the role of user charges as a means of rationing use of existing capacity. First, it is clear that as congestion develops and as the marginal user imposes congestion or pollution costs on other, a public good or service may become "private" in the sense there is a need for rationing to make efficient use of a limited facility or watershed or airshed. Here the question turns to how best to ration. The second point to make here is the price is not the only means to ration. It could well be that some administrative devices could ration effectively and at less cost. This is the place in this paper to point out that the administrative costs of collection of beneficiary charges, especially user charges, may or may not be low. Clearly, this may be an important factor governing their use.[44]

The third point is that zero price for some public services may not always result in excess demand and congestion if there is some complementary factor of production employed in the process which may be rationed itself. Perhaps some examples will make the point clear. The provision of flood control protection to lands in a flood plain is a public good. The service is equally available to all persons in the plain and my consumption does not interfere with your consumption. However, the service of flood control protection is rationed to the extent that to enjoy the service one must buy or rent land (a location). The price of land will reflect the value of the service and serve to ration demand for the limited supply of lands protected.

In fact, most of the benefit over the life of the facility will tend to be capitalized in land values the day the project is built. If the flood control protection is provided free of charge to the flood plain,

---

[43] Mason Gaffney, "The Valuation of Public Goods—Discussion" in Garnsey and Hibbs (eds.) *Social Sciences and The Environment,* University of Colorado Press, 1967, pp. 154–160.

[44] Although many economists are sympathetic to Vickrey's plea for user-charges to ration the use of city streets, they are skeptical of the proposal because the administrative costs of collection are likely to be high in relation to the efficiency gains perceived. Similarly, the desirability of water meters in underdeveloped countries to ration water supply may be doubtful if the costs of the meters themselves, their servicing, and the billing system are quite high in relation to the value of water saved and water investment avoided.

the gainers will be the initial landowners. Subsequent land users will be required to pay land prices or rents that reflect changing opportunity costs and which equate supply and demand. In most cases, it would be difficult to claim that such landowners should be classified as a disadvantaged group so that the case for benefit taxation based upon land values to recover marginal costs would seem to be especially strong.

However, the fact that such windfall gains have a long history of political acceptance is indicative of the general observation that attempts to implement beneficiary charges for efficiency reasons may not be very successful. For the politician and for the man of affairs, the efficiency arguments for beneficiary charges are likely to carry less "clout" than possible rising demands for equity or for the needs for revenue production.

We might observe that the rationing effect of rents for lands or houses which may serve to ration the use of a public service may need attention for another reason. In the case of merit goods we are presumably trying to subsidize a public service, e.g., public schools. If, however, excess demand develops, land values and housing rentals may rise so that the disadvantaged groups may pay the "equilibrium" price for the service in higher rent or else they may be forced to live elsewhere. Zero-prices of services may thus generate "shadow" prices in complementary factors which are inelastic in supply. Although the rationing effect may be commendable from an efficiency point of view, we may not achieve the redistribution of income we sought in provision of the merit good. This "perversity" would be accentuated if the landowners and landlords received untaxed windfall gains in increases in housing and land values.[45]

Several interesting points seem to come out of this discussion of the fact that zero-prices to users do not necessarily imply that there is no rationing involved in the consumption of the service. One observation is that the real or shadow price to the user may be positive. Therefore, it is not always clear, as the literature often asserts, that the marginal benefits will be zero leading to waste because the service is "free" when marginal costs are positive. In fact, the effect of nonprice rationing or the rationing of auxiliary factors may serve to stop consumption far short of the point where marginal values in use are zero, even though user-charges are zero.

A second point concerns the conclusions we can draw from attempts to construct demand curves and estimate benefits on the basis of revealed behavior (consumption data) at zero prices.* Suppose, for example, that actual consumption of a service is Q_1 in a zero-price situation. Without additional information on the degree of rationing (administrative, congestion, or the indirect rationing in the price of complementary resources) we do not know whether to place Q_1 at point A on the implicit consumer demand curve or whether it lies somewhere to the left of point A, say at B in figure I below. The nature of the conclusions we can draw about marginal benefits in zero-price cases,

[45] See, J. W. Milliman, "Land Values as Measures of Primary Irrigation Benefits," *Journal of Farm Economics*, Vol. 41, No. 2 (May, 1959); also "Land Values—A Further Comment," *Journal of Farm Economics*, Vol. 42, No. 1 (February, 1960); Margolis, *op. cit.*, p. 546; and Gaffney, *op. cit.*

*Further discussion of this issue is found in the paper by Margolis in this volume.

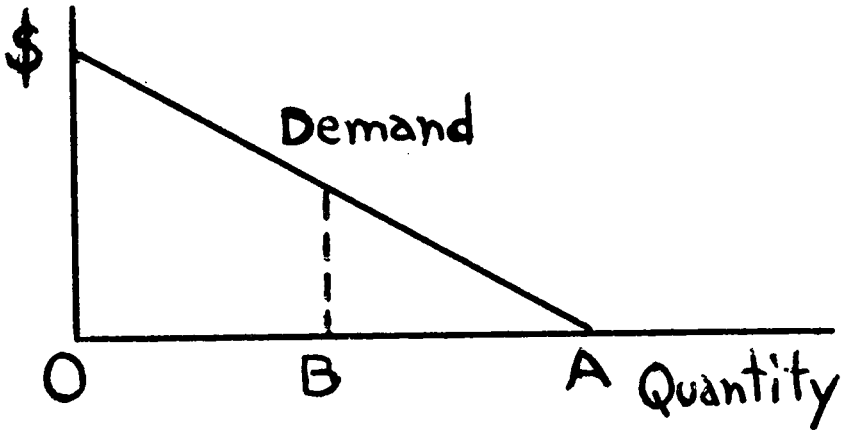


FIGURE I

and hence what we can infer about waste, is thus not intuitively obvious. Greater consideration needs to be given to the possibilities of hidden or implicit rationing in the analysis of revealed behavior in zero-price situations.

Clearly, this discussion does not refute the point made by Krutilla that subsidized facilities for one set of producer goods may distort the demands for auxiliary factors of production so that the inefficiency effects may be found in the use of other inputs and investments, not just in particular service provided.[46] But, it does mean that the question of possible inefficiencies generated in zero-price situations is apparently much more complicated than the existing literature suggests.

LONG-RUN EFFICIENCY*

The efficacy of beneficiary charges for long-run efficiency is based upon their role in the promoting proper scales and levels of investment. Here the role is twofold. The fact that beneficiaries will be charged will provide some "discipline" on the claims of benefits and will tend to force beneficiaries to consider these benefits and costs (the charges) in relation to alternative uses of their resources. If the burden of costs cannot be shifted, there will be less tendency to "puff" claims of benefits and the relative allocation of factors will be improved.

The arguments made earlier that historical costs are not relevant to price policy based upon opportunity costs clearly raises many troublesome questions when we leave the static world and view investment in public facilities as a continuing process. Costs that may not be marginal in respect to output do become "marginal" in respect to the total condition of whether to produce at all. Are the long-run marginal costs that we should be concerned with merely the derivative of the long-run total cost function with respect to output when the question is one of whether to produce at all? Also, are there not serious be-

[46] Krutilla, *op. cit.*, p. 69. See also the paper by Krutilla in this volume.

*Further discussion of this issue is found in the papers by Schultze and Krutilla in this volume.

havioral implications stemming from the dictum of ignoring historical costs the day after a project is constructed? Can you tell beneficiaries and public agencies that costs are important before a project is built and then not require cost-recovery the day after?

Unless substantial payment for benefits is required from beneficiaries or from the jurisdictions in which they reside, the forces to "discipline" public investment decisions will be very weak. Clearly, when the discipline of the market is absent, there are serious problems of how to obtain responsible public investment decisions. We lack measures to reward good decisions and to penalize poor ones.

The "discipline" argument appears to be a powerful one, but there is not a great deal that I can add here. Further study of this line of thinking would be in line with the suggestions to further explore political and behavioral models of the demand for public services.

Closely related to the "discipline" effect is the "information" effect cited above by Fox. That is, the application of beneficiary charges would provide decision-makers with valuable information about the benefit functions for public services. Provision of services at zero prices and failure to impose cost-sharing makes it extremely difficult to estimate benefits received by users and to make careful project evaluations.[47]

Just how difficult it is to estimate benefit functions in zero-price cases is not usually appreciated. Yes, there does exist some literature on the problem of how to get the public to reveal preferences, as we suggest below, but the empirical problems of benefit estimation in zero-price situations are quite formidable. For example, suppose we have some revealed behavior on library circulation at zero-price in an attempt to quantify some of the benefits of a public library. If the circulation is Q1, we have to decide first whether to place it at point A or point B on the diagram above. But, then, how do we generate some sort of demand curve as a basis for benefit estimation? Unless we have another point, it is clear that we have little notion of the area to be measured and that points A and B could be consistent with an infinite number of demand functions. Benefit-cost analysis of public services supplied at zero prices is an incredibly difficult task, even when we have acceptable measures of the output unit. Revealed behavior at zero-prices is extremely difficult to translate into meaningful benefit functions.

In general, I find these two arguments persuasive, yet I think that a great deal more study of the matter is required before we recommend the adoption of beneficiary charges for a greatly expanded range of public services. For goods that have substantial "public" content, the practical questions are how do we get the public to reveal their preferences and how do we collect beneficiary charges that may promote long-run efficiency and yet not cause short-run inefficiencies.

For "private" goods, there is somewhat less of a problem. The people or consumers have to reveal their preferences because no other beneficiaries will do so and if "I do not pay, I won't get the good." The efficiency problem with "private" good production in the public sector

[47] Roland McKean argues that admission charges supply important information about the nature of public demands and might be justified on the grounds of being the cheapest way for decision makers to gain information about marginal evaluations by consumers. See *Public Spending*, McGraw Hill, Inc., New York, 1968, p. 73.

is how to reconcile desire to recover costs with the need to meet opportunity costs. As I suggested above, the theoretical conditions necessary for the reconciliation of these two kinds of "cost-recovery" are not likely to be found in actual practice. However, we would hope that revenues generated by "correct" pricing policies would still recover some or all of the original costs so that the users would be more likely to bear the costs (as well as enjoy the benefits) rather than shifting the burden to other groups.

The application of beneficiary charges to public or merit-type goods may not be quite as bleak as might appear from the general discussion of public goods. As Brazer, Gaffney, and others have pointed out, many kinds of public goods are, in fact, not equally available to all, but available only to particular regions, areas, or user groups. Although it may be difficult to estimate the amount of a policeman's service used by a particular family on the beat, it is clear that the policeman on a given beat provides protection primarily to that beat and only secondarily to other parts of the city and to other cities in the same region. For the commonly used example of a public good, the lighthouse, the service is not provided to all, it is provided only to boats in the vicinity of a given lighthouse. Flood-control protection provides benefits mainly to users of the flood plain, fire protection to one area is not fire protection for the whole city, television signals may benefit only a limited service area and so on.[48]

All of this suggests that it may be easier than many writers believe to identify beneficiaries of a particular public good. If this is the case, the classic ingenuity of the American political system could be called on to devise kinds of collective decisionmaking institutions to get preferences revealed and to see what types of beneficary charges are desired. Clearly, the formation of regional water quality authorities or regional air-pollution authorities can make steps in this direction. Collective services can be provided for the area as a whole, yet financed in large part by taxes on local wealth and incomes. There is often little reason to assume that the benefits are so widespread as to be national in scope and therefore to be paid for out of the Federal treasury.

VII. CONCLUDING REMARKS

When I agreed to write this paper, I thought that the subject of the relation of efficiency to beneficiary charges was relatively straightforward and that a paper of about 20 pages would probably suffice. However, I was worried about the lack of empirical studies to illustrate the great inefficiencies that are often claimed to result from the failure to employ beneficiary charges.

The more I examined the literature and the world about me, the more I realized that the question of beneficiary charges and efficiency was not well worked out in the literature. The cases discussed were highly oversimplified, and they also skirted over the conflicts between long-run and short-run efficiency. Some of the lack of empirical studies can undoubtedly be attributed to the argument that "we do not know

[48] James Buchanan has recently advocated that people should have two roles—one as a taxpayer-purchaser and one as a consumer-purchaser with separate marginal tax prices and marginal user prices. These suggestions have many attractive properties, but it is difficult to see how they could be implemented. See: James M. Buchanan, A Public Choice Approach to Public Utility Pricing," *Public Choice*, Vol. V, Fall, 1968.

what would have happened to resource allocation if we had employed user-charges instead of zero-prices." This argument is a valid one. But I now believe that the lack of empirical work also stems from the fact that the theoretical relations between beneficiary charges and optimal resource allocation are simply not well understood. Until more work is done on the theory of beneficiary charges in relation to the theory of demand generation for public services, the empirical work will lag and our confidence in public policy recommendations on the matter will remain uncertain.

Even though the literature proved to be less helpful than I had first thought, an opposing reaction also soon became apparent. I began to see possible opportunities for the application and misapplication of beneficiary taxes and of direct user-charges all about me. Possibilities for the application of beneficiary charges were literally everywhere—to ration airsheds, national parks, water quality, highway use, highway investment, city streets, most urban public services, public housing, urban renewal, postal services, hunting and fishing licenses, college facilities, hospitals, doctors, and so on. In fact, the whole gamut of public functions usually comprehended by the allocation branch of government seemed to be involved.

At this point, it became obvious to me that I had a "bear by the tail" and that the paper would soon evolve into a lifetime job. I have, therefore, written an essay which hops around the general terrain coming across a commonplace idea here and an interesting idea there. I hope that the paper is sufficiently well developed to give the man of affairs a feel for the problem, and yet to stimulate my fellow economists to do some definitive work on the theory and practice of beneficiary charges.

However, theoretical advances will not be enough; possibilities for greater application of beneficiary charges to achieve greater efficiency in the allocation of resources are contingent upon two fiscal reforms: (1) greater revenue sharing by the Federal Government to other levels of government, and (2) greater assumption by the Federal Government of broad distributive types of programs, such as the negative income tax. Until greater progress is made along these two fronts, the questions related to beneficiary charges and efficiency will be subsidiary to questions relating to the influence of beneficiary charges on equity and revenue production.

Section C

Attaining Public Expenditure Efficiency in a Federal System

THE OPTIMAL ALLOCATION OF JURISDICTIONAL RESPONSIBILITY: THE PRINCIPLE OF "FISCAL EQUIVALENCE"

BY MANCUR OLSON, Jr.

Mancur L. Olson, Jr., is Associate Professor of Economics at the University of Maryland. Until March, 1969, he was Deputy Assistant Secretary (Social Indicators), at the Department of Health, Education, and Welfare.

The establishment of institutional arrangements with appropriate functions, sizes, and incentives is a basic step in implementing sound public policy. Professor Olson stresses that it is essential that the economic analysis of proposed public undertakings consider explicitly these institutional questions for "if a particular function is assigned to the wrong type of institution, the incentives that the relevant decision-maker faces may keep him from doing all that he could to serve the public, however superb his analytical apparatus might be."

The disparate pattern of jurisdictions among national, State, county, and city governments, with their overlapping network of functions raises the question of the optimum pattern of jurisdictional responsibility. In this paper, Professor Olson discusses questions pertinent to determining the optimal match of Government activities and appropriate governmental boundaries. He discusses the conditions which tend to lead to both over- and under-production of public outputs. Both the boundaries of the Government supplying the service and the patterns of beneficiary charges and cost-sharing are pertinent to this discussion. He concludes that the wide differences in the extent to which the impacts of public goods are localized implies the desirability of a substantial number of various-sized, independent governmental jurisdictions. These considerations of the boundaries of public goods outputs together with costs of different sizes of governments, provide a rationale for and guidance to the implementation of grants-in-aid programs.

I. *Introduction*

The analytic, problem-solving approach of the planning-programing-budgeting system cannot realize its full potential unless it considers, not only alternative public programs and expenditure levels, but also alternative institutional arrangements for the achievement of public purposes. These purposes can sometimes best be achieved by local governments, sometimes by State governments, sometimes by the central government, and sometimes by a special public or semi-public authority or corporation. If a particular function is assigned to the wrong type of institution, the incentives that the relevant decision-maker faces may keep him from doing all he could to serve the public, however superb his analytical apparatus might be.*

The question of what the division of responsibilities among different levels and types of governments ought to be is one that very much needs to be asked, for it has never been adequately answered. Though we have good economic theory that can help us decide what functions might call for public action, and what ought to be left to free market

*Further discussion of this issue is found in the paper by Schultze in this volume.

action*, we have no general theory, in economics or any other discipline, that will give us much help in deciding which level of government should handle a given public function.[1] Neither is there any model that will tell us what role subsides from one level of government to another ought to play in an ideally structured system of government, or whether any subsidies should be of the "categorical" variety or of the "block grant" type**.

Perhaps partly because of the lack of satisfactory theory, there is a wide range of thought (or ideology) in this area. There are advocates of the maximum possible reliance on the level of government that is (physically) closest to the people, of metropolitan government, of "states rights," of unitary national government, and of world government. At various times and places there have also been many advocates of "functional" units of government, which have responsibility for a particular problem, occupation, or industry, rather than a given geographical area. This point of view was characteristic of many pluralists, syndicatists, guild socialists, and believers in the corporate state, and is still part of the official ideology of the Gaullist movement. The debates on this subject have lately taken on a new twist as the "new left" has forsaken the "old left's" emphasis on central planning and nationalization in favor of "participatory democracy," which appears to involve a degree of decentralization at least as great as the old right or classical liberals have advocated. The modern demands for what might be called "black power separatism," or separate governments in Negro ghettoes, have added yet another approach, which we shall give special attention here.

Just as opinions on this subject are diverse, so is the practice, even within the United States. There are, by one count, about 80,000 different governments in this country.[2] The New York Metropolitan region alone is supposed to have 1,400 governments. The typical citizen is not only under the jurisdiction of national, state, county, and city governments, but sometimes also subject to a metropolitan transport commission, a port authority, a sewage or sanitary district, a Soil Conservation District, a pollution control district, a school district, an airport commission, or a metropolitan planning commission or council of governments. The citizens of Fridley, Minnesota, for example, are subject to eleven different governments: The city of Fridley, an independent school district, the North Suburban Sanitary Sewer District, the Minneapolis-St. Paul Sanitary District, the North Suburban Hospital District, the local Soil Conservation District, Anoka County, the Minneapolis-St. Paul Airports Commission, the Metropolitan Mosquito Control District, the State of Minnesota, and the United States

[1] There is, however, one respect in which economic theory is not yet sufficient for answering this sort of question. Though it tells us when unregulated private enterprise will not work optimally, it does not tell us whether public action in a given case will improve the situation or make it worse. Since markets are almost always imperfect in one way or another, and therefore do not normally work optimally, this lack is rather serious. Any presumption that private enterprise should be used whenever feasible must therefore rest, not simply on the virtues of market mechanism, but also on some general conception of the shortcomings of public enterprise. Economic theory offers no general conception of the strengths and shortcomings of public enterprise, and if therefore far from a sufficient basis for decisions on which functions ought to be handled in the public sector.

[2] Committee for Economic Development, *Modernizing Local Government* (New York: 1966), p. 17.

*Further discussion of this issue is found in the papers in part I of this volume.

**Further discussion of this issue is found in the paper by Mushkin & Cotton in this volume.

of America.[3] Many of the governments named are obviously *ad hoc* institutions, and such single purpose governments are commonplace throughout the land.[4]

At the same time that the nation has this bewildering and overlapping network of local and functional governments, it also relies to a great extent upon the federal government, whose direct general expenditures exceed those of all state and local governments put together.[5] The national government has also been coming to have a larger share of responsibility over time for some types of governmental activity.

The situation is, however, not simply or mainly one of Federal activity supplanting a multiplicity of decaying state and local governments. New local governments and functional authorities are being created, and since 1948 the Federal government's share of total government expenditures (under any one of six alternative definitions), and of total government employees, has declined, especially in relation to that of local governments.[6]

The network of grants or subsidies from higher levels of government to lower makes the situation even more complex. By one count, there are over 400 different Federal grant-in-aid programs. Such grants are also becoming more important. Since 1957, the amount of Federal grants-in-aid to state and local governments has been increasing at a compound rate of more than 14 percent per year.[7]

The labyrinthian, overlapping network of state, local and functional governments, dependent to an increasing degree on Federal subsidies, is widely criticized. It is considered a waste of resources and an obstacle to intelligent planning and coherent public policy. However much their recommendations differ, most observers seem to agree that the present pattern of governments is duplicative and unsatisfactory. The prestigious Committee for Economic Development, for example, has recommended an 80 percent reduction in the number of local governments.[8]

What principles ought to guide the development of a rational pattern of jurisdictional responsibility? Is a unitary national government a necessary condition of efficiency? Or a systematic reliance on small, local governments with rational boundaries? Is there any general or theoretical case to be made for subsidies from central to subordinate governments? Or for functional governments or other single purpose public authorities or institutions? It is to these questions that this paper is addressed.

II

There is one minor issue that must be settled immediately, lest the search for an optimal pattern of jurisdictional responsibility become trivial. This issue grows out of some stimulating writings on how free

[3] CED, *Modernizing Local Government*, p. 12.

[4] Some professions in some states are also governed by organizations that are ostensibly private, but nonetheless have the legal authority to regulate a particular profession.

[5] George F. Break, "Changing Roles of Different Levels of Government" (forthcoming), Table I, page 3.

[6] Break, *op. cit.* U.S. Department of Commerce, *Congress and the Nation*, p. 1393; and Robert A. Dahl, *Pluralist Democracy in the United States* (Chicago: Rand McNally, 1967), p. 177.

[7] Break, *op. cit.*

[8] *Modernizing Local Governments*, p. 17.

bargaining between firms and individuals can solve some problems that economists previously assumed could not be optimally solved without government intervention. These writings emphasize that even when there are external economies or diseconomies (social benefits or social costs which do not show up in the prices the relevant actor receives or the costs he has to pay), *laissez faire* can still yield an optimal result for the society, since the parties involved with the externalities can bargain among themselves until they arrive at an arrangement which takes account of the externalities.* There is, it is argued, reason to hope that such bargains will in fact deal optimally with externalities, since the parties concerned will always have an incentive to continue bargaining until all mutually advantageous bargains have been struck, at which time Pareto-optimality (a situation which is fully efficient, in that no one can be made better off without someone being made worse off) will have been achieved. The line of argument has naturally had considerable appeal to those who believe in *laissez faire*, since it seems to show that even those problems that were supposed to demand government action will be ideally solved if only the parties are left alone.

Most of those who emphasize the foregoing argument are, however, careful not to press it to the extreme, and are quick to bring in bargaining costs or other considerations that limit the applicability of their argument. In this they are wise, for the bargaining-will-continue-until-Pareto-optimality-is-achieved assumption could prove the efficiency, not only of *laissez-faire*, but of anarchy as well. If free bargaining would always lead to Pareto-optimality, all desirable public goods could be provided through voluntary agreements. Since violence is never Pareto-optimal (unless it be between a sadist and a masochist), anarchy would be peaceful. The distributional implications of an anarchic but Pareto-optimal society would not be appealing,[9] but then neither might those free bargaining leads to in a *laissez-faire* economy.

What has just been said is not altogether frivolous, as is evident when we think of the implications of the bargaining-brings-Pareto-optimality assumption for our search for principles of government structure. If, for example, there were no central government there would be no loss of efficiency under the assumption of utopian bargaining, since local governments could through bargaining make national arrangements to take care of any uniquely national problems. Similarly (as will be obvious later) ideal logrolling in the national legislature could deal optimally with any local problem even in absence of local government.

Another reason why the probably frequent failure of bargaining must be emphasized here is that where problems of government are at issue, the number of interested parties often numbers in the thousands, or even millions. These large numbers not only raise what might be called the "transactions" cost of bargaining, but also pose an additional problem. A large group can bargain with others only if it is organized, so that it can have representatives. But such an organization is itself a public good to the group it would help, and won't

[9] Those who could coerce others would enjoy extortionate gains. The anarchy might also be unstable, and could degenerate into tyranny.

*Further discussion of this issue is found in the paper by Demsetz in this volume.

exist unless it enjoys the power of coercion, or has some other way of giving its members an incentive, on an *individual* basis, to support the organization.[10] Thus in many contexts where governmental action is at issue there may be no real possibility of a "bargaining" solution, since the groups among whom bargains could be struck aren't organized.

III

The subsequent analysis will be furthered by some more special assumptions, some of which, as we shall see, can later be dropped. The first assumption is that the governments, authorities, and other institutions we consider produce only collective or public goods, which are defined for *this* purpose as goods such that it is not feasible to exclude non-purchasers from their consumption (defense and smog control, for example).* A second assumption is that every collective good affects some clearly delineated group or area, which can be as large as the entire population of the earth, or as small as the population of the smallest community. A breakthrough in pure science could benefit the entire world, and a program to combat air pollution could help only a particular community. Both of these examples satisfy our assumption rather nicely, but some other governmental services that are sometimes called public goods would not fit this assumption well, and we shall later have to relax it. A third initial assumption is that there is no complementarity in production among different public goods, in the sense that a government that provides one collective good cannot therefore provide a second collective good at lower costs than a unifunctional government could have. This assumption is probably also unrealistic in certain cases, and will later have to be relaxed.

The argument that will be developed with the aid of these and other assumptions would better be described as a "pre-model" than a "model," since its purpose is to set out a rudimentary and hopefully fundamental approach that should inspire the construction of more elaborate and realistic models based on less restrictive assumptions. There is perhaps some basis for the hope that this approach has promise in the fact that the two other economists, Jerome Rothenberg and Gordon Tullock, have independently set out the essentials of arguments which, however different in some respects, in some ways parallel my own.[11]

The approach to be set out focuses on the problem of allocative efficiency, rather than questions of stabilization and income redistribution, principally because stabilization and redistribution are obviously most appropriate to central governments. We ask here, "What are necessary conditions for the allocative efficiency in the provision of collective goods?"

In every case, one or more of the following three logically possible

[10] This point is demonstrated in my *Logic of Collective Action* (Cambridge: Harvard University Press, 1965).

[11] See the forthcoming Universities-National Bureau of Economic Research volume on The Economics of Public Output, where Rothenberg, Tullock, and I discussed our approaches to this problem. See also my discussion of papers on "Efficiency in the Government Sector," in *American Economic Review, Papers and Proceedings* (May 1964), LIV, 250–251, and Gordon Tullock's article on "Problems of Scale."

*Further discussion of this issue is found in the papers by Steiner and Arrow in this volume.

relationships between the "boundaries" of a collective good and the boundaries of the government that provides it will apply: (1) the collective good reaches beyond the boundaries of the government that provides it; (2) the collective good reaches only a part of the constituency that provides it; or (3) the boundaries of the collective good are the same as those of the jurisdiction that provides it.[12]

In the first type of situation, there is an obvious cause of inefficiency. The fact that the benefits of a given government's activity reach beyond its borders means that it provides an external economy or "spillover," and when any maximizing unit reaps only a part of the benefits of its action, it tends to carry on its activity at a less than Pareto-optimal level. In other words, situations of this kind leave a society with a smaller supply of the collective goods in question than would be optimal. There is some evidence that this actually happens in the case of military alliances, where one nation's military expenditures are a spillover to its allies, and in city-center governments which provide services to the whole of a metropolitan area.[13]

Now let us turn to the second possible type of situation, in which a collective good reaches only a subset of the population in a jurisdiction. Here there is what might be called an "internality," which normally has the same effect as the externality just described. To see this type of case in its most obvious form, suppose that the United States had a unitary government with no state or local jurisdictions. There would still be a need for local collective goods, such as air pollution control in particular metropolitan areas.

In a situation of this type, and a democratic political system with voting by majority rule, the provision of a collective good for a local area will hurt more people than it helps, even if the total benefits from that good exceed its costs, and Pareto-optimality would thus require that the collective good be provided. If the benefits are local and the taxes national, even a collective good which it is Pareto-optimal to provide will still create more losers than gainers. It is always possible to set up a special local taxing authority, but this will be at least a *de facto* government.

If all mutually advantageous bargains were struck, logrolling would insure that all collective goods that it was Pareto-optimal to provide would be provided. But we have already argued that, especially where large groups of people are at issue, it will very often be the case that logrolling will not happen, and that there will not be a Pareto-optimal supply of public goods.[14]

---

[12] Possibilities (1) and (2) could both hold at the same time. It will soon be clear that this is particularly likely to lead to inefficiency.

[13] See Mancur Olson and Richard Zeckhauser, "An Economic Theory of Alliances," *Review of Economics and Statistics*, XLVIII (August 1966), 266–279; and unpublished research on central city and suburban public expenditure patterns by Mr. Harvey Botwin of Claremont College.

[14] In the United States Congress, logrolling probably leads to a greater than optimal expenditure when projects of a "pork barrel" type are at issue. In most of these cases the projects are of a tangible, if not monumental, type, and a Congressman is more likely to be identified in his district with such a project than with a general tax increase, which could not in any case usually be traced to any one package of local projects. The logrolled projects, moreover, normally fall under the jurisdiction of only one committee in each house, and this is often necessary for logrolling.

The difficulty of logrolling across committee lines is illustrated by the fact that legislation which makes "sense" only if it applies only to a small number of areas, often can pass only after most Congressional districts are made eligible for its direct benefits. If logrolling could readily occur with legislative proposals involving a number of committees, there would presumably be no such tendency. I am thankful to Charles Schultze for calling the difficulties of logrolling across committee lines to my attention.

Even when logrolling does occur, there is a likelihood that it will not provide a Pareto-optimal supply of public goods. This is because any sharing of the marginal costs of a bundle of Pareto-optimal projects which is not exactly proportional to the sharing of the gains at the margin will lead to a less than Pareto-optimal expenditure.

This, in turn, is because, under any other sharing of the costs, some member or members of coalition will be paying more than their share, and will find that marginal costs to them come to exceed their marginal benefits *before* a Pareto-optimal amount has been provided for the society. They will then have an incentive to hold back.[15]

There is one special case in which benefit boundaries smaller than jurisdictional boundaries will not lead to less than a Pareto-optimal level of public expenditure. If X percent of the voters in a constituency are required to vote affirmatively before a measure is declared passed, and a given collective good benefits more than X percent of the voters, there may be a tendency for provision of a supra-optimal level of this good. This is because (as Gordon Tullock has shown) the minority that does not benefit from the good also pays taxes, and those who do enjoy the good can vote for a level of provision that equates only their own tax burden at the margin with the total benefits of the collective good.

If there is this exception to the generalization that a failure of government boundaries to match the boundaries of collective goods leads to a less-than-Pareto-optimal supply of collective goods, it is still true that there are systematic forces, of one direction or another, which work against allocative efficiency in any situation where the boundaries of a government and a collective good it provides do not coincide. We must, then, argue provisionally (the argument will be modified somewhat later) that there is a need for a separate governmental institution for every collective good with a unique boundary, so that there can be a match between those who receive the benefits of a collective good and those who pay for it. This match we define as "fiscal equivalence."

Though it is much too early to make policy recommendations, it is already evident that *both* the "centralizing" and "decentralizing" ideologies are wrong, or at any event entail inefficiency. Only if there are several levels of government, and a large number of governments, can immense disparities between the boundaries of jurisdictions and the boundaries of collective goods be avoided. There is a case for every type of institution from the international organization to the smallest local government. It is a merit of the present approach that it can help explain the need for both centralized and decentralized units of government in the same framework.

It is also already evident that some of the complaints about the proliferation of governments, and the overlapping boundaries of different types of governments, need to be greeted with skepticism. Though there are undoubtedly some redundant governments, and some that are too small to serve their purposes properly, there is also a need for new governments as well, including some small ones. The metropolitan transport commissions, the port authorities, the Soil

[15] If bargaining proceeds solely in "all or nothing" terms, and marginal increments are never considered separately, the situation becomes more complex. But there would presumably still be some tendency in the same direction.

Conservation Districts, the suburban towns, and the neighborhood school boards that have been created are in many cases probably a response to the forces described in this paper. The forces we have described presumably have meant not only that expenditures on some public problems were at a less than Pareto-optimal level, but also that some public problems weren't dealt with at all. This, in turn, must have led to the creation of new authorities or governments.

V

Patterns of residential segregation can also justify additional governments, if (and only if) it is assumed that integrated housing areas are not now a feasible (or desirable) alternative. Since different racial and ethnic groups often have different cultural backgrounds and tastes, they may want different types of collective goods. In cases where the sense of ethnic identity is very strong, or where there is antagonism among different social groups, this is particularly important. People may then care passionately, not only about the level of expenditures on policemen, teachers, and other public authorities, but also about whether there is a civilian police review board or courses in Afro-American history, and even about the race or religion of public officials at every level. The characteristics of the collective goods may affect not only the subjective evaluation of the services provided, but also their objective effects, as is evident whenever differences in language, accent, or social background impede communication between officials and citizens. The huge role of ethnic factors in American political history, especially at the level of city government, and the tense state of police-community relations in many Negro ghettoes today, suggest that ethnic and religious factors can have a decisive impact on the nature of the collective goods that are demanded.

Though this question is much too complex to be settled here, it is clear that there is a case for a separate governmental institution for each ghetto, whenever group differences in taste for collective goods are important and the patterns of segregation in housing are not amenable to change. The collective goods then have relatively well defined boundaries given by the boundaries of the ghetto, and governmental institutions which conform to these boundaries are a necessary condition for Pareto-optimal provision of public services.

There is no assertion here that patterns of segregation in housing should or must be maintained, and no denial of the possible public interest in common arrangements for education and law enforcement which might ultimately lessen cultural and educational differences among ethnic and religious groups. Where education of children is at issue, the desire to satisfy consumer (parent?) preferences may be subordinated to a more general interest in the future strength and productivity of the nation. For these and other reasons, no general policy conclusion can be derived here. But the possibility of cases where matching boundaries of governments and ghettoes are a necessary condition for economic efficiency cannot be denied.

VI

We have not yet considered the relationship between the size of a government and the production cost per unit of the collective good it

provides. This relationship obviously must have something to do with defining the optimal structure of government and the division of responsibilities among different levels of government, so we must consider it now.

In many cases, a government with boundaries that match those of the collective good it provides will find that it produces that collective good under conditions of decreasing unit costs. The fact that the good could be provided to a larger group at lower costs per person does not, however, modify the conclusion that there is a need for the "fiscal equivalence" that this paper has advocated. If all of those who would benefit from a given collective good are already in the jurisdiction that provides it, there could be no point in providing it to others.

The only case where the relationship between the costs per unit of the collective good and the size of the clientele served would modify the argument developed above is that in which a government big enough to comprehend all the benefits of its activity would have higher unit costs of production for the collective good than a smaller government. Here we have the only case where economies and diseconomies of scale in production play an independent role.

An example of this sort of situation *might* be education. Suppose it is true that a national school system would be so bureaucratic and cumbersome that it would have higher costs per student educated than a system with thousands of independent school boards. Even granted this supposition, there still appears to be some case for Federal Government involvement in education, since some of the benefits of education spread throughout the land (and, with "brain drains," even abroad). If children in one locality get a poor education, this could make the national democracy ultimately work less well. If the children of that locality often migrate to another locality, this poor education becomes a problem for the recipient community. Thus each local school district finds that its educational expenditures provide an external economy to the rest of the Nation, and it therefore will spend too little on education.

What normative precept is available when the desire to internalize externalities calls for central government provision, and diseconomies of large scale operation call for local provision?

The necessary condition for Pareto-optimality is then local governments of a size that minimize unit costs,[16] and central government grants to these local governments that are just large enough to compensate the local government for the external benefits of its expenditures. The grants to the local governments must involve the sharing of the marginal costs of each additional unit of service in the same proportion as the benefits of that additional unit are shared between the locality and the Nation at large. If a local government activity provides no benefits beyond its borders it should get no Federal subsidy (except for redistributive purposes, and this should be in the form of a lump sum or block grant, rather than one in which marginal costs are shared). If a local government produces something from which it gets no special benefit, the Federal government should pay the full

[16] Unit costs may be a function, among other things, of the spatial distribution of the population and the level of per capita income in each area, both of which may be affected in turn by the level of taxes and public services in different jurisdictions. A formulation as simple as that in the text is satisfactory only if the possible interdependence between decisions about government boundaries and the provision and financing of collective goods, on the one hand, and the spatial distribution of population and economic activity, on the other, is ignored. I am indebted to Robert Hartman for helpful discussions of these issues.

costs, as it does with other contractors who produce goods needed for national purposes.

There is, then sometimes a case in principle for Federal subsidies to state and local governments,* because only through such subsidies can we obtain both the advantages of "fiscal equivalence" and also those savings that occasionally accrue from small scale production of collective goods. In a period when improved transportation and communication, and increased geographic mobility, probably cause the external benefits of some local government activities to rise, we should not be surprised to see an increase in Federal grants-in-aid. Nor should we assume that great increases in subsidies of this type indicate that the local governments are obsolete: they could mean they are the most efficient purveyors of some types of collective goods. The demands for Heller-Pechman revenue-sharing plans may also owe something to this apparent increase in the external economies of local governments. But revenue-sharing, though sometimes appropriate for other purposes, is not a suitable means for dealing with this phenomena, since only grants which involve sharing the marginal costs of the activities with spill-overs will provide the appropriate incentives to local governments.

VII

If some of the assumptions of the argument are relaxed, the situation becomes complicated. There is ultimately a need for an elaborate model that explores the implications of a wide variety of assumptions. But a rough idea of the effect of relaxing particular assumptions is easily gained.

It is clear that a relaxation of the assumption that there is no complementarity in the processes by which collective goods are produced tends to weaken the presumption that there should be a separate government for every collective good with a unique boundary. Some boundary discrepancies would be justified if collective goods could be provided more cheaply by governments that provided two or more such goods.[17]

The assumption that the boundaries of a collective good are given also needs to be relaxed, if the full range of what are usually called public goods is to be considered. Some of the goods that governments characteristically provide are such that it is sometimes practical to exclude non-purchasers from their consumption. These are usually goods which can be provided either by the private or the public sector. Garbage collection might be an example of such a good. Though uncollected garbage at one residence may create problems for those next door, it is also true that garbage collection is left to private enterprise in some communities, and that a public garbage collection activity can have a wide variety of boundaries.

What normative precepts would be relevant in such cases? The assumptions would have to be spelled out in greater detail to give a fully satisfactory answer. Where a government supplies a wholly private good, the logic of the argument here would demand that the provision

[17] One way in which a multi-functional government may be able to cut costs is by reducing the need for legislators, officials, and elections, and saving voters the trouble of keeping informed about elected officials.

*Further discussion of this issue is found in the paper by Mushkin & Cotton in this volume.

of the good be left to the private sector. Where there is a quasi-collective good that can feasibly be provided to areas with diverse boundaries, there is a good case for having the boundaries of the jurisdiction determined by the consideration of production cost. The size of jurisdiction, and its particular boundaries, would be set in such a way as to minimize the total cost of producing the desired good. Such external benefits as then spilled beyond the boundaries so determined should be covered by subsidies from a higher or more inclusive level of government.

VIII

It is rarely appropriate to go directly from theory to policy. Reality is almost always too complex to permit policy recommendations derived solely from a single model, much less a "pre-model" of the sort adumbrated here. The argument to the effect that governments that were either too small or too large would tend to provide a less than Pareto-optimal supply of public goods might be countervailed, or even more than offset, by other considerations, in any particular case. For example, as James Buchanan has shown, a progressive tax structure such that the median voter in a country pays less than the average rate of tax could lead to a more than optimal level of public expenditure. Countless other considerations might also be relevant in any given case.

Nonetheless, it seems highly likely that a major discrepancy between the boundaries of a collective good, and those of the jurisdiction that provides it, leads to problems of the kind described in this paper. This, in turn, argues that we can be reasonably certain that a broad array of governmental institutions is a necessary condition of the Pareto-optimal provision of collective goods, and that neither the extreme centralist nor the extreme decentralist position makes sense. It is also likely that many of the thousands of governments that have been set up in this country to deal with particular problems have in fact been needed, notwithstanding the complaints about the duplication and complexity in our system of government. The existing network of governments has many serious faults, but is nonetheless probably better than most of the diverse arrangements that have been proposed to replace it.

27-877—69—vol. 1——23

SYSTEMATIC ANALYSIS AND GRANTS-IN-AID IN A FEDERAL SYSTEM

BY SELMA MUSHKIN and JOHN COTTON

Selma J. Mushkin is Program Director, State-Local Studies, of The Urban Institute; John F. Cotton is a Consultant to The Urban Institute.

They here examine the effect of public expenditure analysis on federal grant-in-aid programs. They assess the impact on program evaluation "when emphasis is shifted, as it is in the implementation of PPB systems, from program inputs . . . . to program outputs." In their discussion, the extent to which this shift in analytical emphasis may enable the objectives and standards of Federal grant-in-aid programs to be designed more effectively is appraised. They emphasize the need for PPB systems in State and community within a Federal system which would "(1) ensure that funds are used effectively to deal with identified national problems, (2) allow the states to tailor methods of dealing with a problem to their own situation, and (3) avoid the need for tight federal control of program expenditure."

The authors point to the manner in which the PPB frame-work can be employed to aid both the federal government in developing effective incentives for local action, and the local governments in constructing and implementing efficient programs. They also discuss the impact which PPB analysis is already having on federal grant programming, the appearance in federal grants of requirements for program evaluation, systems analysis and cost-effectiveness measurement, the reinterpretation of federal planning assistance in the vocabulary of PPB systems, and the encouragement of experimental projects to develop the methodology and data necessary for program evaluation.

Introduction

The development of integrated systems for PPB implementation and the execution of the core of such systems—analysis of relative costs and effectiveness—have been altering traditional approaches toward intergovernmental fiscal relations. In the present paper, we ask: What changes may be expected in the patterns and processes of intergovernmental relations as national, State, and local governments begin or continue to implement PPB systems? We first consider the conceptual changes that will affect both grant-in-aid design and intergovernmental program policies as a process is set in motion of (1) defining public products in terms of objectives, (2) searching for and specifying program alternatives to be subjected to analysis of costs and effectiveness, and (3) quantifying relative costs of various ways of achieving program purposes. Finally, (4) we review briefly the changes that can even now be observed in the emphasis of grant programs, their requirements or conditions, and the types of programs funded.

The principal purpose of a PPB system is to improve program policy decisions and concomitant budgetary allocations. Implementing the system calls for establishing a routine for program planning and for budgeting that brings relevant benefit-cost type information, organized and documented, to bear on public decisions. The routine is procedural, starting with emphasis on identification of program objec-

tives—i.e., defining the results that are expected—and looking toward a structured grouping of public services and activities into a hierarchy of categories, subcategories, and elements that can provide a framework for considering and assessing relative program results and resource use. The routine calls further for (1) the preparation of analytical studies that, for a range of feasible alternatives, will compare costs, on the one hand, and results or output, on the other, and (2) the development of program and financial plans in terms of outputs, costs, and financial requirements over a period ahead which includes the budget year and several years forward.

Although the documents and procedures of PPB constitute a routine, they cannot be applied in a "routine" way if the purpose of improving program and budget decisions is to be achieved. They must be seen as a built-in stimulus for a specific questioning of specific programs—why? for whom? with what effect? and with what full cost implications? Most importantly, a PPB system is a procedure postulated on the imaginative asking: What might alternatively be done? It is a stimulus for gathering of facts that can shed light on the questions posed, and for assembling such facts in a way that will illuminate and quantify the relevant considerations. The routine obviously cannot be applied in the same way in all governments. A jurisdiction's ecology, power structure, governmental organization and size, and its PPB staff's capacity all have an important bearing on the system's results, and how and where the questioning of purpose, the design of measuring rods, and the formulating and testing of options take place.

The processes of reassessment involved in PPB implementation are clearly relevant to intergovernmental programs: why cooperative programs?, toward what end?, for which groups in the population?, at what cost?, with what progress in meeting the defined purposes?, what alternatives are available? A comparing process is not new in intergovernmental policy; comparisons of various levels of government expenditures and of revenues for all public purposes or for specific public programs have been customary guides to policy formulations. Disparities in education have been measured, for example, by school expenditures per child in one state compared with that of the national average, or with the average for a region. Or disparities have been shown by comparing school expenditures per child in suburb and core city. Comparisons also have been made of staffing patterns across governments—for example, pupils per teacher, welfare cases per social worker, or proportion of public health districts with full-time public health officers. These comparisons have quantified the disparities that have called for intervention by the National Government, or by the State.

Focusing on the Program "Outputs"

When emphasis is shifted, as it is in the implementation of PPB systems, from program inputs (e.g., school expenditures per child), or program process variables (e.g., pupil-teacher ratios) to program outputs and the products produced for the citizen, the measuring rod applied in quantification of intergovernmental disparities changes [1]. The functions toward which Federal aid directs its support may con-

tinue to be the treatment of liquid wastes, the provision of improved nutrition for children in school, or preventive public health services; however, the definition of purposes, and the yardsticks for quantifying progress toward satisfying the defined purposes, are changed. Program standards come to be viewed in public product terms or as outputs that represent the quantifier of the defined objectives. In place of dollars per pupil for education, the standard may become learning achievements; in place of public health staffing or expenditures, measures of "good health," "positive health," "physical and emotional well being" may be sought. Measures of "satisfactory employment and earnings" could take the place of dollars per capita spent for vocational rehabilitation or vocational education.

Prior preoccupation with inputs into the process of producing public services or with tax dollars spent cannot be turned about without considerable study and experimentation. The public product or output must be defined, and its dimensions and characteristics described; methods have to be designed for measuring those dimensions and characteristics; ways of collecting information on the measurements need to be devised and tested; and the data must be gathered. Additionally, if a concept of a nationwide standard of output underlies the grant-in-aid concept, the measurements have to be uniformly designed and uniformly applied throughout the Nation.

Types of output measurements—The definition of public outputs, the methodology for measuring these outputs, the strategy for carrying out and testing of data collection are for most public products a must in an initial exploratory phase. A beginning, however, is being made in accord with the changing emphasis from inputs to outputs.

In these initial efforts at assessing output, measurements of varying depths and specificity are being discussed, ranging from fairly simple indicators of output to more complex aggregative indexes. Relatively simple indicators of public program outputs have been formulated that are akin to production measures in the private sector. These output indicators include such quantity measures as, for example, numbers of persons served, or average daily number of patients treated, or number of public housing units provided, and also such quality indexes as number of patients cured, nutritional levels of assisted families, decline in rates of measles, polio, and reductions in length or period of unemployment. (Tables 1 and 2 reproduce examples of such output indicators shown for selected program elements in illustrative materials prepared for State and local officials) [2]:

TABLE 1.—*Illustrative volume indicators*

Selected program elements	*Volume of services*
Natural resource programs:	
Forestry	Acres of forest lands maintained; acres planted.
Agricultural extension services	Number of farmers assisted.
Conservation	Number of acres with conservation practices.
Reclamation	Irrigated acreage.

TABLE 1.—*Illustrative volume indicators*—Continued

Selected program elements	*Volume of services*
City environmental programs:	
Air pollution control	Number of violations corrected.
Water supply system	Gallons of water consumed.
Solid waste disposal	Tons of trash removed.
Rodent control	Reduction in rat population.
Mosquito control	Area sprayed and irrigated.
Waste treatment works	Gallons of water treated.
Public housing	Number of housing units.
Urban residential renewal	Number of housing units renovated; acreage renewed.
Beautification	Trees planted; flowerbeds maintained.
Recreational programs:	
Parks	Number of persons using parks at peak time of day; number of parks; acreage maintained.
Museums	Number of persons attending.
Libraries	Number of library users; number of volumes on loan.
Zoos	Number of persons attending.
Recreation centers	Number of persons using or attending.
Performing arts centers	Number of persons attending.
Transportation programs:	
Airports	Number of passengers embarking; number of aircraft takeoffs.
Harbors and ship terminals	Dock area; number of berths.
Highways	Miles of paved highways maintained.
Mass transit facilities	Number of passengers per day.
Parking facilities	Number of motor vehicles parked.
Ship terminal warehouses	Tons of freight stored.
Public safety programs:	
Parole activities	Number of persons paroled.
Police surveillance	Number of surveillance hours.
Traffic controls	Number of intersections controlled.
Fire station activities	Number of alarms answered.
Courts	Number of cases cleared.
Jails and other detention institutions	Number of prisoners released; number housed.
Street lighting	Number of street lights maintained; kilowatt hours.
Child care and education programs:	
Child care centers	Number of children in centers.
Day care centers	Do.
Preprimary education	Number of children enrolled.
Regular day programs:	
Elementary schools	Number of students enrolled.
Secondary schools	Do.
Vocational schools	Do.
Higher education (student teaching).	Number of students enrolled; number of degrees granted, total, or by field of specialization.
Manpower training	Number of persons enrolled; number of persons employed at close of training.
Afterschool hour programs	Number of persons enrolled.
Summer school programs	Do.
Exceptional children programs	Number of children served.
School lunch programs	Number of meals served.
School health programs	Number of examinations; number of children screened, by types of screening.

TABLE 1.—*Illustrative volume indicators*—Continued

Selected program elements	*Volume of services*
Job opportunity programs:	
Neighborhood youth programs	Number of persons employed.
Job counseling	Number of persons assisted.
Employment services	Number of persons placed; number of persons interviewed.
Sheltered workshops	Number of persons employed.
Vocational rehabilitation	Number of persons receiving services; number of persons reemployed.
Health programs:	
General hospitals (inpatient care)	Average daily inhospital patients; number of patients treated in outpatient departments.
Mental hospitals (inpatient care)	Average daily patient loads.
Institutions for the mentally retarded.	Do.
Mental health clinics	Number of patient-hours of care provided.
Alcoholic clinics	Number of patients treated.
Family planning clinics	Number of women counseled.
Disease-screening clinics	Number of persons screened.
Immunization programs	Number of persons receiving services.
Referral services	Number of families served.
Home health services	Number of persons served; number of hours of care provided.
Nursing homes	Average daily patient loads; number of institutions inspected; number of institutions participating in staff training sessions.
Welfare programs:	
Family counseling services	Number of families assisted; average daily caseload.
Welfare payments	Number of persons and families receiving assistance, total and by type.
Food distributions	Number of families receiving food surpluses.
Foster home care	Number of families assisted; number of children or aged receiving care.
Halfway houses	Number of persons assisted; number of persons restored to community.

TABLE 2.—*Illustrative quality indicators*

Selected program elements	*Quality of service*
Natural resource programs:	
Forestry	Standing timber growth, by type of timber.
Agricultural extension services	Farm output per acre, by crop.
Conservation	Topsoil depth, width, and quality; water levels.
Reclamation	Farm output per acre in reclamation areas.
City environmental programs:	
Air pollution control	Current pollutant levels (carbon monoxide, sulfur dioxide, hydrocarbons, etc.).
Water supply system	Water rationing days; water quality indexes (dissolved oxygen, temperature, hardness, chloride, etc.).
Solid waste disposal	Periods elapsing between collections; proportion of wastes used as landfills.
Rodent control	Rat bite cases reported.
Mosquito control	Area cleared.
Waste treatment works	Water quality levels attained. (See water supply system.)
Public housing	Physical condition of public housing (working condition of plumbing, lighting, heating, or air conditioning and other equipment; noise).
Urban residential renewal	Number of families rehoused in satisfactory housing.
Beautification	Measures of physical appearance of selected sample areas.

TABLE 2.—*Illustrative quality indicators*—Continued

Selected program elements	*Volume of services*
Recreational programs:	
Parks	Lawn, trees, planted areas in satisfactory condition.
Museums	Number of exhibits or paintings; scope of exhibit coverage; special shows.
Libraries	Number of volumes requested not available; waiting lists of borrowers.
Zoos	Number of species represented.
Recreation centers	Range of activities available for different age groups.
Performing arts centers	Number of performances of companies with national reputations.
Transportation programs:	
Airports	Delay time; ground transport time.
Harbors and ship terminals	Size of vessels accommodated; tons of freight handled.
Highways	Traffic delays; traffic accidents; vehicle capacity per hour.
Mass transit facilities	Extent of crowding at peak hours; frequency of service during offpeak hours.
Parking facilities	Number of parked vehicle violations; number of parked vehicles towed from city streets.
Ship terminal warehouses	Cubic area, by type.
Public safety programs:	
Parole activities	Number of persons restored to community; number of repeaters among parolees.
Police surveillance	Age adjusted crime rates; number of offenders, by number of prior arrests.
Traffic controls	Number of motor car accidents; number of traffic fatalities and injuries.
Fire station activities	Time elapsing between calls and firefighting; fire insurance premium rates.
Courts	Length of court dockets; number of decisions sustained on appeal.
Jails and other detention institutions.	Number of repeaters; numbers rehabilitated and employed.
Street lighting	Illumination indexes (kilowatts).
Child care and education programs:	
Child care centers	Hours of education, health, and counseling services provided.
Day care centers	Do.
Preprimary education	Number of children by reading-readiness scores.
Regular day programs:	
Elementary	Number of children with achievement scores at or above grade level.
Secondary	Number of graduates; number of school leavers; number admitted to colleges.
Vocational schools	Number of persons receiving training who are employed.
Higher education (student teaching).	Number of college graduates admitted to graduate or professional school; attrition rates; number of graduate students receiving fellowships.
Manpower training	Number of persons receiving training who are employed.
Afterschool hour programs	Changes in school achievement scores.
Summer school programs	Do.
Exceptional children programs	Number of children, by achievement or adjustment score level.
School lunch programs	Number of children purchasing school lunches; nutritional status of children.
School health programs	Number of children with correctable deficiencies who have received treatment.

TABLE 1.—*Illustrative volume indicators*—Continued

Selective program elements	*Volume of services*
Job opportunity programs:	
Neighborhood youth programs	Number of persons employed by skill levels.
Job counseling	Number of persons employed by skill levels; length of unemployment period.
Employment services	Length of unemployment period.
Sheltered workshops	Number of persons reemployed in community.
Vocational rehabilitation	Earning levels of rehabilitated persons.
Health programs:	
General hospitals (inpatient care)	Number of appendectomy, hysterectomy cases reviewed by medical boards; waiting lists; special care units.
Mental hospitals (inpatient care)	Number of therapy hours of care provided; length of stay.
Institutions for the mentally retarded.	Number of persons functioning in community; discharges.
Mental health clinics	Number of persons functioning in community.
Alcoholic clinics	Number of patients cured; number of nonrepeaters.
Family planning clinics	Number of women planning families.
Disease screening clinics	Number of positive cases receiving treatment.
Immunization programs	Decline in rates of smallpox, tetanus, whooping cough, measles, polio.
Referral services	Number of persons receiving appropriate types of care or services.
Home health services	Range of services provided; hospitalizations avoided.
Nursing homes	Hours of professional nursing time provided; accessibility of physician care.
Welfare programs:	
Family counseling services	Changes in separation and divorce rates.
Welfare payments	Number of families moving across the poverty line; number at defined budget levels.
Food distributions	Nutritional levels of assisted families.
Foster home care	Number of cases of institutionalization avoided.
Halfway houses	Number of persons restored to community.

Another study, addressing the development of "evaluation criteria" for State and local public services, explores output measurements in terms of underlying purposes of governmental production of consumer goods and services [3]. In this discussion, as in the case of output indicators mentioned above, multiple measures of public products are presented. For example, the ten criteria in regard to law enforcement include such measures as annual changes in the number of offenses for each class of crime, crime rates per 1,000 inhabitants, and the annual value of property lost through crime.

Still another effort is directed at the formulation of composite "social indicators" that can summarize the well-being of individuals and families, their living and home conditions, their health status, the quality of their environment, their social mobility, their personal freedom and security.* A social indicator, as defined in the volume *Toward a Social Report*, recently released by the Secretary of the Department of Health, Education, and Welfare, is a statistic that facilitates concise, comprehensive, and balanced judgments about the condition of

*Further discussion of this issue is found in the paper by Sawhill in this volume.

major aspects of a society and "is subject to the interpretation that, if it changes in the 'right direction,' while other things remain equal, things have gotten better, or people are 'better off.' " [4] Among the indicators developed are the expectancy of healthy life (or life expectancy free of bed-disability and institutionalization) and extent of criminality (crimes weighted by a measure of harm suffered by the victim either through property losses or personal injury). Social indicators are summary measurements of social health or illness, but they do not directly relate public programs to outputs.

Still another summary measure calls for estimates of "imputed economic value"; imputations of changes in the quality of life reflected in the adjustment of economic accounts. Attempts have been made to discover how income measures might convey, in monetary terms, aspects of levels of living and of individual or family well-being that are not adequately represented in current economic concepts of gross national product or monetary income. While imputations are made at present in such accounts for rental value of owner-occupied dwellings and for food produced for own consumption, other factors that influence the enjoyment of income are not counted. Air pollution, for example, may lower satisfaction derived from income; thus it might appear reasonable that incomes should be adjusted for differences in volume and type of pollutants in the atmosphere.

A monetary measure of impact of public programs that warrants special attention is an estimate of the total effect of those programs on future earnings and income levels in a community. To provide a consistent method of counting, these future earnings are discounted to show present value of the future earning stream. The comparative effects of various programs can be aligned in the same way.

Range of outputs—We need to be realistic in asserting how far we can hope to proceed in this change from inputs to outputs. For it is evident that the prerequisite for a "complete" changeover in the emphasis of measuring programs is the establishment of adequate estimates of causal relations between inputs and outputs. In this regard we are severely limited in most programs, particularly those in which the research base has been weak.

As an example, if we were to attempt to measure the output of a "law enforcement" program in a given city, we would want to know such things as the effect of increasing the number of police officers or of establishing more police patrol car units on the criteria used to measure the benefits of law enforcement. However, it is known that crime rates and value of property loss depend in an unpredictable (to us) way on many other factors. Consequently, we would have difficulty in establishing a useful relation between inputs and "end" outputs. It does appear feasible to estimate the impact of more police officers on a measure of the time elapsing between the occurrence of a crime and the arrival of a police officer at the scene. Further research also may permit us to relate average "reaction time" to the likelihood of successful apprehension of the offender.

It would be reasonably accurate to say that the change in emphasis from inputs to outputs in PPBS consists of two things: (1) a clear identification of objectives and the criteria by which one would judge the extent of progress toward those objectives, and (2) more mean-

ingful estimates of the relation between program inputs and "end' outputs, so that the policies pursued will more closely approximate the end-results that are sought.

A measure of reaction time is better than a measure of the number of police officers, and a measure of likelihood of successful apprehension is better than a measure of reaction time. Ultimately as we learn more, we might be able to relate these measures to crime rates and the other criteria that more directly reflect attainment of society's objective of reducing the incidence and effects of crime.

Program effectiveness as a determinant of grant awards—As the work on output measurement progresses, the meaningful character of the information—for instance, on increased learning achievement, higher productivity, reduced losses of property due to crime, lowered travel time to work, and so forth—will suggest reorientation of intergovernmental programs. The emphasis will shift toward grants-in-aid that are conditioned on the outputs or results obtained rather than dollars spent or personnel employed. To illustrate, and to make a comparison between the newer output measurements and the customary input approach to grant formulation, we select a public program area on which sizeable expenditures are made—correctional institutions. We assume that there is federal government concern about improving such institutions as one method of reducing the crime-breeding conditions sometimes attributed to them.

Past procedures would possibly have called for (1) identification of some Federal grant amount that might reasonably be expected to contribute to the improvement of corrections work, and (possibly) for (2) a specification of the kinds of staffing and activities that would serve that purpose—e.g., psychologists' services, occupational training. As more work is done on outputs and output measurement, the *purposes* of correctional institutions come under scrutiny. When we ask what the purposes are, the answers might include: (1) punishment of the criminal as a matter of justice; (2) crime prevention, encompassing punishment of the criminal as a deterrent to others, rehabilitative efforts when possible to restore criminals to legitimate and socially useful activity, and separation of the criminally ill (i.e., the mentally ill bent on their own and society's destruction) from the rest of the community. If the National Government's purpose in its aid is essentially crime prevention, the emphasis would be placed on item No. 2 in this listing.

But how is progress in satisfying this purpose measured? We would need to examine the problem of measuring progress in (a) crime deterrence for those outside of correctional institutions, and in (b) restoration of those released from correctional institutions. Direct measurements of "the deterrence potential" are difficult to design and apply. Crime rates adjusted to yield comparability before and after a change in correctional policies are proxy yardsticks, but they are far from precise. "Restorations" for specific persons released from institutions are not as complex, even though the causal relation between individual behavior after release from prison and activities in the prison may be difficult to establish. We might select two indicators to assess program output for the "restorations" component, namely rates of recidivism, and proportion of persons released from prison who are employed in a gainful occupation 6 months after release (in-

cluding, for women, work as a housewife). These indicators of output become the measures of achievement of the program.

Rates for each of the outputs can be specified as a basic national "norm," toward which a grant-in-aid system could be geared. Or, rates of their improvement can be specified as yardsticks of progress, permitting a differentiation among States and communities in the level of program effectiveness. The amount of Federal support required to achieve the national "norm" or the rates of improvement still needs to be determined experimentally by a "laboratory-type" test in terms of inputs of resources and their dollar costs, or on a trial-and-error basis that seeks to evaluate the gains made relative to the increased inputs in different states or communities.

As an illustration of the process of establishing a national "norm," we can draw upon a program area that has been the subject of analysis within the Federal Government—Maternal and Child Health. Here, as in the case of law enforcement cited earlier, there is a sequence of potential measures for the program. At the one end of the sequence we have program inputs; that is, average number of dollars spent in family planning per target group family. Proceeding onward we might measure the magnitude of activities supported, that is, number of family planning centers established, or number of women provided with services. Or measures of direct program effect might be used, such as changes in birth rate among the target population. Finally, one might hope to be able to identify the expected decrease in infant mortality and birth defect rates accompanying the changes in birth rate.*

Thus, in the maternal and child health example, we can see an array of measurable factors which might be considered as bases for setting standards. But how does one go about establishing a national norm for any one time period?

Let us examine an idealized procedure that would be consistent with the philosophy of a PPB system. It is known that a problem exists in regard to the subject of maternal and child health—an undesirable disparity between its status in our nation and what seems to be achievable from at least two comparative measures: (1) the differing infant mortality and birth defect rates among various socioeconomic groups in the United States and (2) the difference between the average rates in the United States and those of other advanced nations.

The program analysis staff of the U.S. Department of Health, Education, and Welfare might be (and, in fact, was) charged with the responsibility of analyzing Maternal and Child Health programs, to seek information on the most effective ways to attack the problem [5].** The analysis would estimate that one might expect to accomplish in reduction of infant mortality rates and birth defect rates as a function of total cost for a broad array of program approaches. Ideally, the analysis would show the best combination of program approaches for any given commitment of total resources and would estimate the impacts of the approaches. If the research base is adequate, the im-

*Further discussion of this issue is found in the papers by Wholey, in this volume and Grosse in vol. 3 of this collection.

**Further discussion of this issue is found in the paper by Wholey in this volume.

pacts would be estimated in terms of the estimated reductions in infant mortality and birth defect rates that could be produced by different spending levels for varied packages of programs. If knowledge of the necessary relations is not available, the "highest level" measures available would be estimated.

On the basis of this *and* other relevant information and judgment, the responsible agency officials would decide what would be the most appropriate program package to pursue in the present time period. In principle, the decision at this point is in terms not of the level of federal outlays, but of total resource commitments required. Thus, a specific total resource commitment to a program package made up of (1) prenatal care and improved nutrition for expectant mothers and (2) family planning services might be judged appropriate.

The Federal Government would then have to try translating the hypothetically achievable into fact. Ideally, the effectiveness criteria (e.g., mortality and defect rates) would become the standards for the grant-in-aid program. The control points would shift from program inputs to outputs with standards set for a desired national output "norm." The year-to-year funding for a given State would then be set on the basis of its performance in relation to the output standard without Federal controls on how the State put together resources to carry out the program purpose.

Two hard questions have to be considered in this realignment on Federal controls: (1) the possible divergence of the program package that is developed by the States from that estimated by the National Government when it decided upon its grant-in-aid commitment, and (2) the difficulties of administering controls based on output standards when there is a large uncertainty about how to achieve given program results.

The judgment that the Nation as a whole should make a specific commitment of resources to maternal and child health was made in part on the basis of how much could be accomplished by that commitment. If the program is defined and controlled broadly as maternal and child health impacts, there is no assurance that the resources will be put together at the State level in a way that corresponds to the analysis carried out by the National Government.

Programs would have to be more tightly defined, for example, with a certain fraction of funds constrained to be used for family planning centers, to assure that the program desired will be implemented in fact. This type of constraint points to continued categorical approaches in program design. If the National Government wishes to ensure the outcome it has projected based on the national assessment of relative costs and effectiveness, it must then control the detailed way that the funds are spent by the system.

There is no easy way out of this dilemma. Strengthening of program analysis and planning at the State level should help assure that funds which are not tightly controlled will be used well. But this use will not solve the problem of divergence from a national assessment and possible need for continuing narrow definitions of program areas.

We assume that program analysis has developed an estimate of a causal chain: from dollar inputs to resource inputs such as facilities and personnel, to services provided to people through these facilities, to the effect of these services on birth rate and family size, and finally

to the effect of birth rates and family size on mortality and defect rates. As one proceeds along this chain, the uncertainty of the relations and the decrease in control between the links increases. In other words, it is much easier to specify that a given amount of dollars will in fact produce a certain set of facilities and personnel than to specify that those facilities and personnel will service a given group of people in a certain way, and that such services will produce a standard of maternal and child health.

If one wishes to avoid a set of rigid, narrow categories, on the one hand, but does not, on the other hand, want to sacrifice assurance that funds are reasonably well spent, it will be necessary to change the control points to standards for a desired national "norm" of outputs. Given the uncertainties about how to achieve a given output and also the many program and factor interactions, this kind of standard setting could not be subjected to rigid controls for some time. Certainly a cutting off or reduction of funds for failure to meet the standards under these circumstances would be a harsh penalty. Conditions for grant receipt would have to be limited to a requirement for evaluation of programs, out of which should come, over time, the knowledge required to establish controls in accord with output standards.

The changing approach to grants-in-aid under a uniform standard is illustrated by the legal action now being taken by Detroit under Michigan's Constitutional guarantee of public education for all the children in the State. This provision is being interpreted as an assurance of "equality of educational opportunity" with "equality" being reviewed, not solely in terms of expenditures, or amounts of state grants, but also in terms of school "achievement" levels.

A more ambitious and more intriguing approach to grant design is one that would abandon the national "norm" and in its place rely to a much greater degree on the individual State's assessment of its problem and the proper way to deal with it. In this approach, the Congress might take the position that maternal and child health is a significant problem and establish a broad funding authority to deal with it. However, at this point, the States, on the basis of their own analysis of their problem in maternal and child health, would negotiate with the National Government the magnitude of the program, within specified limits, that is appropriate to that State. The Federal share of the costs for carrying out the negotiated program would then be determined yearly on the basis of the State's performance in relation to the mutually agreed upon standards of performance. An example of the type of agreement that might be reached could be: if the State achieves the output level it projected or is within some specified limit of that output level, the Federal share would be an amount x; if the State exceeds the projected output level, the Federal share would be increased proportionately; and if the state falls too far short of the level, the Federal share would be decreased proportionately (within limits). A conditioning of the grant on the results of an evaluation would be essential to this approach [6].

While this approach to the grants-in-aid complex would require a great deal of effort to establish, it might present a feasible way to: (1) ensure that funds are used effectively to deal with identified national problems, (2) allow the States to tailor methods of dealing with a problem to their own situation, and (3) avoid the need for tight Federal control of program expenditures.

Program Incentives Through Price Reduction*

Quite separate from the potential alteration of criteria for grant awards discussed above, the development of PPB systems could have a large effect on the design of the matching provisions of Federal grants-in-aid. The development of PPB systems could alter the degree to which States and localities respond in a rational manner to the Federal incentives (hopefully rationally designed) that are incorporated in grant programs. Since the Federal Government cannot generally legislate standards for the States (as States can for local governments) the design of incentives necessarily must take account of (1) differences in objectives between Nation and State or locality and (2) intergovernmental differences in priorities among objectives.

Each governmental unit has its view of public purposes and its own emphasis (there is no "natural" sequence or order of ranking). Preferences of citizens in different communities may vary, reflecting, among other things, diverse sectional history, geography, economics, and population characteristics. With the introduction of PPB systems, the similarities and differences in objectives and priorities among governments can be more clearly set forth. Assume, for example, that the National Government is seeking to encourage promotion of major child health services. Then assume that in one State a large majority of the population is opposed on religious principle to such medical care. What types of grant provisions would gain the national purpose in that State? The conflict in basic objectives may require the enforcing of large penalities or grants well in excess of 100 percent of program cost that go beyond the concept of a Federal grant offering of a share of costs on a "take-it-or-leave-it" basis.

For most federally aided programs, differences in basic objectives are not at issue, but differences in the priorities assigned are. A State Governor's judgments about the relative utility of a public product may vary from those of a city mayor or of the President of the Nation. Intergovernmental comparisons of budget allocations for defined objectives can become a guide to the National Government on the priorities assigned by the various State and local jurisdictions. With this type of information, incentives to encourage a reordering of priorities and matching ratios tailored to achieve State and city action in accord with national purposes and priorities could be designed less blindly.

The effect of price reduction—As suggested earlier, the Federal grant-in-aid is an incentive to States and local governments to achieve a nationally defined purpose. The grant-in-aid works as a stimulus by altering the budgetary constraints and by changing the relative prices of activities, making it less costly to undertake the aided program than a non-aided one.

The following illustration is concerned with the question of using a grant-in-aid to promote achievement of a specified nationwide standard within the budget year. We expect the State to increase the level of the aided program if the "price is right." The size of the increase will depend on the nature of the decision process in the government and on underlying program information. For example, if a State or

*Further discussion of this issue is found in the papers by Schultze and Olson in this volume.

locality routinely makes use of cost-effectiveness analysis in its program decisions, the prior program level of performance would have been based in part on criteria comparing the payoff from the program to its costs. The response to the grant offering would then depend on how the budget constraints are altered as well as how the change in program costs affects the decision criteria. Uncertainty will be substantial and the task of estimating State or local response to grants will not be easy in any case.

Figure 1 illustrates one program of national concern (program X) and a second program which represents the aggregate of all other State programs (program Y). We assume that, in the absence of Federal support, a State would have a budget constraint AB, and would adopt a program combination X_0Y_0 after comparing relative costs and effectiveness. We assume further that a national program standard has been established for program X at X_S greater than X_0.

The matching provisions or Federal grant incentive need to be set so as to gain the national program standard X_S for the State.

By providing a carrot of the right magnitude, the Federal Government hopes to lead the states or localities to the desired program resource allocation. The appropriate criteria for the grant design are: (1) the establishment of matching conditions that just induce the State to move to program level X_S, and (2) the allotment to the state for that program of an amount that, at the established matching ratio, would provide the required input dollars to achieve the outputs designed.

Two different grant options are shown in Figure 1. Both are closed-end grants at the defined standard program level X_S. As illustrated in the figure, the grant producing an apparent budget constraint ACD provides insufficient incentives for the State to achieve the minimum standard level, given its comparisons of relative gains from programs X and Y and the altered relative prices. On the other hand, if the grant is sufficient to raise the budget constraint to AEF, the State would be induced to reach the standard level for program X.

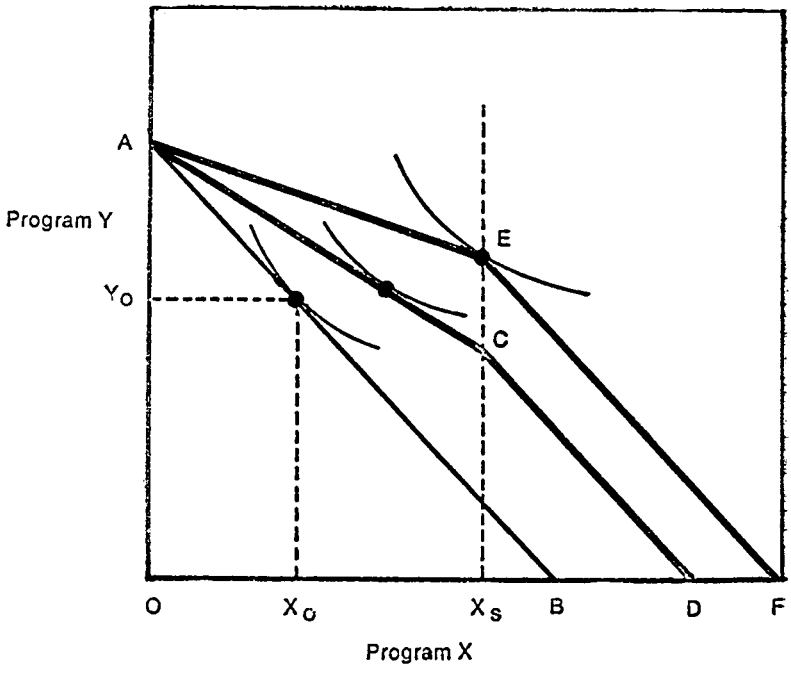
The matching share required would depend upon the relative own revenue costs to the State or localities compared to the expected gains. External flows of revenue and benefits through exported taxes, on the one hand, and spillover of benefits, on the other, would be taken into account in the before-grant and after-grant comparisons.[1] The federal grant can be designed to obtain a nationwide standard of performance, or to encourage progress toward the achievement of a result that requires, overall, a larger commitment of resources than any of the State or local governments are immediately prepared to make. More recent discussions on new directions in Federal aids suggest use of the Federal taxing system not as a program incentive but as a method for funding State and local services without program strings attached [7].

We have indicated that strong financial inducements are required to gain State action when the basic purpose of a State are at variance with those of the National Government. Far smaller financial incentives are required when the objectives of both State and Nation are similar but their priorities among programs differ.

---

[1] The options for setting the Federal share in terms of correcting distortions produced by the benefit spillovers are presented by the authors in *Functional Federalism*, Washington, D.C., The George Washington University—State-Local Finances Project, November 1968, pp. 52–54. See also the paper by Knetsch in vol. 3 of this collection.

FIGURE 1



Moreover, in designing grants, the general commitment of resources to obtain desired results within a specified time frame has to be determined. Funding arrangements would be different for grants that are designed to obtain some defined national standard within a specified period than for grants that seek to gain progress at a defined rate or at negotiated rates, toward a long-term goal.

Current matching ratios in the perspective of PPB—Matching ratios have come to be varied as if they were precision instruments for encouraging organizational changes in government or the provision of particular services considered of high national priority. In fact, the precision does not exist without the type of analytical framework assumed in the implementation of a PPB system. Current matching ratios are widely disparate, depending more on the period of adoption of the grant than upon some rational concepts of appropriate incentives to achieve national standards. Moreover, incremental changes in matching have increasingly been used as if these ratios were finely tuned to public decision, but the conditions have not existed for such precision. Many illustrations may be given: Federal shares for college construction are 33⅓ percent of total costs; however, if a public community college facility is being constructed, they are raised to 40 per-

cent. Administrative costs for welfare are shared on a 50–50 basis, but for administrative services that offer the possibility of preventing poverty, a 75-percent matching is provided. For certain other programs, the Federal share for certain community facilities is increased if the facilities are provided on a metropolitanwide basis.

Responses of the States and localities to these varied matching ratios are, at present, poor indicators of the adequacy of the incentives provided for expenditure decisions. A strengthening of program analysis at the national and State levels would permit the responses recorded from states and localities to grant-in-aid offerings to become a better measure of the appropriateness of the ratios established. Moreover, the National Government would have less need for tightly controlling resource inputs when grants are structured and assessed in terms of the results achieved.[2]

Program Options in a Federal System

The same (or a similar) national objective can be pursued in a number of different ways by a range of both public and private activities. An important phase of the analytical process of PPB implementation is, thus, the design of program alternatives. The health of the child, for example, can be improved by medical services, by nutrition care, or by health education. The choices in methods to achieve defined objectives depend upon the options that are created. This emphasis on alternatives or options alters the decision making from that of "confirming" or "denying" a particular method, to that of choosing on the basis of an assessment of the trade-offs involved in allocating scarce resources. A large commitment to build analytical capacity within State and local governments, to appraise results of public programs, and to design new methods where old ones do not yield results could alter the operation of our Federal system in a major way. As we shall point out below, grants-in-aid can play a significant role in allowing many options to be considered.*

States by and large have become reactors to rather than creators of social invention and innovation. However, generalization about the fifty States in this respect, as in others, is likely to be only partially accurate. Many States have innovated at one time or another. Wisconsin introduced unemployment compensation before the nationwide system was on the drafting board; California introduced hospital insurance benefits long before Medicare; Rhode Island adopted the first temporary disability benefit. Equalization of grants-in-aid is a State invention adapted by the National Government for its own grant-in-aid procedures. The negative income tax proposal is an adaptation and extension of experience in Indiana, Minnesota, Nebraska, and Wisconsin, with income tax credits for sales or property taxes paid by the aged or the poor. "Home health care" is an extension of hospital practices in Ohio and New York. Tax sharing, which is now receiving such widespread national attention, holds a State patent. Nevertheless, the volume of inventive or innovative activity on the State level is less than impressive.

[2] There may continue to be need for narrowly defined categories of Federal aid to overcome jurisdictional barriers, as indicated later.

*Further discussion of this issue is found in the paper by Olson in this volume.

In the private market place, day-to-day competitive forces encourage creative design of new products and new production methods. Large sums are spent by industries in research and development activities to generate new products and methods so that competitive positions may be maintained or improved. For governments, a certain amount of incentive for new idea generation has been provided by the political process; however, that process focuses on novel products that can become the "issue of the year" rather than on a continuing research and development effort. With the introduction of analytically based planning-programing-budgeting systems, a potential procedural basis is established for seeking out program options at all levels of government.

Classes of options—In understanding the possible contribution of program analysis to the generation of alternatives and new public products at all levels of government, various classes of program alternatives may be distinguished by their point of origin. One class follows from the logical procedure of the analysis itself; a second represents a program proposal in the public domain—i.e., an idea that has not been implemented in the particular government; and a third is the new social invention [8].

The "search for alternatives" necessarily starts by definition of components of the system that help to identify the various ways a problem may be attacked. In the example presented earlier of improving correctional institutions to prevent crime, the components that affect the output of correctional institutions include: the correctional institution itself (physical facilities, staffing, guard services, food, clothing, educational training, psychological services and other medical care, etc., for the prisoner); the courts; the police and probation system; the families of the prisoners; the community; and industry and its hiring practices. We also identified several measures of output to assess progress toward achieving the defined result. These include (a) the rates of recidivism, and (b) the proportion of those released from prison in satisfactory employment after a defined period. The programs and activities which might favorably affect those measurements of results include changes to (1) improve correctional institutions, (2) develop the socially acceptable attributes of prisoners, (3) make law enforcement fairer and speedier, (4) motivate families, community, and employers to re-absorb the prisoner into family and community living.

Some of the program options logically follow from identification of the components of the system. Still other options might be devised, building on proposals that have been advanced; for example, more extensive use of half-way houses and community-based correctional programs for adjustment to community living after release from prison. Additional options essentially require social invention or new ideas about how to improve the components of detention programs and correctional institutions, police, and probation systems—or how to develop new and more effective incentives to modify family, community, or employer behavior.

For each kind of option there is a range of program levels; that is, variations in doses (the more or less) of each activity. And there are possible variations in level for different persons within the general group for whom services are provided. An important range of options can result from questioning about the smallest "doses" that will yield the greatest result for each of the types of persons involved. Even

without a great degree of precision in measurement, some quantification of changes in effectiveness that are produced by adding small increments to costs, or by reducing costs by some increment, may provide new guides to policy decisions. Focusing on marginal gains and marginal costs accompanying changes in the level of a particular activity is then a possibility open to each governmental unit, although not all types of activities or programs are equally accessible to each unit.

Fragmentation of government and constraints on program options—In a centralized government, whose public responsibilites extend over the entire spectrum of the system, the options are open ended. The government is in a position to define the range of possibilities, to gather the facts required, to experiment, and to invent. And it could, on the basis of all the information on results and costs, select a combination of programs that in specified doses would change prisoner behavior, prisons, courts, police, parole, and incentives for families and employers to yield the lowest rate of recidivism, the highest percentage of employment—and that would lower crime rates.

In a decentralized system, some options are foreclosed to the National Government, or are shared with the States; frequently many local governments are involved, including overlapping local governments in a single geographic area.* In such cases, the selection of feasible options for program analysis takes on a somewhat different meaning. For example, a city government may have responsibility for police, and the State for parole; the county government may have responsibility for the jails, the State for correctional institutions and also for training programs or employment incentives. A particular local government's power to deal with a program problem may be constrained by State laws. Such constraints are not necessarily absolute but change would certainly involve time and persuasion and, for the immediate present, may be blocked.

To illustrate the effect of jurisdictional responsibilities on alternative program designs, we follow somewhat further the example cited earlier of rehabilitation of criminal offenders. For a number of possible ways of gaining the results desired in crime reduction, we identify the agency whose program would be affected and the government responsible for the agency's program (table 3). Resource allocation of State, city, county, and special districts would be variously influenced. Moreover, if within a governmental jurisdiction program options are restricted further to the particular mission of a single agency, program possibilities that might yield greater results may be foreclosed because the option falls outside the responsibility of that agency.

We have not selected the most complex instances to show the jurisdictional constraints in the search for alternatives. Federal authorities may be directly involved, for example, in cases involving drug addicts. "Agencies" may be independent commissions that have direct access to legislative bodies and are not responsible to the governor, mayor, or county executive.

Two aspects of the search for alternatives in a decentralized system require special emphasis. First, as illustrated above, options are constrained by the fragmentation of government. Second, these constraints may point to special federal encouragement by narrowly defined categorical grants to provide incentives for overcoming the barriers of such constraints.

*Further discussion of this issue is found in the paper by Achinstein in this volume.

TABLE 3.—REHABILITATION OF CRIMINAL OFFENDERS TO REDUCE RECIDIVISM AND INCREASE EMPLOYMENT IN SOCIALLY ACCEPTED EMPLOYMENT

Program type	Agency	Assumed governmen
Improving detention services and facilities	Jail	County.
	Welfare agency	Do.
Improving correctional institutions	Superintendent of prisons	State.
Staff	do	Do.
Physical facilities and security	do	Do.
Vocational training and equipment	do	Do.
Counseling	do	Do.
Therapy	Health department	City.
Cooperative arrangements with industry	Employment office	State or private.
Improving parole arrangements	Board of parole	State.
Counseling	Social services	Do.
Providing access to training	Education agency	Special district.
Referral to community agencies	Referral agency	City.
Surveillance	Board of parole	State.
Improving police deterrants of past offenders through more certain apprehension.	Police department	City.
Improving deterrants of past offenders through more certain and faster sentencing.	Court	County.
Programs for nonimprisonment	Courts	County.
	Probation department	State.
	Social services	City.
	Educational	Special district.
	Department of institutions	State.
Community acceptance and adjustment:		
Vocational placement program	Employment exchange	Do.
Employment followup	Social services, police	City.
Family aids	Various	State, city, county.
Program of selection among offenders: Analysis of characteristics of offenders in relation to rehabilitation methods.	Mental health agency	City.
	Health department	Do.
	Department of institutions	State.
Opening employment opportunities	do	Do.

Strengthening the Federal system—Over the longer run, procedures that call for options give momentum to a change in the way of thinking about intergovernmental programs. In the past, many State and local government programs have been responses to legislation enacted by the U.S. Congress making new offers of Federal aid. The initiative, increasingly, has been a National Government initiative. State, city, and county have begun, however, to implement PPB systems and to undertake analytical searches of better ways to get the results desired. The State and local governments should be able to initiate programs and develop innovative production methods. Some of the Federal grant offerings, on careful assessment, may be deemed contrary to community purposes, or insufficient.[3] New methods more in harmony with the political, demographic, economic, and social conditions in a locality may be formulated.

Perhaps no more fundamental impact of PPB systems on intergovernmental relations will be generated than through the processes calling for a search for alternatives and the testing of them. It has been a long-standing notion of federalism that the division of responsibility provides government laboratories for experimentation. With the introduction of PPB procedures, a new meaning is given to experimentation. Out of such experimentation may come laboratory findings and a body of experience varied to meet the special needs in particular communities and giving added support to greater flexibility in Federal grant-in-aid designs.

[3] See section above: Program Incentives Through Price Reduction.

Immediate Impacts of PPB on Intergovernmental Programs

The questionraising of PPB is having its immediate impact on Federal grant programing. A number of newer Federal grants require program evaluation, application of systems analysis, cost-effectiveness measurement either as a condition for Federal aid, or as a criterion for project selection. Federal planning assistance is being reinterpreted in the vocabulary of PPB systems, and extended to give support to PPB implementation in State and city. And experimental projects or research are being funded to yield data necessary to assess potential program payoff for resources spent. These three developments are sketched briefly below.

Program evaluation: selection criteria and grant requirements—Early in the national PPB effort, programs of the Office of Economic Opportunity were subjected to a study of the pay-off for amounts spent. The nationally administrated Job Corps program, as well as the intergovernmental Neighborhood Youth Corps, were compared to determine the cost and effect of each on employability of economically disadvantaged youth. "Head Start," as a cooperative intergovernmental preschool program, was launched but not without the strings of assessing the effort to show what was gained by way of educational achievement. Title I of the Elementary and Secondary Education Act of 1965 required project evaluation as part of the reporting conditions attached to the grant. The Demonstration Cities and Metropolitan Development Act of 1966 calls for evaluation of the model cities program. To assist cities, the act authorizes HUD to provide technical assistance directly or by contract to city demonstration agencies. The 1970 budget calls for over a tenfold increase in the range of specialized technical assistance, including contracts to assist both the cities and the Federal agency in evaluating the effectiveness of the city program. Grants enacted in 1968 make even plainer the effect of the processes of program consideration that characterize PPB systems on grant requirements and selection criteria. The Juvenile Delinquency Prevention and Control Act of 1968, for example, sets forth—as a criterion for project approval—an assessment of relative costs and effectiveness of the programs proposed for achieving rehabilitation of juvenile delinquents.

Planning assistance—As the processes of planning, programing, budgeting systems are being implemented, a new direction is being given to planning provisions in federally aided programs. Federal planning assistance is in a gradual, but nevertheless important, transition in which "planning" is increasingly coming to be defined in the terminology and content of PPB. Assessment of costs and effectiveness of alternative activities or programs is being encompassed within the context of planning. In some instances, new planning requirements are being adopted that call for cost-effectiveness analysis as a part of the planning process. Planning requirements of the Omnibus Crime Control and Safe Streets Act of 1968 are being defined in this way, for example. In other cases, existing planning grant support is being reinterpreted, including guidelines for comprehensive health planning and for educational planning. A pulling together of planning activities by a central unit in State or city is part of the new awareness about an integrated system of program planning. For example, Federal funds for comprehensive health planning may be made available to a central planning agency by the governor's decision.[4]

[4] But these funds continue to be earmarked for health planning purposes.

Central government-wide planning support through the 701 grants under the Housing and Urban Development Act is increasingly becoming available to the States for implementation of PPB. The Department of Housing and Urban Development is now providing encouragement to states to undertake studies for adapting and testing PPB concepts in the operation of State government. A 1969 set of guidelines for State PPB studies makes plain the thrust of strengthening "the governmental decisionmaking process by integrating the planning and budgeting functions and, thereby, including systematic program design and analysis within a multiyear time frame." The Model Cities program—Title I, Demonstration Cities and Metropolitan Development Act of 1968—is giving support to systematic analysis of programs by its guidelines for goals and objectives setting, the design of strategies for achieving those goals and objectives, and by its requirement of a multiyear financial and project forecast. A linking of analytical processes in model neighborhoods to city-wide PPB efforts is encouraged. The act specifically calls for "systematic analysis of the costs and benefits of alternative courses of action."*

The planning assistance requirements and the Federal guidelines set for States and localities are being altered from the early emphasis on compilation of data on resources, and estimates of need, to a new emphasis on formulation of objectives, development of alternatives, quantification of costs and effectiveness of alternative methods of satisfying those objectives.

The steps that have been taken to encourage PPB systems in State and locality fall short of a continuing Federal grant support for central staff work on programs analysis or central review of such analyses in State, city, and county. But proposals are being advanced for such Federal grant support as an underpinning for greater flexibility in Federal grants for substantive functional programs, and also through nonconditional aids.

Federal support of experimentation—Experimentation and research on a vastly enlarged scale is a major prerequisite to meaningful analysis. The process of implementation of PPB places a heavy demand on information about the operation of public programs and activities and the effects they produce. The question-asking proclivities of the analytical process are enormous. How are children motivated to learn? Are different groups of children motivated in different ways? What is the best way, for what types of children? Can we identify the characteristics of the children who are best motivated through the various methods? What are appropriate and efficient testing methods to determine those characteristics? What is the margin of error in applying such tests? How could these errors be reduced?

Or, to change the subject matter: What are the factors that contribute to physical and emotional well being? How are persons in different ages, economic conditions, with differing habits and cultural patterns encouraged to eat, sleep, drink, play, exercise, work, and seek health care so as to gain physical and emotional well being in the population? We could expand this type of question raising over the entire range of human and environmental problems of society.

*Further discussion of this issue is found in the paper by Ross in vol. 3 of this collection.

Many of the questions are familiar ones that have been raised before. But the concept of testing program results and of relating public resource commitments to the promise of results obtained, is different. And in an analytical framework in which programs are tied to budget decisions, the questions take on new urgency, and the findings have a direct application that was not available earlier.

There is a need to adopt a truly scientific approach to our quest for answers. Governments are learning, sometimes at considerable expense, that the necessary answers do not "pop out" of massive fact gathering exercises. We must identify what it is that we need to know and design our real life experiments to help us find the way.

Steps are being taken to get answers to the many unanswered questions. A new Federal encouragement is being given to research. Examples are the newly created Law Enforcement Institute in the Department of Justice, and the federally supported nonprofit Urban Institute. Experimental programs designed as *experiments* with appropriate controls are also being fostered. These experimental programs differ from the demonstrations supported through Federal grants heretofore. Most of the earlier demonstrations were intended to foster innovation, to encourage the trying out of ideas in a number of communities, and to make some inroads, in a prescribed way, on problem areas that the National Government was not prepared to launch on a nationwide basis because of costs involved. The demonstration programs were designed to gain more widespread use of program methods and had no clearly designed experimental content. Evaluation procedures were not built in, and provision was not made for dissemination of results of the demonstrations.

Demonstration grants to State and local governments, as well as universities, had been made, for example, for well over a decade before discussion took place on how to design an economic opportunity program. Despite the long period of demonstration support for a wide range of public services, there was not a ready body of material that could be used in developing the antipoverty program.

The purpose no longer is innovation; it is measuring the results of innovation. The notion of building into a series of demonstration projects planned variations in program and measuring effects of the modifications or variations came only as analytical procedures on a systematic basis were adopted. There is now not only a need to know, but a way to apply and to alter programs as a consequence of the information obtained.

The stimulus to research and experimentation may be expected to gather momentum with heightened emphasis on program design prior to large-scale commitment of resources. Some examples are already plain, such as in Head Start and Follow Through programs in some communities that are planned to yield experimental findings, and the new models for experimenting with alternate approaches for up-grading vocational education.

By the careful recording of these and other experiments, a deepened knowledge becomes available for analysis of expenditures in local communities. Concomitantly, networks of clearinghouses are being created to transfer information from research and experimental projects.

Some Next Steps

A new path is being taken in intergovernmental programs as the National Government, States, and localities undertake to implement integrated systems of planning, programing, and budgeting. Advance down this path would be greatly facilitated by specific legislative authority for a formula grant-in-aid offering especially designed to encourage and support staff efforts and staff training on PPB in the States, cities, and counties. The intergovernmental 5–5–5 demonstration, financed by the Ford Foundation, called on five States, five cities, and five counties to begin the process of PPB implementation. This project provides a body of experience on which other States and localities may draw in organizing and building the necessary staff competence.

Grant-in-aid methods and procedures for substantive programs cannot but be affected by more widespread adoption of processes for improving the information on which program and budget decisions are made, and especially by the new emphasis on program results. At this stage of understanding of program outputs and combinations of personnel, material, and facilities that are required to achieve the results sought, primary emphasis needs to be given to program evaluation, research, and experimentation that can yield information for appropriately designed grants-in-aid. But even now, steps could be taken to review the entire grant-in-aid mechanism in use and to ask whether greater latitude can be given to States, cities, and counties.

Adoption of an unconditional grant as supplementary to categorical aids would help to bridge the gaps in program covered by present categorical grants as well as to improve the fairness and efficiency of the tax sources used in financing of State and local public services. The comparing processes of program analysis applied to these services would help assure that the augmented Federal tax funds would be spent to provide the most wanted public services at the lowest cost.

REFERENCES

(1) (In) National Committee for Support of the Public Schools: Education in the States. Washington, D.C., The National Committee, 1966. p. 29.

(2) State-Local Finances Project: PPB Note 7. Washington, D.C., The State-Local Finances Project, The George Washington University, June 1968.

(3) Hatry, Harry: Criteria for Evaluation In Planning State and Local Programs. A study submitted by the Subcommittee on Intergovernmental Relations, U.S. Senate, 90th Cong., 1st sess., 1967.

(4) U.S. Department of Health, Education, and Welfare: Toward a Social Report. Washington, D.C., the Department of HEW, January 1969.

(5) Wholey, Joseph S. and Silver, George A., M.D.: Maternal and Child Health Care Programs. Washington, D.C., U.S. Department of Health, Education, and Welfare. Program Analysis 1966–6- October 1966.

(6) Kirsch, Lawrence: Draft preliminary memorandum on a new approach to grants-in-aid. Prepared for the U.S. Bureau of the Budget, 1968.

(7) Mushkin, Selma J. and Cotton, John F.: *Functional Federalism: Grants-In-Aid and PPB Systems.* Washington, D.C., State-Local Finances Project, The George Washington University, November 1968.

(8) Mushkin, Selma J. and Herman, Brian: The Search for Alternatives. Washington, D.C., State-Local Finances Project, The George Washington University, October 1968.

Section D

Limits on the Consideration of Expenditure Alternatives: The Budget and the Bureaucracy

BUDGET "UNCONTROLLABILITY" AS AN OBSTACLE TO IMPROVING THE ALLOCATION OF GOVERNMENT RESOURCES [1]

BY MURRAY L. WEIDENBAUM

Murray L. Weidenbaum is Professor of Economics at Washington University in St. Louis. He is currently on leave, serving as an Assistant Secretary of the Treasury [redacted].

The application of economic analysis to Federal expenditure policy can lead to improved effectiveness of public programs only if expenditures can be reallocated in response to the findings of analysis. Currently, a substantial share of the Federal budget cannot be transferred from one program to another through the budget review and appropriations process. Professor Weidenbaum cites four categories of Federal expenditure for which budgetary allocation cannot be controlled by annual Congressional action on appropriation bills. These categories are trust funds, permanent and indefinite appropriations, fixed charges, and ongoing projects. In analyzing the budgets of each of the Federal agencies for 1969, Professor Weidenbaum finds that only 52 percent of the total Federal budget is "relatively controllable." The degree of control present in agency budgets ranges from 0 to 100 percent. He states: "Under present law it is almost futile to perform benefit/cost . . . analyses which may demonstrate that the Government obtains a lower return on its investments in highway transportation than in air transportation or some other alternative and, hence, that some shifting of funds might improve economic welfare." The analysis concludes with recommendations for reducing these institutional obstacles to the implementation of sound public expenditure policy.

Introduction

The increased efforts that economists and others have been making in recent years to improve the concepts and procedures for allocating public resources make it especially necessary and desirable to focus greater attention on the obstacles to making these improvements operational. One major set of obstacles to improving public resource allocation is the legal and other institutional constraints that limit the discretion of governmental policymakers.

For example, under present law it is almost futile to perform benefit/cost or similar analyses which may demonstrate that the Government obtains a lower return on its investments in highway transportation than in air transportation or some other alternative and, hence, that some shifting of funds might improve economic welfare. The futility arises from the simple fact that the major financial authorizations for highway programs are not contained in the appropriation bills requested by the President and enacted by the Congress, but in the relatively long-term legislation which authorizes the Federal-aid highway program. Thus, the Congress cannot, through the budget review and appropriations process, in practice effect a transfer of funds from surface to air transportation by reducing the appropriations for the Bureau of Public Roads and increasing those

[1] An earlier version of this analysis appears in M. L. Weidenbaum, "On the Effectiveness of Congressional Control of the Public Purse," *National Tax Journal*, December 1965. The author is indebted to Mr. Suk Tai Suh for assistance in developing the statistical materials used here.

for the Federal Aviation Agency, two component units of the Department of Transportation.

Similarly, there is no discretion through the budget process to shift funds from an income-maintenance program such as public assistance to aid to education, both functions of the Department of Health, Education, and Welfare—or to any other purpose whether it involves expenditures or tax reduction. This rigidity arises because the expenditures under the public assistance program are in the nature of fixed charges; they are predetermined by statutory formulas governing Federal matching of State disbursements for public assistance. Given the permanent statute on the books, the amount that the Federal Government spends on this income-maintenance activity each year is determined by the pattern of state welfare disbursements. Neither the President nor the Congress can much influence the amount of Federal expenditures in this area within the confines of the budget process. Changes in the basic social security legislation would be necessary.

There are many other examples of these institutional obstacles to improving the allocation of public resources, as will be shown later on a more comprehensive basis. The end result of course is that the process of public resource allocation is hardly that deliberate and systematic choice among alternatives that economists try to envision. Rather, it is a fragmented and compartmentalized affair. Many of the key decisions are not made during the budget process or within the budgetary framework at all.

It is an earlier stage of the process which is the effective point of decision-making on numerous government spending programs—the enactment of substantive and often permanent legislation. This is the birth stage, and rebirth and growth stages, of a substantial proportion of Federal spending. This is the stage where many of the basic policy decisions are made—the nature of farm subsidies, the types of public assistance payments, and the level of highway grants. However, since it is the substantive committees of the Congress which handle enabling or authorizing legislation (e.g. Commerce or Foreign Relations or Public Works), rather than the appropriations committees, cost implications of the new programs often are relegated to secondary consideration or even ignored.

As will be demonstrated below in quantitative terms, the effectiveness of appropriations control over Federal Government expenditures is far less than it superficially appears to be.[2]

This study of the techniques of governmental budgeting may shed some light on the substantive issues involved in the allocation of government resources. It may help to explain, for example, why the military budget goes through cycles of alternate expansions and contractions, while the expenditures of domestic-civilian activities—notably the social welfare programs—continue to rise almost without interruption. The basic explanation presented here is in terms of the differences in the relative controllability, through the appropriations review process, of the different types of government spending programs.

---

[2] This substantive point is developed more fully in M. L. Weidenbaum, *Federal Budgeting: The Choice of Government Programs*, Washington, D.C., American Enterprise Institute for Public Policy Research, 1964.

Types of Budget Controllability

This study focuses on the effectiveness of congressional power over the public purse, as measured by the degree to which the Presidential budgetary recommendations are subject to substantial modification through the appropriations process. In most cases, the discretion of the Executive Branch in preparing the budget estimates is also limited by similar institutional obstacles.

The rather narrow definition of controllability of government funding used here needs to be emphasized. The analysis is being made from the viewpoint of annual action by the Congress on the appropriation bills that finance the various government agencies. Given a long enough time span and the support of the Congress as a whole, virtually all Federal spending programs are susceptible to modification, if not elimination. If it so wished, the Congress could repeal the substantive, permanent legislation requiring public assistance grants or veterans pensions or farm price supports, or at least modify the statutes to make them more permissive. In time, it could conceivably retire the public debt and thus obviate the need for annual interest payments or at least reduce the size of the debt to be serviced.

Nevertheless, in practice the President and the Congress do not face each year's budget preparation and review cycle with a clean slate; they must take account of large accumulations of legal restraints within which they must operate.

From the viewpoint of appropriations review, there are thus numerous exogenous forces and factors which they must take account of and cannot effectively control: the number of eligible veterans who apply for pensions or compensation, the amount of public assistance payments made by the states and for which they must be partially reimbursed according to prescribed matching formulas, and so forth. The relatively controllable portion of the budget, from this viewpoint, consists of those government spending programs where the determining factors are endogenous to the appropriations process, which may modify them, at least to a considerable extent.

Four categories of exogenous institutional barriers to improving (or at least changing) the allocation of government resources are identified here: trust funds, permanent and indefinite appropriations, fixed charges, and ongoing projects. These categories are not mutually exclusive and thus individual programs have been assigned to them sequentially; that is, all Federal Government activities operated through trust funds have been assigned to that category, even though the great bulk is financed through permanent and indefinite appropriations. Thus the category of permanent or indefinite appropriations is limited to Federal activities not operated through trust funds. Similarly, activities financed under permanent appropriations may be viewed as a fixed charge on the annual budget. Nevertheless, only programs which do not fall within the two categories mentioned previously (trust funds and permanent or indefinite appropriations) are shown as fixed charges. Thus, double counting is avoided.

TRUST FUNDS

The first category of relatively uncontrollable items dealt with here are the so-called trust funds. These vary from the large social insurance type of mechanisms, such as the old-age, survivors', and disability insurance program, to the gift fund for the Library of Congress. The common characteristic of these trust funds which is relevant for the present inquiry is that they are generally financed through permanent appropriations which do not require annual action by the Congress. As stated in one recent Budget Document:

> "Most trust fund receipts are made available for use by permanent law, without requiring further action by Congress."[3]

Another clear indication of the relative uncontrollability of these trust funds through the budget process is that they generally do not even appear in the annual appropriation bills. In the case of the social insurance funds, the actual level of expenditures is determined by the number of eligible persons who apply for benefits during a given year.

For grants to states for highways, the Federal-aid Highway Act of 1954 and amendments to it not only authorize the program but also provide authority to enter into obligations, in this case to commit the Federal Government to make grants to the States at a later date.* This bypassing of the appropriations process is often referred to as "backdoor spending." Technically, however, it is backdoor financing. The actually disbursements of the Federal funds to the states require the Congress to enact so-called "appropriations to liquidate contract authorizations." The latter is a mere formality. There is virtually no Presidential or Congressional discretion over these liquidating appropriations—the Government was financially committed at an earlier point, at the time the obligations were incurred.

In the Federal budget for the fiscal year 1969, trust funds accounted for $55.1 billion or 27 percent of the total budget authorizations requested for the year.

PERMANENT AND INDEFINITE APPROPRIATIONS

In addition to the trust funds, there are numerous permanent appropriations which are contained in budget funds. The largest of these is the permanent and indefinite appropriation for the payment of interest on the national debt: "Such amounts are appropriated as may be necessary to pay the interest each year on the public debt" (31 U.S.C. 711 (2) and 732).

Other permanent accounts cover such items as the appropriations to the Department of Agriculture for removal of surplus farm commodities and to the Department of the Interior for range improvements. Thirty percent of gross customs receipts is automatically available to finance the agriculture program each year, regardless of estimated need or relative desirability vis-à-vis the changing mix of pub-

---

[3] *The Budget of the United States Government for the Fiscal Year Ending June 30, 1965*, Appendix, Washington, D.C. U.S. Government Printing Office, 1964, p. 898.

*Further discussion of this issue is found in the papers by Achinstein in this volume, and Nelson in vol. 3 of this collection.

lic sector activities. One-third of grazing revenues from Federal lands are similarly available for range improvement work.

A related category of funding is the "indefinite" appropriations. Although these are contained in the annual appropriation bills, they are in the nature of a blank check good for one year. Indefinite appropriations authorize a government agency to spend the sums necessary to meet a given specified requirement. For example, the Post Office Department is financed through an annual indefinite appropriation. So is the retired pay of commissioned officers of the Public Health Service.

In the fiscal 1969 budget, permanent or indefinite appropriations (other than to trust funds) accounted for $20.2 billion or 10 percent of the total budget authorizations requested.

OTHER FIXED CHARGES

A third type of budget request which is relatively uncontrollable through the appropriations process is often termed a "fixed charge." These are programs where the level of spending is determined effectively by basic statutes rather than through the review of annual appropriation requests. The largest programs in this category are the appropriations for public assistance and for veterans' compensation and pension payments. The Department of Health, Education, and Welfare makes grants to states to reimburse them for a fixed share of the public assistance payments that they make. Similarly, the Veterans' Administration provides statutorily determined benefits to all qualifying veterans or their widows and children who apply.

Although programs such as these are funded through annual definite appropriations, there is little effective control over the actual level of disbursements. Frequently, the initial appropriations turn out to be too low and supplemental appropriations are subsequently requested and routinely approved. There is considerable incentive for the Congress to appropriate less than the initial amount requested in the budget for these items. Thus, it gains some political benefit for supposedly "cutting" the budget. They then can later and much more quietly vote supplemental funds.

In the fiscal 1969 budget, fixed charges (other than those arising from trust funds and other permanent appropriations) amounted to $19.0 billion or 9 percent of budget requests.

PARTIALLY-COMPLETED PROJECTS

The final type of relatively uncontrollable budget activity analyzed here is the amount of new funds requested to continue or complete construction and similar long-term projects started with money voted in the budgets of earlier years.* The almost unassailable justification for these appropriations is the old question, "What is the value of just half a bridge?" Typically for government agencies with large construction programs, such as the Army Corps of Engineers and the Department of the Interior, each year's budget request is dominated by funds needed for projects begun under prior year budgets.

*Further discussion of this issue is found in the paper by Knetsch in vol. 3 of this collection.

One indication of this influence of previous commitments is the fact that the Federal Budget for 1969 estimated that $2.4 billion would be spent in that year to carry on construction projects previously begun and for which a total of $28.8 billion already had been spent prior to the budget year. Even though these expenditure figures are not directly comparable to the appropriation or budget authority estimates used in the present study, the contrast between large amounts of what in effect are sunk costs and relatively small increments of additional funding is clear.[4]

The National Aeronautics and Space Administration (NASA) may constitute a special case at the present time.* The great bulk of its current expenditures is devoted to completion of Project Apollo, the effort to land a man on the moon prior to 1970. Theoretically, the program can be reduced or stretched out and thus the President or the Congress could reduce the funds requested for Apollo. In practice, there is a very natural reluctance to interfere with the successful completion of an undertaking in which the Nation already has invested such sizeable funds (over $15 billion for Apollo during the fiscal years 1959–68 alone).

The data for funds requested to continue or complete ongoing projects, as shown in the tables that follow, are incomplete. In many cases it was not possible from publicly available information to identify the specific long-term projects of many agencies. Thus, the funds shown as relatively controllable are overstated, and the uncontrolled funds understated.

The Department of Defense (military functions) constitutes the major example of this gap in our knowledge and thus no military projects are shown in this category of relative uncontrollable programs.** On occasion individual weapon systems have been cancelled after substantial investment of development and production funds. Nevertheless, budget reviewers in both the executive and legislative branches often are reluctant to terminate a large project, even though the changing course of events indicates that the returns may not be as attractive as originally envisioned.

It may be that nonstatutory, implied commitments may be of overriding importance in military budgets from time to time. During the Vietnam War, for example, the Congress has appropriated virtually all of the funds requested in support of that specific and costly military endeavor. Formally, the $30 billion a year request for Vietnam was subject to substantial reduction by the appropriations committee, and is therefore included in the controllable portion of the budget in this analysis; in practice no substantial modifications of the Vietnam estimates were considered by the Congress. However, the Congress did critically review and modify the non-Vietnam portions of the budget of the Department of Defense.[5]

---

[4] *Special Analyses, Budget of the United States, Fiscal year 1969,* Washington, U.S. Government Printing Office, 1968, p. 82.

[5] See U.S. House of Representatives, Committee on Appropriations, *Department of Defense Appropriations for 1969,* Washington, U.S. Government Printing Office, 1968.

* Further discussion of this issue is found in the paper by Augenstein in vol. 3 of this collection.

** Further discussion of this issue is found in the paper by Enthoven in vol. 3 of this collection.

In essence, what is involved here in justifying this military situation, and comparable civil ones, is an implicit incremental benefit-cost analysis: will the returns from the completion of the total project exceed the additional cost to be incurred in completing it? Clearly, many projects midway in the construction state may show incremental benefit-cost ratios substantially in excess of unity, whereas freshly computed total benefit-cost ratios would indicate far less attractive results. There may be substantial political onus attached to abandoning an effort after the investment of substantial public funds. The completion and operation of a public undertaking where the newly determined estimated costs are greater than the estimated benefits is hardly likely to attract great public attention.

Estimates of Relative Budget Controllability

On the basis of the foregoing analysis, Table 1 was prepared in an effort to indicate the relative controllability of the budget requests of the various Federal departments and agencies. The data cover all of the recommended budget authority (new obligational authority as well as loan authority) contained in the Federal Budget for the fiscal year 1969. Table 1 includes both budget and trust funds and is based on the unified budget concept, the most comprehensive measure of Federal finance available at the present time.

TABLE 1.—CONTROLLABILITY OF FEDERAL GOVERNMENT BUDGET REQUESTS—FISCAL YEAR 1969

[In millions of dollars]

Department or agency	Relatively uncontrollable				Relatively controllable	Total
	Trust funds	Permanents, indefinites	Fixed charges	Ongoing projects		
Funds appropriated to the President	1,324				4,819	6,143
Agriculture	68	735	3,831		2,896	7,530
Commerce	134	214			679	1,027
Defense—military	7		2,313		76,796	79,116
Defense—civil	9	4		950	344	1,307
Health, Education, and Welfare	37,670	41	7,456	13	6,190	51,370
Housing and Urban Development	159	1,821	358		3,004	5,342
Interior	97	268		180	312	857
Justice					542	542
Labor	4,095		145		596	4,836
Post Office		920				920
State	12	2			414	428
Transportation	4,703	70	51		1,701	6,525
Treasury	39	15,425			−54	15,410
Civil Service Commission	3,626		42		131	3,799
General Services Administration		1	2		327	330
Railroad Retirement Board	1,064		18			1,082
Veterans' Administration	746	12	4,664		2,368	7,790
NASA	1			2,133	2,235	4,369
Export-Import Bank		608				608
Farm Credit Administration	535					535
All other	773	97	91		896	1,857
Total	55,062	20,218	18,971	3,276	104,196	201,723

Note: Includes requested new obligational authority and loan authority.

Source: Based on data contained in Budget of the United States Government, Fiscal Year 1969, and appendix.

In the aggregate, the trust funds, the ongoing construction projects, and the other permanent and indefinite appropriations and fixed charges account for a major share of the budget—$97.5 billion or 48 percent of the total budget authority requested in the fiscal year 1969. It should be emphasized that where the budget document and available

supporting materials did not provide sufficient detail, or where any doubtful cases existed, the items in question were treated as controllable. Hence, there may be some significant underestimation of the relatively uncontrollable portion of the budget shown here. As mentioned earlier, there undoubtedly is an underestimate in the ongoing project category.[6]

VARIATIONS BY AGENCY AND PROGRAM

Were the fixed charges and other relatively uncontrollable items distributed proportionally to the size of the budgets of the various Government agencies, the interference with the allocation of Government resources might be less than is presently the case. However, as shown in table 2, this is hardly the case. Some agency programs virtually escape the scrutiny of effective annual budgetary review—the Post Office, the Export-Import Bank, the Railroad Retirement Board, the Farm Credit Administration, and the great bulk of the Treasury Department.[7]

TABLE 2.—*Relatively controllable portions of agency budgets—Fiscal year 1969 budget requests*

Justice	100
General Service Administration	99
Defense (military)	97
State	97
Funds Appropriated to the President	78
Commerce	66
Housing and Urban Development	56
NASA	51
All other	48
Agriculture	38
Interior	36
Veterans' Administration	30
Defense (civil)	26
Transportation	26
Health, Education, and Welfare	12
Labor	12
Civil Service Commission	3
Post Office	0
Treasury	0
Railroad Retirement Board	0
Export-Import Bank	0
Farm Credit Administration	0
Average for Federal Government	52

Source: Table 1.

At the other end of the controllability spectrum, all or almost the entire annual budgets of the Department of Defense (excluding civil

[6] For what was perhaps the pioneering attempt to analyze the controllability of Federal spending, but limited to the administrative budget, see "Controllability of 1952 Budget Expenditures," in Joint Committee on the Economic Report, U.S. Congress, *January 1951 Economic Report of the President,* Washington, D.C., U.S. Government Printing Office, 1951, pp. 89–103.

[7] Interfund adjustments complicate the Treasury figures. In practice, the budgets of the operating bureaus are generally subject to effective annual review.

functions such as the Corps of Engineers' construction work), the Departments of Justice and State, and the General Services Administration are subject to effective control through the annual budget process.

An interesting contrast appears between the two departments with the largest budgets, one military and the other civilian. The Department of Defense—which received most of the funds appropriated for national defense purposes—operates with very few and very small trust funds and other fixed charges. Almost all of its budget is subject to annual scrutiny.[8] In comparison, only one-tenth of the HEW budget can effectively be altered during the annual budget cycle. Most of the funds spent are insulated by permanent and indefinite appropriations and other long-term statutory commitments.

Upon further examination, it can be seen that a relatively small number of large programs account for the bulk of the funds which are relatively immune to effective budgetary control. The following 12 programs of over $1 billion each account for over $85 billion or 88 percent of the portion of the fiscal year 1969 Budget which is here estimated to be "relatively uncontrollable":

	In millions
Social security trust funds	$37,670
Interest on the public debt	15,200
Public assistance	5,765
Veterans' pensions and compensation	4,654
Highway grants to states	4,650
Unemployment insurance	4,095
Civil service retirement payments	3,626
CCC (Farm price supports)	3,362
Military retired pay	2,275
Project Apollo	2,133
Medicare (Treasury contribution)	1,360
Railroad retirement payments	1,064

THE RELATIVELY CONTROLLABLE PORTION OF THE FEDERAL BUDGET

Table 3 shows the distribution by agency of the relatively controllable portion of the Federal budget authorizations requested for the fiscal year 1969. It is apparent that the Department of Defense accounts for the great bulk of the funds where the President and the Congress possess substantial discretion over the amounts initially requested (74 percent). For purposes of comparison, it can be noted that the DOD represents 38 percent of the total Federal budget.

[8] As pointed out earlier, the Congress may be reluctant to exercise this potential control over the military budget during wartime and similar emergency periods.

TABLE 3.—*Distribution of relatively controllable budget requests—Fiscal year 1969 budget requests*

	Percent
Defense (military)	74
Health, Education, and Welfare	6
Funds Appropriated to the President	5
Agriculture	3
Housing and Urban Development	3
Transportation	2
Veterans' Administration	2
NASA	2
Commerce	1
Labor	1
All other	1
Defense (civil)	*
Interior	*
Justice	*
State	*
Civil Service Commission	*
General Services Administration	*
Post Office	0
Treasury	0
Railroad Retirement Board	0
Export-Import Bank	0
Farm Credit Administration	0
Total	100

*Less than one-half of 1 percent.

Source: Table 1.

A handful of other departments and agencies—Agriculture, HEW, Transportation, NASA, and the Veterans' Administration account for the bulk of the remainder of the relatively controlled portion of the budget.

REDUCING THE INSTITUTIONAL OBSTACLES

The data presented earlier lead to the rather striking conclusion that the great bulk of the expenditures for the domestic civilian agencies of the Federal Government is authorized virtually automatically as a result of the basic, continuing commitments previously enacted by the Congress, rather than through the deliberations of the annual budgetary process. Somewhat less conclusively, it appears that the military programs are susceptible to effective budgetary review to a far greater extent.

For most of the nondefense programs, the effective point of control appears to occur not at the time that the appropriations are voted, but at the earlier period where the Congress enacts the basic legislative commitments, that is, the rates of veterans' pensions or social security benefits.

For purposes of analysis, it may be helpful to divide the various uncontrollable items into two categories, "natural" and "artificial" (this attempt at labeling by no means exhausts the possibilities).

The "natural" type of uncontrollable item is exemplified by the permanent, indefinite appropriation for the payment of interest on the public debt. These payments arise directly from the amount and types of public debt issues which are currently outstanding. There is no discretion left at the disbursement phase of the process; the Federal Government simply must honor its promise to pay the interest on its obligations as it falls due. The natural uncontrollability of this

item expense is acknowledged by the Congress in the form of a permanent appropriation to pay interest with no fixed dollar limit.

Similarly, the making of monthly compensation payments to veterans on account of service-connected disabilities is a program which is naturally uncontrollable within the confines of the budget process. The law requires monthly payments to all those certified by VA doctors as possessing a given percentage impairment of earnings. However, in this case the Congress insists on annually reviewing the appropriation for the payment of veterans' pensions and compensation. It is hard to characterize this congressional review as anything other than wheelspinning or having "fun and games" with the budget. Moreover, this exercise in futility diverts executive branch and congressional time and attention to the budget away from the areas where they can significantly alter the results.

In sharp contrast, there are numerous government programs which are artificially uncontrollable as a result of statutory law, but which lend themselves, through changes in substantive legislation, to effective annual budgetary review. For example, under section 32 of the act of August 24, 1935 (U.S.C. 612 C) an amount equal to 30 percent of annual customs receipts is automatically appropriated into a permanent, indefinite special fund for the "removal of surplus agricultural commodities." These amounts bear little relationship to the requirements for such funds. In fact, recent appropriation acts have authorized transfers of funds to the school lunch program and for related activities. Clearly, the amount of funds automatically appropriated exceeds the needs of the basic activity financed by the appropriation.

The annual grants of $50,000 paid to each state and Puerto Rico for A & M colleges similarly are made under a permanent appropriation act. Neither the Bureau of the Budget nor the President nor the Congress has any opportunity to review the annual appropriation request and thus annually redetermine the continued need for or desirability of these payments.

There are numerous other examples. Many permanent indefinite appropriations to the Department of the Interior are tied to a portion of revenues from sales or rentals of government assets and bear little relationship to the current requirements for Federal expenditures for the activity to which they are earmarked. Thus, visitor fees at Yellowstone National Park are automatically used to provide educational expenses for dependents of park personnel, while visitor fees at Grand Teton National Park are used automatically as payments to the state of Wyoming, in effect in lieu of taxes.

Conclusions

Although the analysis of individual government programs presented here is incomplete (partly due to the lack of available data), it is clear that the effectiveness of appropriations control over Federal Government expenditures is far less than is generally appreciated. The following changes might be considered toward reducing these institutional obstacles to improve the allocation of public resources.

1. *A review of the necessity for the numerous trust funds that have been established.*—Some of them—such as those for the financing of social security benefits—appear to somewhat approximate the general

notion of funds held in trust. In many other cases—such as the Federal-aid highway program—it is hard to make a case for segregating the activity from ordinary budget operations. In that particular case, the program of Federal grants to the states did operate out of general revenues until 1954. In good measure, the highway-related excises which are now funneled through the highway trust fund may be viewed more properly as a form of earmarked taxes and treated as a special fund within the regular budget procedure.*

2. *A reevaluation of the need for the various permanent and indefinite appropriations.*—Some of them may have outlived their usefulness. However, there is no automatic or periodic review of their status and a clean slate examination might be most useful.

3. *A reexamination of the "fixed charges" on the budget.*—Some of them might usefully be converted into permanent or indefinite appropriations. In other cases, discretion might be restored to the appropriations committee to determine annually the amount to be voted for the stipulated purpose, in the light of then current conditions and competing requirements. This latter action, of course, would require changing the substantive legislation governing the program.

4. *A focusing of greater attention on "new starts" of construction and other long-term projects.*—It is a natural tendency to place greater emphasis in the budgetary review process on the items with the largest price tags. However, as has been shown, most of the appropriation requests in this category of long-term projects are to continue or complete projects already underway. The point of most effective control is at the outset, prior to the investment of public resources in the project. However, it is precisely at the starting-up stage where the appropriation requests are most modest and thus perhaps more readily approved. A careful weighing of the expected full or long-term costs and benefits is thus extremely important at the outset.[9]

The reduction of these institutional obstacles to maximizing the taxpayers' return on their investment will not of itself result in eliminating relatively low priority and less efficient government activity, but it should make efforts in that direction less difficult.

---

[9] In recent years, the budget requests for military and selected other areas have been prepared on the basis of "full funding" of proposed projects, that is of appropriating the entire estimated cost of a project at the time it is started. This procedure helps to enable the Congress to ascertain the total cost of a project before the work actually begins. However, water resource projects continue to be an important exception to this desirable change. See *Special Analyses, Budget of the United States, Fiscal Year 1970,* Washington, U.S. Government Printing Office, 1969, p. 81.

*Further discussion of this issue is found in the papers by Achinstein in this volume, and Nelson in vol. 3 of this collection.

CONSTRAINTS ON POLICY ANALYSIS AND POLICY IMPLEMENTATION IN THE FEDERAL AGENCIES

By Asher Achinstein

Asher Achinstein is a Senior Specialist in Price Economics at the Legislative Reference Service of the Library of Congress. His paper is addressed to a widely recognized problem which inhibits effective public policy analysis, namely the constraints which prevent agencies from systematically evaluating a wide range of alternative means for accomplishing an objective. The method used by Dr. Achinstein to ascertain the nature of these constraints was to question Federal agencies directly on how they feel themselves to be constricted in applying economic analysis to policy planning and implementation. In their answers, agency officials cited a number of significant constraints. In this paper, they are discussed under the topics of institutional, budgetary, personnel, financing, methodological, and statutory constraints. As an appendix to the paper, excerpts from each of the agencies responding to the questions are presented.

What Do Agencies Find To Be the Major Obstacles in Policy Formation and Policy Decision?

On November 22, 1968 Senator William Proxmire, then Chairman of the Joint Economic Committee, sent a letter to 21 agencies of the Federal Government requesting information concerning the impediments to the use of economic analysis in policy formation and policy decision. Recognizing that improvement in the planning and execution of Federal programs was dependent on the systematic comparison of a wide range of alternative methods for meeting the objectives of expenditure programs, the committee was interested in learning what were the obstacles which prevented agencies from fully exploring alternatives.

Each of the agencies was requested to furnish the committee with a statement showing:

1. The principal legal and institutional constraints affecting the scope and effectiveness of expenditure policy analysis;
2. The ways in which these constraints hinder sound economic analysis; and
3. What changes—legislative, administrative, or others—could be made that might substantially contribute to lessening the influence of the constraints on effective economic analysis.

To illustrate the kind of information desired there was attached to the chairman's letter a copy of a statement submitted by the Department of Transportation in answer to similar questions that he had raised a few months earlier during the course of public hearings of the Subcommittee on Economy in Government. The statement is reprinted on page 412.

It was anticipated that there would be considerable variation in the replies of the agencies—both in length and content—since they

differ in the character and magnitude of their operations, and in the interests and zeal with which top management is prepared to use analytical tools in the development of policy. To be sure, the request came at a time when the influences of the Christmas-New Year holiday season was soon to be felt, in addition to the fact that the Government was in transition from an Old to a new administration. However, in accounting for the differences in replies, the holidays and the election appeared to have played a minor role as compared to such influences as types of agency and management.

Statements were received from the following 20 agencies:

Department of Agriculture.
Department of Commerce.
Department of Health, Education, and Welfare.
Department of Housing and Urban Development.
Department of the Interior.
Department of Labor.
Post Office Department.
Department of the Treasury.
Corps of Engineers.
Agency for International Development.
Atomic Energy Commission.
General Service Administration.
National Science Foundation.
Office of Economic Opportunity.
Peace Corps.
Veterans' Administration.
Federal Communications Commission.
Federal Power Commission.
Small Business Administration.
Tennessee Valley Authority.

These agencies represent about four-fifths of those listed in section 1, and more than two-fifths in section 2, of the 1968 Planning-Programing-Budgeting (PPB) Guidelines Bulletin No. 68–9 of the Bureau of the Budget.

This paper is concerned with reporting on the views of the Federal agencies as to the major obstacles they encounter in developing and implementing program objectives based on the consideration of a wide range of alternatives. After classifying and summarizing each agency's experience with the various factors that limit the role played by economic analysis in the planning and execution of expenditure programs, we present the statements submitted by the heads of the agencies.

Two Agency Biases in Policy Formation

Examination of the agency statements reveals that they generally assert that there are no legislative constraints which are specifically and directly aimed at preventing them from conducting their economic research and analysis. They are free to consider as full a range of alternative ways of meeting the objectives of their programs as imagination, technical skill, and good organization will permit. Having said this much we must hasten to add that there are a number of factors which do influence the scope and effectiveness of the use of analytical tools. The rest of this paper will be concerned with these influences.

Before we begin our review of the specific factors emphasized by the different agencies as limiting their economic analyses, reference should be made to two influences which appear, in varying degrees, to affect top-management's activities in this field. The first constraining influence is the prospect of implementation of the analytical results of an agency's activities. Administrators are practical men and are always faced by the program of limited resources. The scope and nature of the alternatives likely to be considered are frequently influenced by the probability of successfully introducing change or innovation. This point was touched upon in the statement of the Department of Agriculture, as follows:

> Obviously if a program alternative would require major legislative action, substantially alter longstanding institutional arrangements, and appear to require an administrative and political effort out of proportion to the anticipated program improvements, there is hesitation in applying scarce analytical resources to the detailed evaluation of that alternative.

There is a second influence which tends to constrain agency policymakers in implementing the sound economic analysis of programs. In the on-going administration of an agency's activities, there are understandable pressures to concentrate on the justification of existing programs rather than to think in terms of program modifications, new program emphasis, or other possible alternatives to existing policy. This bias is referred to by the Undersecretary of Interior.

> It is recognized that analysis performed by individuals, divisions, bureaus or other organizational units can reflect an agency bias. Those responsible for management of existing programs may have an understandable tendency to concentrate on justification of such programs rather than full consideration of possible alternatives.

Institutional Constraints

Governmental economic analysis is subject to the influence of institutional constraints. The consideration of alternatives may be adversely affected by the organizational structure within the executive branch of the Government, the relations between the agency and the Congress, the extent to which the planning and execution of programs is dependent upon the cooperation of Federal, State and local governments, and other countries, and by the absence of user charges or other cost sharing arrangements when these modes of financing are available.[1]

INTERDEPARTMENTAL CONSIDERATIONS

Several agencies comment on the fact of the limited horizon of the single department or agency when the problem of more effective pro-

[1] Analysis is constrained because there is no gauge of the value which beneficiaries place on the output of the expenditure when, through pricing or cost sharing, such a gauge is available. The implementation of economically sound decisions is made difficult because those who benefit from an expenditure make a very small contribution toward its financing. When such "subsidization" occurs, political pressures may overwhelm economic considerations.

gram analysis calls for an interdepartmental focus. For example, the Department of Commerce points out that almost every program in the Department has significant ties to other Federal programs. It then goes on to say that when two or more Federal agencies share similar objectives, such as pollution control, or economic development, it is essential that the analytical work on cost-benefit analysis should, at least, be done jointly.

LEVELS OF GOVERNMENT

The Department of Labor refers to constraints arising from the fact that several levels of government are involved in many programs.* The impact of such involvement varies with the program. For example, while the Manpower Development and Training Act is implemented by the States, it provides for Federal financing and the establishment of Federal standards of applicant eligibility, program operation and reporting requirements. Federal orienation makes possible data collection and analysis that is comparable on a nationwide basis. By comparison, the Unemployment Insurance system is both administered and primarily financed by individual States. While minimum Federal standards guide the system, the State oriented nature of the program hinders reporting procedures. For example, it is extremely difficult to obtain data on the characteristics of unemployment insurance recipients. The problem is even more acute in the Workmen's Compensation system—a total State program.

ORGANIZATIONAL RIGIDITIES

The Department of Agriculture refers to institutional constraints developed over the years and now well established in custom and practice. Formal agreements with non-Federal cooperating organizations made by Secretaries of Agriculture years ago tend to limit the program and organizational flexibility of incumbent Secretaries. Then too, longstanding assignment of complementary, or supplementary, components of programs to different departments, agencies, and organizations, often results in imbalances that are difficult to correct. It should be added that other government agencies have these problems in common with the Agriculture Department.

NONECONOMIC CONSIDERATIONS AND PRIORITIES IN THE COOPERATING COUNTRY

The Director of the Peace Corps emphasizes that the criteria for program planning cover a broad range of considerations that involve other than economic categories of analysis. Peace Corps programs must relate to one or more of the host country's most critical needs as identified by the country itself. Because Peace Corps Volunteers are integrated into host country programs, it is often difficult to separate any unique Peace Corps output of a project. This limitation is characteristic of a wide variety and large number of projects which the agency now has overseas.

---

*Further discussion of this issue is found in the papers by Mushkin & Cotton in this volume, and Mangum in vol. 3 of this collection.

INFLEXIBILIES CAUSED BY REQUIREMENT OF PRIOR APPROVAL OF PROJECT BY CONGRESS

Several agencies comment on Congressional practices that limit management's ability to put into effect changes indicated by economic analysis. The Post Office Department maintains that cost-saving operational improvements could be made were it not for the fact that management is required to obtain formal or informal approval of the Congress prior to introducing the change. Congress also determines the total number of postal facilities to be acquired and whether they are to be built by the government or leased. The Agency for International Development states that prompt implementation of new alternatives is prevented by the requirement that all new technical assistance and international organization projects must be justified to the Congress before they are launched. The experience of other agencies with the requirement of prior Congressional approval before implementation is referred to in the sections on budgetary and legislative constraints.

Budgetary Constraints

As might be expected, many of the agencies stressed funding and the budgetary process as among the most important factors restricting the planning and implementation of programs. Among the items restricting the consideration of alternatives are the high cost of economic analysis, the limited resources that are made available, the poor timing of appropriations, the rigidities resulting from the earmarking of funds for particular purposes and the inability to switch funds between programs, forced stretch-outs in the case of apropriation-financed projects, the drain on personnel in preparation of appropriation and performance budgets, and the limitations of planning resulting from authorizations made on a year-to-year basis.

HIGH COST OF ANALYSIS

The Department of Interior points out that good analysis takes a great amount of time and effort. The Department's oil shale study, which had considerable bearing on major policy issues and decisons, took approximately one year. To the extent that top people in the agency were involved in the study, they were not available for other studies or program activities.

The Department of Labor underscores the fact that there is a lack of realization of the magnitude of the expense involved in conducting careful and thorough analytical work. Among the examples cited by the Secretary are the improvement of the national employment statistics, and of the management information reporting system. To accomplish the former would require an increase of $10 to $15 million a year above current expenditures, and to implement the latter would cost about $10 million a year.

INADEQUATE FUNDING

The Internal Revenue Service of the Department of the Treasury states rather vehemently that sharp reductions in recent years in budget resources and appropriations have played havoc with program objec-

tives and operational accomplishments. These are well illustrated in the case of auditing tax returns. As a result, the Government is not realizing as much tax revenue as has been legislated and needed to fund its policies and operate its programs. In addition, there is the danger that the general willingness of taxpayers to comply may be eroding under enforcement programs limited in effectiveness due to lack of funds.

The General Services Administration complains that funding levels are imposed upon the GSA, independently of consideration of the quantity of services the agency is called upon to provide for expanding or new Federal programs of other agencies. Because the GSA has had great difficulty in developing and carrying out an effective program to meet Federal office space requirements through construction of Federal buildings, it has had to provide space through the more costly and less satisfactory method of leasing privately owned buildings. Moreover, there has been a growing tendency of Federal agencies to obtain independent authorization and funds for constructing their own facilities. This fragmentation of responsibility for building construction, says the GSA, impairs the effectiveness of a planned public building program, since the appropriation is often made without regard to more urgently needed construction projects which remain unfunded. The GSA proposes legislation which would provide a more consistent level of funding for public building purposes. This calls for "user" rates to be charged occupant agencies for all space provided by GSA, and that the agencies be required to obtain appropriations necessary to reimburse GSA for public building services performed.

POOR TIMING OF APPROPRIATIONS

If the full benefit of the PPB system is to be realized, action on program funding must be taken in time to be fitted to the programs in an efficient and economic way. According to the Treasury Department, the timing of actions on the budget, and in recent years on the availability of funds already appropriated, have seriously limited the utility and effectiveness of program analysis. This effect develops from damage to the relationship between program plans and program execution, from the lack of relationship between actions taken and the assumptions involved in program design, and from the imbalances among and within programs that result from sudden changes.

RIGIDITIES FROM EARMARKING OF FUNDS AND INABILITY TO SWITCH FUNDS*

The Department of Agriculture and the Agency for International Development point to the earmarking of funds for particular purposes as considerably reducing the exploration of alternatives. It may cause overprogramming in one area, with a corresponding deleterious effect on other fields. The Bureau of the Mint and the Post Office Department also comment on the limitations of the appropriations structure which do not permit an agency to switch funds between programs. In

*Further discussion of this issue is found in the papers by Weidenbaum in this volume, and Nelson and Haldi in vol. 3 of this collection.

the case of the Post Office, funds for Departmental and regional personnel are provided by appropriations separate from field personnel. This militates against the flexible use of manpower and the lowering of overall costs. In the case of the Bureau of the Mint, the inability to switch funds, and the necessity for requesting reprogramming authority or supplementary appropriations, has led the Bureau to recommend legislation to set up a fund which would permit the Mint to finance its operations from Mint revenues derived from coin production.

The Department of Transportation mentions another form of rigidity associated with the fact that Federal Highway expenditures are determined by the amounts collected from motor vehicle user taxes as they are placed in the Highway Trust Fund. Moreover, the required allocation of funds to States on the basis of legislated formulas and without regard to relative cost-benefit relationships of projects, is a further constraint on the implementation of sound economic analysis.

STRETCHOUTS IN APPROPRIATION-FINANCED PROJECTS

While the Tennessee Valley Authority states that the Act under which it operates provides it with a high degree of flexibility, nevertheless, in the case of appropriation-financed projects, it suffers from budgetary restrictions resulting in construction stretchouts. It maintains that it plans such projects on the basis of the most efficient schedules, which will, in turn, produce the lowest possible costs and largest possible excess of benefits over costs. Budget-produced stretchouts disrupt these plans and inevitably result in less efficient construction, inflated costs, and a reduction of the ratio of realized benefits to costs.

CONSTRAINTS OF THE BUDGET PROCESS

The Post Office Department refers to a limitation on the performance of economic analysis which must be true of the other departments of the Government, namely, constraints that stem from the budget process itself. Since Congressional budget review is still performed along appropriation lines, the limited analytical resources of an agency must be divided among program analysis, appropriation-based analysis, and the reconciliation of the two approaches. This division of personnel produces delay in the development of the PPB system.

PLANNING RESTRICTIONS RESULTING FROM YEARLY AUTHORIZATIONS

The Agency for International Development maintains that foreign aid authorizations on a 1-year basis constrains the Agency's ability to plan ahead over a meaningful time frame, which for development purposes should really be three to five years. It is willing to settle for a minimum of two years, since respite from the long-drawn out yearly fights over AID authorizations would give top management more time to spend on meaningful planning, programming and operations.

Personnel Constraints

A number of agencies, such as the Departments of Commerce, Interior, and Labor, stress the shortage in the supply of economic ana-

lysts as an important limiting factor in conducting a continuous and comprehensive program of analysis. At the same time, these and other agencies are faced by expenditure and personnel hiring limitations.

SHORTAGE OF QUALIFIED ANALYSTS

The Department of Labor refers to the fact that the highest grades are, in general, given to individuals with administrative as opposed to analytical responsibility. It recommends making more supergrade positions available to facilitate the recruitment and retention of persons with analytical capability of the kind required for successful economic analysis.

INEFFICIENT EMPLOYMENT POLICIES AND PRACTICES

The Internal Revenue Service points out that Federal employment policies and practices limit the range of alternative approaches available in meeting objectives. The limits established on year-end employment, temporary employment, and overtime, together with uniform, nationwide salary schedules in the face of geographical differences in the cost of living and in the demand for professional skills, result in the acceleration of employee turnovers and in the reduction of the Service's ability to compete with the private sector.

ARTIFICIAL EMPLOYMENT CEILINGS

The Post Office Department, the Internal Revenue Service, and the Bureau of Engraving and Printing, refer specifically to the deleterious effect of Public Law 90–364, the Revenue and Expenditure Control Act of 1968. This act, which stipulates that the number of employees of an agency shall not exceed the number of employees at a base period, is cited by the Internal Revenue Service as the most drastic example of hiring restrictions which make orderly program planning virtually impossible.

Financing Constraints

Two principal constraints, which are mentioned as discouraging consideration of the full range of alternatives in developing economic policy, are statutory limits on debt obligations and statutory interest rate ceilings. The Treasury Department refers to the 4¼-percent interest rate ceiling on Treasury debt issues of more than 7 years maturity, and the statutory limit on the dollar amount of public debt obligations, as restraining influences on program planning.

The Tennessee Valley Authority urges that the ceiling on the amount of power revenue bonds the TVA may have outstanding at any time be eliminated, in the interests of efficient long-range planning for increasing the power capacity to meet the growing needs of the area served by the TVA.

The Agency for International Development recommends that it be given greater flexibility in setting the terms of aid to fit individual country situations, particularly, the minimum interest rates at which development funds may be loaned.

The Small Business Administration refers to the fact that several different statutory provisions govern the rate of interest charged on direct loans placed by SBA. In the interests of a more effective use of resources, it recommends greater uniformity of rates among the different programs, with a broader range of discretion allowed to the Administrator in setting specific rates.

Methodological Constraints

One of the problems faced by agencies attempting to develop PPB types of analyses is that they vary considerably in the extent to which their programs lend themselves to quantification in terms of cost-benefit analysis. We will refer to three different agencies which comment on their difficulties.

According to the Agency for International Development, there are intrinsic complications in applying the PPB approach. Objectives can be identified in general terms without too much difficulty, but defining the specific "outputs" in measurable terms useful for analysis is not an easy task. Analyzing just how "inputs" of U.S. foreign resources—which are marginal in relation to the resources of the recipient country—affect the outputs and objectives is even more complex.

The Department of Labor refers to certain methodological problems, especially in manpower-related fields, which hinder satisfactory economic analysis. These relate to such questions as defining skill needs in the economy, projecting these and other needs, quantifying the more intangible costs and benefits connected with manpower development, and determining the desirability of manpower development or income maintenance programs for different population groups.

The National Science Foundation states that while the PPB effort has been useful as a formal mechanism to reexamine objectives, it has not been able to utilize economic analysis extensively in its operations. Its difficulties with economic analysis stem from limitations in the state of the art rather than from legislative or institutional constraints. The "outputs" of the agency are primarily: newly gained scientific knowledge, men and women educated in science, and institutions better equipped and staffed to conduct research and education in the sciences. These do not readily lend themselves to economic analyses requiring systematic comparison of resources with expected outputs, together with appropriate measures of effectiveness.

Statutory Constraints

It is quite apparent that many of the limitations on policy analysis and policy decisions which have been referred to as institutional, budgetary, personnel, and financial, are also legislative in character. In this section we will summarize statutory constraints not previously mentioned. What they have in common with many of the earlier restrictions is that legislative action for their removal, or for lessening their influence, is generally required.

With respect to the influence of statutory requirements concerning grant-in-aid programs, agencies may differ in their evaluations.*

*Further discussion of this issue is found in the paper by Mushkin & Cotton in this volume.

According to the Department of Agriculture, grant-in-aid programs with allocation formulas specified by law may restrict consideration of alternatives, and reduce the possibility of successfully introducing change or innovation. The Department of Agriculture operates several grant programs where the statutory formula specifies allocations to the States or local governments based on criteria related to social needs at the time the laws were enacted. Sometimes these formulas are not appropriate to meet the needs as new and quite different problems arise.

The Department of Housing and Urban Development emphasizes different aspects of grant-in-aid programs. It points out that if some statutory provisions constrain its program operations, it is generally in ways which are provided deliberately by the Congress to make certain that programs are operated or distributed as intended. Examples of these provisions are limitations on the percentage of funds which can be used in any single State, and restrictions of eligibility for financial assistance to cities of specific population sizes. However, these constraints do not prevent analysis of alternatives, or the submission of proposals for eliminating statutory provisions which are deemed to restrict effective program operations unwisely.

The Department of Labor mentions two types of statutory constraint which limit its analytical work. There are statutes, among them the Social Security Act, which provides strict limitation by formula on the amount of available funds which may be used for administration. Since it is from these funds that economic analysis is usually conducted, these limitations impair the effectiveness of that analysis. In addition to the constraint on administrative ability there is the constraint on research capability. The Wage and Labor Standards Administration lacks the legal authority for making research grants. As a result, many areas vital to economic analysis which lend themselves to small projects suitable for grants rather than contracts are left unexplored.

In the case of the Post Office Department, the Postmaster General states that experimentation essential to thorough economic analysis is often impossible because of the fact that many postal matters are governed by statutory requirements. As an example, legislative action is required for the Post Office Department to conduct a regional test of combined air first-class letter rate to determine public reaction and the cost effects of a merged service. Then again, many statutory restrictions on the use of transportation services severely limit transportation planning. As an example, the Department's inability to negotiate with air carriers has hindered the development of containerization, with the result that costly manual handling procedures are still in use at most air transfer points. In the matter of rates, economic analysis has demonstrated that some postal rates are below the level at which postal out-of-pocket costs are covered. Since rates are presently established by statute, the opportunity for change is limited.

The Office of Economic Opportunity cites one type of constraint which prematurely forecloses the consideration of other possibilities in dealing with the problems of the poor. The Economic Opportunity Act prevents the OEO from conducting experimental and demonstration projects that might encourage rural poor to migrate to urban

areas. It thereby prevents the agency from giving full consideration to a major policy alternative.

For the Treasury Department, the major statutory constraints have been referred to under budgetary and personnel limitations. The Department mentions other statutory provisions, perhaps of lesser importance, which restrict the operations of the Bureau of Customs and the Internal Revenue Service, and whose removal is recommended.

The Agency for International Development mentions two types of legal constraints against the actual provision of assistance to a country under the Foreign Assistance Appropriation Act. First, the Foreign Assistance Appropriation Act sets arbitrary limits on the number of countries to be aided through development loans, technical assistance, and supporting assistance. Second, the Act contains a large number of constraints which are designed to achieve desirable purposes other than development. We shall refer to only a few of the restrictive nondevelopment purposes. The act states that no assistance can be provided, or aid may be terminated, if the country furnishes, or permits ships or aircraft under its registry to transport materials to Cuba or North Vietnam, if its government owes a debt to a U.S. citizen, if its government has nationalized property of a U.S. citizen without compensation, or if a country is delinquent with respect to its U.N. obligations, and the like.

The Department of Transportation states that the PPB system makes unnecessary such provisions as section 7(a) and 4(b)(2)(B) of the Department of Transportation Act.* The former limits the role of the Department in developing criteria for the formulation and evaluation of proposals for the investment of Federal funds when those proposals are concerned with certain types of projects—for example, water resource projects. The latter provides that the adoption, revision, or implementation of any transportation policy, or any investment standards or criteria, must receive prior authorization by the Congress. These constraints, together with certain provisions in law relating to the Federal-aid highway program and the Federal Airport Act, unnecessarily restrict the Department's authority to manage its programs.

The statement of the Corps of Engineers appears to be the most optimistic of any received by the Joint Economic Committee in answer to its inquiry on constraints of existing legislation. In its view, the Water Resources Planning Act enables the agencies of the Federal Government "to formulate truly comprehensive plans, and in doing so to give consideration to all alternative means of utilizing, developing and conserving the Nation's water resources." This is in contrast to the view of the Department of Transportation as indicated in the previous paragraph. In fact the Chief of Engineers questions the Department of Transportation's interpretation of section 7(a) as follows:

> Strictly speaking section 7(a) does not, in my opinion constitute a legislative directive prohibiting sound economic analysis. . . . It has been the view of the Corps of Engineers that the real purpose of section 7(a) was to establish a "standard" for use in

*Further discusssion of this issue is found in the papers by Nelson and Knetsch in vol. 3 of this collection.

> determining whether a particular waterway does or does not, qualify for inclusion in the National transportation system. Such a standard expresses a policy of Congress. It seems probable, therefore, that what the Department of Transportation intended to convey to the Committee was the fact that application of the standards of section 7(a) will not necessarily assure the selection of that alternative which produces the greatest net economic benefit. This is quite different than saying that section 7(a) constitutes a constraint upon sound economic evaluation by the Executive Branch. This section requires the Executive Branch to inform the Congress whether a contemplated waterway would, or would not, qualify under the standards set by that section. It does not, in my opinion, prevent the Executive Branch from applying any other tests it may consider desirable, and informing the Congress of the results of these tests.

While the interpretation of this provision by the Corps of Engineers differs from that of the Department of Transportation, both agencies would presumably agree that the very existence of this uneconomic statutory criterion is a serious impediment to instituting appropriate ongoing analysis of waterway investments. While it does not "prevent" such analysis, it does impede it.

We have thus far attempted to highlight the obstacles encountered by agencies of the Federal Government in program planning and implementation. It has not been our purpose to draw up a statistical balance of what agencies regard as their major impediments. PPB analysis is too new for such an attempt. Suffice it to note that there are many restrictive factors with which the agencies must deal. At this stage in the development of the PPB system, it may be useful to know where many of the agencies find the "shoe pinching" in the areas of their responsibility. At the same time, it may also be useful to note the following statement made by the Under Secretary of the Interior:

> We have observed that there often is a tendency to use legal or institutional factors as an excuse for not undertaking thorough analysis when, in fact, it may be that such constraints are the primary reason that analysis should be made as a basis for recommending legislative or program changes.

Finally, it is instructive to examine the actual statements submitted by the various agencies, which we now reproduce.

AGENCY STATEMENTS

Statement of John A. Schnittker, Under Secretary of the Department of Agriculture

There are no specific legal prohibitions on the scope of analysis performed in the Department of Agriculture. There are, however, a number of legal and institutional arrangements which place significant limitations on the ability of any Secretary of Agriculture to put into immediate effect new or modified program practices which might be indicated by the results of analysis.

Because these limitations affect the probability of successfully introducing change or innovation, they also influence decisions on the scope and nature of the alternatives likely to be considered. Obviously if a program alternative would require major legislative action, substantially alter long-standing institutional arrangements, and appear to require an administrative and political effort out of proportion to the anticipated program improvements, there is hesitation in applying scarce analytical resources to the detailed evaluation of that alternative.

Grant-in-aid programs with allocation formulas specified by law are a case in point. The Department of Agriculture operates several grant programs where the statutory formula specifies allocations to the States or local governments based on criteria related to social needs at the time the laws were enacted. Sometimes these formulas are not appropriate to meet the needs as new and quite different problems arise.

Permanent appropriations of earmarked revenues also tend to set some programs apart in the program evaluation and budgeting process. With financial resources tied up by law for specified uses, and therefore not available for reallocation, the incentive for rigorous exploration of alternatives is considerably reduced.

However, I want to emphasize that, while certain restrictions may inhibit or preclude short-term adjustments in programs, we do not refrain from considering alternatives beyond the scope of present laws whenever the social need is sufficiently pressing. Program evaluation often leads to legislative proposals to change or to remove restrictive legal provisions—or to provide added authorities for new programs.

In addition to legal constraints, there are many institutional constraints developed over the years and now well established in custom and practice. Formal memorandums of agreement with non-Federal cooperating organizations signed by Secretaries of Agriculture many years ago tend to limit the program and organizational flexibility of incumbent Secretaries. Long-standing assignment of complementary, or supplementary, components of programs to different departments, agencies, and organizations, often results in imbalances that are difficult to correct. These problems are so common to all Government operations as to not warrant detailed discussion here.

STATEMENT OF C. R. SMITH, SECRETARY OF THE DEPARTMENT OF COMMERCE

I know of no legislative or institutional limitation specifically aimed at the Department or any of its agencies on doing economic research or analysis. We are not enjoined from considering a full range of alternative ways of carrying out our program missions. Nonetheless, like all Federal agencies we operate under two general kinds of constraints.

First, economic analysts are in scarce supply and like all other agencies we face expenditure and personnel hiring limitations. While this is patent, it would be a mistake to underestimate the limits these factors place on mounting continuing and comprehensive program analysis.

Second, a single Federal department or agency is often not the proper vantage point for economic analysis. Almost every program in the Department of Commerce has significant ties to other Federal programs. Where two or more Federal agencies share similar missions, such as pollution control or economic development, the analysis done to determine the most cost-effective way of carrying out the program involved should, at least, be done jointly. The horizons of a single department are very limited in cases like these. In my view, program analysis should move toward an interdepartmental focus as rapidly as possible.

The second part of your inquiry was on the legal and institutional limitations for implementing the results of economic analysis. Here again, I can respond on a department-wide basis. I know of no peculiar limitations on the Department of Commerce. But in the short run, all enabling legislation must be considered not only a grant of authority, but a limitation on its use. For example, the Economic Development Administration is not simply empowered to promote economic development, the act under which it operates specifies the way the agency is to promote economic development. These are, at the same time, limitations as well as grants of power.

In the long run, however, provisions in enabling legislation or executive orders can only be considered a limitation if the agency has done the economic analysis, discovered that the present law prevented it from pursuing the most cost-effective course of action, requested the needed change, and has been turned down.

STATEMENT OF WILBUR J. COHEN, SECRETARY OF THE DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

As we view economic analysis as a part of our PPB system, there are no legal or institutional constraints which would prevent us from considering a full range of alternative ways of meeting the objectives of our programs. There are, of course, constraints which would prevent us from adopting certain alternatives which analysis might show to be the best way to proceed. In such cases, the analysis might form the basis of proposed legislation to provide the required legal authority to carry out the program in the most effective way.

With the creation of the Office of the Assistant Secretary (Planning and Evaluation), early in 1966, the contribution of economic analysis took an important step forward in HEW. New staffs of analysts have come into the decisionmaking picture, both in the Office of the Secretary and in several constituent agencies. The analytical work done has raised the level of information available concerning budget and legislative issues and affected major and minor decisions. Other work has been less visible; for example, improving our estimates of the costs and benefits of family planning. Still other studies have yet to be translated into decisions; for example, work on alternative income maintenance programs, and programs to support higher education. Both of these studies are evaluating alternatives which could require legislative changes. It is too early to tell what the conclusions will be.

STATEMENT OF ROBERT C. WOOD, SECRETARY OF THE DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

There is nothing in HUD statutes or administrative regulations which inhibits economic analysis and full consideration of alternatives. It is true that some statutory provisions constrain program operations, but generally only in ways which were provided deliberately by the Congress to make certain that programs are operated or distributed as intended. Examples of these provisions would be limitations on the percentage of funds which can be used in any single state and restrictions of eligibility for financial assistance to cities of specific population sizes. However, these constraints do not prevent analysis of alternatives or the submission of proposals for eliminating statutory provisions which are deemed to restrict effective program operations unwisely.

STATEMENT OF DAVID S. BLACK, UNDERSECRETARY OF THE DEPARTMENT OF INTERIOR

We have had each Interior bureau and office review your request and give us its evaluation. These reviews indicate that we have no significant barriers to adequate economic analysis other than our physical capability in terms of staffing and data availability.

The Department's PPB System is based upon the concept that analysis, including the display of alternatives, can provide a more adequate basis for legislative, policy, and budgetary decisions. The intensity of such analysis depends upon the magnitude of the decision, the adequacy of information, and the time and staffing available for analysis before decisions must be made. To be fully effective, such analysis must be made on as objective a basis as possible. Evaluation of existing legal and institutional constraints should be a part of the analysis. Accordingly, we do not feel that we are prevented from giving appropriate consideration to the full range of alternatives. Perhaps a few examples would help.

Land and Water Conservation Fund.—The Act of 1965 (78 Stat. 897) provided that certain revenues be deposited into the Land and Water Conservation Fund for outdoor recreation planning, land ac-

quisition and development. Administration of related programs involved problems similar to those of the Highway Trust Fund. When it became obvious that funding as authorized by this Act would be grossly inadequate, probably due to land price escalation, an analysis was made of needs, land cost trends, and alternative means of meeting program objectives. This study provided the basis for the 1968 amendment to provide funding at the level of $200 million per year. While the existing legislation represented a program constraint, it did not prevent the making of the analysis required to support a legislative revision.

Oil Shale.—The legislation authorizing leasing of Federally-owned oil shale lands is the Mineral Leasing Act, 30 U.S.C. 241 which imposes numerous controlling provisions such as maximum lease acreage, rentals, only one lease to any one leasee, etc. While such constraints may have a significant effect on leasing practices, they did not inhibit our capability to make a thorough analysis of all factors relating to oil shale development. In view of the major policy issues involved, a special task force, made up of the best, multi-disciplined expertise available throughout the Department, was assembled to perform the necessary analysis. Its members dug for facts and developed engineering, economic, environmental, legal, and other data as a basis for evaluation of past assumptions, development of alternative courses of action and for formulation of conclusions and recommendations. The study report served as the basis for Departmental decisions.

One of the primary functions of our PPB system is to provide Secretarial analytic capability, direction, and leadership to carry on in-depth analysis of Interior programs as a basis for evaluating their effectiveness and for development of new programs and program alternatives. We feel we can make systematic comparisons of alternatives. Past, present and prospective analysis has not identified constraints that prohibit consideration of the full array of alternatives. It is recognized that analysis performed by individuals, division, bureaus or other organizational units can reflect an agency bias. Those responsible for management of existing programs may have an understandable tendency to concentrate on justification of such programs rather than full consideration of possible alternatives. However, the present Departmental PPB system provides the means whereby both existing programs and alternative actions can be analyzed thoroughly at both the bureau and Departmental levels as a basis for recommendations and decisions.

While we have both the system and the organizational capability, there is, of course, much more analysis required than has been performed or is currently under way. Good analysis takes a great amount of time and effort. The oil shale study took approximately one year. To the extent top Departmental people were involved in that study, they were not available for other studies or program activities. The broad range of Departmental activities is such that priorities must be set so that the most urgent studies are undertaken first. Several years will be required to complete all of the major study areas already identified. Changing situations, policies, and new problems may require reanalysis as well as identify additional program areas that must be analyzed. Physical capability, not institutional or legal factors, will

be the primary constraint to timely accomplishment of necessary program analysis. We have observed that there often is a tendency to use legal or institutional factors as an excuse for not undertaking thorough analysis when, in fact, it may be that such constraints are the primary reason that analysis should be made as a basis for recommending legislative or program changes.

The usefulness of PPB depends on how the decision-maker elects to use the system. We feel that the Interior approach develops fully objective analysis, including both economic and other factors, which can provide useful and essential background for decisions.

Statement of W. Willard Wirtz, Secretary of the Department of Labor

The principal legal and institutional constraints affecting the Department of Labor's ability to perform effective expenditure policy analysis are considered under the following categories: financial and intergovernmental considerations which limit data collection necessary to sound analysis, constraints on personnel recruitment and retention, inadequate methodology for analyzing certain programs, and statutes which impose certain limitations on administrative authority and research capability.

FINANCIAL CONSIDERATIONS

It is generally accepted that the bulk of our social program expenditures should be directed toward providing services to individuals. In relation to the size of those expenditures the amount spent for economic analysis is admittedly very small. Yet the cost of conducting careful and thorough analysis would be very large. In fact, I believe that we have not yet begun to realize the magnitude of the expense involved for such economic analysis.

For example, it we were to improve our appraisal of this nation's employment problems, collecting separate data for several hundred geographic areas, and using definitions of employment that would permit quantification and analysis of the full range of employment problems, the cost would be about $10–15 million a year above our current expenditures ($8 million) for national employment statistics.

Or, if we were to fully implement a management information reporting system so that it could provide both up-to-date data on individuals and the services performed for them, the cost would be about $10 million a year.

Finally, if we were to analyze fully what we are achieving through our manpower efforts, we would need a more complete program participant followup system. This would include tracing individuals over a period of time in order to determine the effects of varying mixes of services.

Carrying out effective follow-up places heavy demands on a variety of reporters (the contractors, and the local offices of USES, etc.) who must be capable of securing reliable data. Without doubt, the effectiveness and adequacy of such follow-up systems would be directly related to the monetary investment made in them.

In short, economic analysis is costly; in and of itself that is a constraint on its use. As we become more adept at data collection, some of the costs are likely to diminish. However, as the scope of our interest and the sophistication of our efforts increase it is likely that the price of these activities will remain considerable and we will continue to struggle with the limitations imposed by this constraint.

INTERGOVERNMENTAL CONSIDERATIONS

The involvement of several levels of government in many programs sometimes gives rise to constraints on the performance of that analysis. This varies among programs.

For example, while the MDTA is implemented by the States, it is federally financed and establishes Federal standards for applicant eligibility, program operation and reporting requirements. Its Federal orientation makes possible data collection and analysis that is comparable on a nationwide basis. By comparison, the Unemployment Insurance system is both administered and primarily financed by individual States. While minimum Federal standards guide the system, the State oriented nature of the program hinders reporting procedures. For example, it is extremely difficult to obtain data on the characteristics of unemployment insurance recipients. Finally, the problem is more acute in the Workmen's Compensation system—a total State program.

CONSTRAINTS ON PERSONNEL RECRUITMENT AND RETENTION

As in any endeavor, an important factor is the quality of personnel. Analytical capability of the kind required for successful economic analysis is in short supply. However, our highest grades are, in general, given to individuals with administrative as opposed to analytical responsibility. The availability of more super-grade positions—requested by the administration and rejected by the last Congress—would facilitate the recruitment and retention of this talent.

METHODOLOGICAL CONSTRAINTS

Certain methodological problems remain at this time, especially in manpower-related fields, which hinder satisfactory economic analysis. These relate to such questions as defining skill needs in the economy, projecting these and other needs, quantifying the more intangible costs and benefits connected with manpower development, and determining the desirability of manpower development or income maintenance programs for different population groups.

STATUTORY CONSTRAINTS ON ADMINISTRATIVE AUTHORITY

There are statutes, among them the Social Security Acts, as amended, title IX, section 901, which provide strict limitation by formula on the amount of available funds which may be used for administration. Since it is from these funds that economic analysis is usually conducted, these limitations impair the effectiveness of that analysis.

STATUTORY CONSTRAINTS ON RESEARCH CAPABILITY

Our Wage and Labor Standards Administration lacks the legal authority for making research grants. As a result, many areas vital to economic analysis which lend themselves to small projects suitable for grants rather than contracts are left unexplored.

STATEMENT OF W. MARVIN WATSON, POSTMASTER GENERAL, THE POST OFFICE DEPARTMENT

Economic analysis in the Post Office Department is conducted through the PPB System and various contributing sub-systems. In the somewhat more than three years the PPB System has operated, there have been some improvements in postal management. The system has not, however, been without problems. Some of these problems were eliminated by a revision of the program structure and other improvements in the PPB System. Some of the problems which we have encountered in our efforts to improve management and to utilize economic analysis are of the type about which you inquired in your letter of November 22.

Legal and institutional constraints which have limited the full effectiveness of economic analysis in the Post Office Department may be divided into two categories. The *first* category consists of those factors which limit the performance of economic analysis. Included in this first category are the budget process, personnel limitations, and various statutory requirements. The budget process acts as a constraint since Congressional budget review is still performed along appropriation lines. Limited analytical resources in the Department must be divided among program analysis, appropriation-based analysis, and the reconciliation of the two approaches. This division of labor has resulted in a delay in the development of an information system. Personnel limitations, such as those imposed by Public Law 90–364 on headquarters and regional personnel, have hindered efforts to develop a greater capability for economic analysis.

Since many postal matters are governed by statutory requirements, the experimentation essential to complete economic analysis is often impossible. As an example, legislative action is required for the Post Office Department to conduct a regional test of combined air first-class letter rate to determine public reaction and the cost effects of a merged service.

The *second* broad category of constraints concerns constraints on implementation. In this category are those statutory and institutional constraints which limit management's ability to put into effect changes indicated by economic analysis. Many operational changes which the Post Office Department could make require the formal or informal approval of Congress. Postal financing is such a case, since operating and capital funds are appropriated separately. Trade-offs between the two accounts in the interest of lower total costs are often difficult to make because of the lack of transfer authority. Another limitation imposed by the present appropriations structure is that funds for Departmental and regional personnel are provided by appropriations

separate from field personnel. These separate appropriations militate against the flexible use of manpower which could lower overall costs.

Economic analysis has demonstrated that some postal rates are below the level at which postal out-of-pocket costs are covered. Since rates are presently established by statute, the opportunity for change is limited.

Congress also determines the total number of postal facilities to be acquired and whether they are to be built by the Government or leased. These decisions may often depend more upon the size of the postal deficit and upon competing claims for Federal funds than upon return-on-investment criteria. In many cases, this has resulted in buildings being leased which should be owned, obsolete buildings, and lack of needed mechanization.

The many statutory restrictions on the use of transportation services severely limit sound transportation planning. As an example, the Department's inability to negotiate with air carriers has hindered the development of containerization, with the result that costly manual handling procedures are still in use at most air transfer points.

The requirements of chapter 45, title 39, U.S. Code, remove compensation of employees entirely from the control of postal management. In many situations, if incentive pay arrangements were available, both the employees and the Government would benefit. Another limitation is the outdated system for compensating rural carriers. This problem is compounded by section 3339, title 39, U.S.C., which prohibits consolidation of rural routes except when a vacancy in a rural carrier position occurs.

In addition to statutory limitations such as those described above, there are also certain traditional public constraints which limit effective management. One of these is the widespread practice of viewing postmarks as a symbol of community identity. This has created many problems for the Department in its efforts to streamline mail processing in growing suburban areas. Another such constraint is the public resistence to the Department's imposing limitation on the size and shape of letters. This has hindered attempts to establish mail standards which would simplify mail processing and reduce personnel costs.

Your letter asks what changes might substantially lower the influence of these constraints on effective economic analysis. As you know, this matter has received a great deal of attention in the Post Office Department in recent months. The recommendation of the President's Commission on Postal Organization that the Department be made into a Government corporation was undoubtedly influenced by the existence of the many constraints on the present postal organization. It has been the Department's position that the question of overall postal organization is a matter for Congress to decide, and that many factors other than economic considerations must be taken into account.

A review of the many severe constraints on postal management would, however, be very useful at this time. The Department has recently worked with the House Post Office and Civil Service Committee on a bill to "review laws relating to the transportation of mail." This legislation would introduce a degree of flexibility to a key postal function now narrowly limited by statutory restrictions. Similar changes in other areas, particularly rate setting, borrowing authority, and personnel compensation, would make economic analysis a far more effective tool than is possible under present circumstances.

STATEMENT OF A. E. WEATHERBEE, ASSISTANT SECRETARY OF THE DEPARTMENT OF THE TREASURY

The legal and institutional constraints that affect and hinder economic analysis under the PPB system are reviewed by the Treasury Department under the following headings: Bureau of Customs, Internal Revenue Service, Bureau of the Mint, Bureau of Engraving and Printing, and Office of Debt Analysis.

BUREAU OF CUSTOMS

Collection of revenue program

(a) Section 1499, title 19, United States Code, requires that Customs examine at least one package of merchandise shipped on each invoice. This rule places a severe restriction on the ability of Customs to efficiently utilize its limited resources toward achieving the goals of collecting the revenue and enforcing the laws for which it is responsible. For many shipments of merchandise, sufficient evidence and experience exist, without an examination, to determine the nature of the shipment and derive the proper classification and value. Customs believes that excellent results could be obtained by adopting a system of selective in-depth examinations based on substantive characteristics of shipments, Customs experience and a random sampling. The examination of merchandise presently requires upward of 5 percent of total Customs resources. A significant savings could be effected were Customs given the leeway to determine which packages require examination. In some cases many more than one package per shipment would be examined; in other cases, no packages would be examined.

(b) Section 1498(a)(1), title 19, United States Code, requires that shipments valued in excess of $250 must be treated under the formal entry proceduces. Shipments valued at less than $250 are handled by informal entry procedures. The informal entry procedures are accomplished more rapidly and require less resources than the formal entry procedures. Better utilization of limited manpower could be made if this informal entry procedure encompassed shipments valued up to $500. Analysis has shown that the cost of administering formal entries often exceeds the duties collected therefrom where the shipments are valued under $500. Additionally, some personal use articles, which yield little or no duty, must also be treated as formal entries. Again, a more economical allocation of resources would be possible were all personal use articles allowed to be entered informally.

(c) Section 1321(a)(1), title 19, United States Code, gives the Secretary of Treasury administrative authority to establish forgiveness of additional duties due resulting from liquidation where these added duties are $3 or less. In many cases, the cost of making this determination and collection of additional duties due will exceed $3. Therefore, in the interest of economy the $3 limit should be raised.

(d) Fees for the entrance and clearance of vesels are much more costly to collect than the revenue they yield. Some fees as small as 10 cents are charged and Customs must spend many dollars in collecting and accounting for them. It is suggested that fees for entrance and clearance be made compensatory so that they at least equal the cost of accounting and collection.

1. Financing of Service programs has been incompatible with accepted program objectives

A. The conduct of an orderly planning, programing, and budgeting system depends in large part on the acceptance and funding of a defined set of program goals, and on the ability to relate those goals to program plans and program accomplishments. In recent years, however, reductions made in the resources available to the Service have been incompatible with accomplishment of the program objectives on which Service plans are based. This has damaged the relationship between goals, plans, and operational accomplishments which is essential to efficient operations.

B. This is especially well illustrated in the Service activity concerned with the audit of tax returns.

1. In this activity, program level and resource requirements are determined by the number of returns that should be audited. In turn, the number of returns to be audited should be based on accepted program goals.

2. In the first years of this decade the Service, with the support of the Bureau of the Budget and the Appropriations Committees of Congress, developed generalized criteria for audit program objectives, based on base period accomplishments and marginal yield/cost ratios.

3. Thus, the Audit coverage enforcement goal for any class of tax returns should be maintained at the same level as in the base year (fiscal 1960), whenever the marginal direct yield from future additional enforcement effort is likely to be less than six times the marginal cost.

However, when the marginal direct yield at the base-year coverage level is likely to be more than six times the marginal cost, then the Audit coverage enforcement goal for the class should be increased to the point where the additional revenue approximates six times the additional cost.

The 6 to 1 goal was selected so as to provide for the economic cost of audit activity to taxpayers as well as to government—and as recognizing that appropriations and manpower availability would be limited, even for the basic tax enforcement activity that is producing substantially more than it is spending.

4. Since then the Service's long-range program plans for the Audit activity have been based on these program objectives. However, in recent years budget resources and appropriations have been cut back sharply. One result is that the program objectives are so far from realization that they verge on being unrealistic.

5. Consequently, what was once considered to be a good, and generally acceptable, measure of program objectives for PPBS purposes has receded so far into the future that the Service is trying to find and rationalize a change in objectives. As a result the Government is not realizing as much tax revenue as has been legislated and needed to fund its policies and operate its programs. In addition, there is the danger that the general willingness of taxpayers to comply may be eroding under the lower enforcement programs.

C. Correction of this problem depends on greater stability in the funding of IRS programs at levels compatible with high taxpayer compliance levels consistent with the tax laws and the Government's tax revenue needs. This course requires effective understanding between the Legislative and Executive branches on the proper program objectives for IRS activities and the annual review and funding of tax administration activities in terms compatible with the accepted objectives. Funding of the Revenue Service and similar revenue-producing activities should also be recognized as an investment which will provide additional resources for application to other Government programs. Revenue-producing activities should not be regarded as competing for resources with these other programs.

D. The objective of a tax system is to underwrite the fiscal soundness of the policies and programs of Government. The objective of tax administration is to implement the tax system. This means determining the resources required to collect the legislated revenues and, then, optimizing the collectible revenue with the resources given the tax administrator. The present financing procedures and concepts do not encourage the most efficient tax administration operations.

II. Employment limitations

A. Federal policies and practices governing employment inevitably limit the range of alternative approaches available in meeting objectives.

1. Many of these policies themselves are designed to accomplish public objectives. However, even in these cases, their effects on agency operations should be considered, and it should be determined whether the adverse effects on operations are justified by benefits gained from the policy.

2. The limits established on year-end employment, temporary employment, overtime, et cetera, are examples. These are very real constraints in the development of alternative methods of accomplishing public objectives.

3. Uniform, nationwide salary schedules make it difficult to operate efficiently in the face of differences between geographical areas in the costs of living and in demands for professional skills.

As a result, employee turnovers are accelerated and Service ability to compete with the private sector is reduced.

B. Uneconomic difficulties imposed in the recruitment of quality manpower for Service programs place stringent limits on near-term program alternatives.

1. An example of this is the Audit program. Ideally, as discussed previously, the constraint on man-year requirements would be the desired program of audit coverage for the year (s) involved.

2. As a practical matter though, for Revenue Agents, the constraint is Service ability to recruit and absorb new employees.

—Given (1) the present availability of accounting graduates, (2) emphasis on quality hires, (3) training facilities, and (4) the problem of integrating new people into the organization, the maximum number of Revenue Agents the Service can recruit and absorb annually is 2,500 to 2,600.

—Of this number, 1,000 to 1,100 are usually needed just to cover attrition.

—Another 1,100 are required to keep pace with the growth in the number and complexity of returns being filed.
—This leaves about 400 available to improve the audit coverage of returns filed.

3. Recruiting 2,500, or even 1,100, new Revenue agents is a difficult task. The target of our recruitment program has to be the bright college graduate—and to mount an effective college recruitment program we need to begin recruitment and give firm job commitments in January and February.

4. The budgetary cycle plus delay and uncertainty in the appropriations process increase the difficulty in getting the necessary employees on the rolls. This introduces an added, unnecessary constraint on the development of programs by inhibiting Service ability to get the people we need when we need them.

C. The most drastic example of this kind of constraint is the effect in fiscal years 1969 and 1970 of P.L. 90–364, the Revenue and Expenditure Control Act of 1968. The hiring restrictions imposed by that act are particularly adverse to effective planning and efficient programming and budgeting.

1. Orderly program planning is virtually impossible. This means there are no program alternatives available. The only alternatives available are additional use of contracts to get work done, or where that is impossible, as in IRS programs, drastic reductions in program levels.

2. The effects of these constraints on the Service's fiscal year 1970 program are already evident.

(*a*) We were unable to hire at the beginning of fiscal year 1969, due to the Revenue and Expenditure Control Act. The act also has prevented us from filling in behind attrition during the year. This means at the end of fiscal year 1969 we will have accumulated a minimum of 1,900 attrition vacancies—these are vacancies that would have to be filled just to audit as many returns in fiscal year 1970 as we originally planned to audit in fiscal year 1969.

(*b*) Under the most favorable conditions this would permit hiring of only 600 additional Revenue agents to keep pace with the growing workload. This is 500 short of the 1,100 we need to hold coverage levels constant.

(*c*) The problem is compounded by our inability to recruit. Until we know where we will stand on fiscal year 1970 as far as the act is concerned we can't really make an effort to recruit—not even for the 70 percent of attrition that we would appear to be able to hire. Demands in the revenue accounting and returns processing area must be met first, and are using nearly all of the hiring authority developed in the Compliance programs, including Audit.

D. Clarification of the future place of employment limitations in planning, and of employment levels available to the Service, is necessary before meaningful planning can continue.

E. The tax revenue losses attributable to these kinds of constraint are many times the gains in reduced Service expenditures. There seem to be no good reasons for these revenue losses:

(1) The tax laws are there, just not being fully implemented.

(2) This is inequitable and in effect taxes the more honest and less daring taxpayer more than his opposite.

(3) Expenditures for tax administration are not the same as expenditures for other programs.

(4) Effective tax administration is a condition precedent to the funding of other programs.

(5) Ineffective tax administration is a source of economic and political instability.

III. Legal constraints

A. Few Government agencies are more completely bound by law and policy established outside of their immediate organization. This results from the complexity of tax law and the narrow limits within which the law can be interpreted.

1. Ideally, this arrangement should assure that Congress as an elected body bears the greatest part of the pressures for changes in law and policy, leaving the IRS relatively free to devote its energies to administration of established law and policy.

2. Operating effects can be seen as well, however, and should be considered and funded.

(*a*) Section 5202(d), Internal Revenue Code, 1954, requires, roughly, that distilled spirits must be stored on bonded premises under the A & TT custody, locked with Government locks, and at no time to be unlocked or opened, unless a Service inspector is on the premises. This provision imposes constraints which preclude the flexible use of manpower in our Permissive activity. As a means of easing these constraints, we suggest legislation which would permit the storage of spirits in Internal Revenue bond under such supervision by A & TT inspectors as they may be prescribed by regulations by the Secretary.

(*b*) Also the number of provisions which extend lack of finality to tax determinations has greatly increased, e.g., deduction and credit carryovers, income-averaging, etc.

(*c*) The basic tax law contains deadlines for completing many actions involved in revenue administration. These are constraints on the alternatives available to the Service in designing methods and procedures.

B. The provisions of the Internal Revenue Code providing Federal income tax exemption for certain classes of organizations result in a significant constraint on alternative methods of accomplishing the public objectives intended to be furthered by these exemptions. These provisions in subchapter F of chapter 1, subtitle A, of the Code, are largely statutory carryovers from the early years of the Federal income tax. The problems are well recognized by tax practitioners.

Enactment of basic statutory changes would make possible a more efficient enforcement operation. Such a revision should be grounded on a recognition of the real economic character of nonprofit corporations, and articulation of a national policy concerning what purposes or endeavors should enjoy Federal subsidy or encouragement through the exemption of specified sources of income and the deductibility of clearly defined expenditures.

IV. Budget timing

A. The timing of actions on the budget, and in recent years on the availability of funds already appropriated, have seriously limited the utility and effectiveness of program analysis.

This effect develops from damage to the relationship between program plans and program execution, from the lack of relationship between actions taken and the assumptions involved in program design, and from imbalances among and within programs that result from sudden changes.

1. In the normal budget cycle, despite the efforts of our Appropriations Committees, the final appropriation action can allow too little time to adapt to program reductions in the most effective way. This is especially true when cuts come late, or are large. Delay in appropriations action, or large cuts, can delay getting the fiscal year's operations on a firm footing.

2. The effect of cuts may be too great to fit in the context of program planning assumptions, so that the approved program cannot be related to the program plans developed on a PPB basis. This would not be so serious if the cuts were clearly based on changes in the planning assumptions, but they seldom are.

3. In the last 2 years, late cuts, below the appropriated levels approved in program and budget review, have caused emergency adjustments in program plans. The cuts were too large, and too late, to be fitted to our program plans in an efficient and economical way. The rather severe imbalances resulting among and within programs have lent an aspect of futility to program planning. They have seriously damaged relationships between the program objectives we have developed through economic analysis techniques and the programs we have been able to carry out.

B. If the full benefit of the planning, programing, and budgeting system is to be realized, action on program funding must be taken in time to be fitted to the programs in an efficient and economic way. Also, funding provided should be compatible with accomplishment of the desired program objectives, and if serious program reductions are intended they should be made explicitly after evaluation of the consequences and alternatives.

BUREAU OF THE MINT

Title 31, U.S.C.A.—"Money and finance" contains the legal basis for the existence of the Bureau of the Mint. This law further defines what business is to be conducted by the Mint, and specifies the manner in which it is to be conducted. This law places no constraints affecting the scope and effectiveness of expenditure analysis.

Effective analysis is hindered by the Budgetary System when considered in light of the Mint's great susceptibility to sudden and unexcepted changes in the demand for its products.

The Mint being a manufacturing operation lends itself ideally to the use of economic analysis in encouraging consideration of the full range of alternatives. However, under the present system of funding, this consideration becomes more of an academic exercise than an effective and valuable tool. We are tied to one alternative rather than being adaptable to the full range.

The problem is caused by the inability of the Mint, under the present Budget System, to switch funds between programs or obtain additional funds as rapidly as needed to adjust to changes in demand. It is necessary to go through the process of requesting reprograming or supplemental appropriation. These procedures, when successful, are oft times too slow to provide effective relief.

One remedy would be the enactment of legislation permitting the Mint to finance its operations through a "Mint Operating Fund." Under such a plan the fund would be financed from Mint revenues and all expenditures would be paid from this fund. This system would permit the Mint to adjust its coin production schedules quickly, when required, to provide increased quantities of coins, without having to resort to supplemental appropriations or to reprograming actions.

Revenues would be deposited in the Mint Operating Fund in such amounts as necessary to finance approved programs, and all expenditures including coinage metal, transportation of coins, and so forth, in addition to amounts now paid from appropriations, would be paid from the Fund.

Under such a system, the Mint would not be tied down to what might "seem to be" the best alternative, at the time the budget request is submitted, but which may no longer be desirable when the time comes to put the program into effect.

Proposed legislation to provide for a "Mint Operating Fund" was submitted to the 90th Congress; however, no action has been taken.

BUREAU OF ENGRAVING AND PRINTING

The Bureau engraves, prints, processes, and delivers United States currency, Treasury bonds, food coupons, postage and revenue stamps, other items of financial character, and a wide variety of miscellaneous items of an engraved nature, as may be ordered by customer agencies.

The Bureau does not receive an annual appropriation. Under the provisions of Public Law 656 of August 4, 1950, the Bureau is required to recover all costs, both direct and indirect, from its customer agencies for services requisitioned by them. That law also provides for a working capital fund method of financing Bureau operations and requires the adoption of business-type accounting and budgeting procedures. The Fund was capitalized on the basis of—

(*a*) an initial appropriation by the Congress of $3,250,000 as working cash;

(*b*) all of the receivables and the investories and other physical assets of the Bureau on hand as of the close of business June 30, 1951, exclusive of buildings occupied, land, and the unexpended balances of appropriations made to the Bureau; and

(*c*) assumption of all of the liabilities of the Bureau as of the close of June 30, 1951.

Financing Bureau operations with the assistance of a revolving fund provides a substantial degree of flexibility and freedom from constraints which might affect the scope and effectiveness of expenditure policy. The size of the Fund and the amount of money obtainable from depreciation of fixed assets determines how fast the Bureau can proceed with technological improvements. Greater flexibility could be obtained through an increase in the amount of the Fund. Availability

of additional money for the purpose of buying fixed assets would speed the modernization of the Bureau's printing and processing operations with an attendant result of substantial savings at an earlier date.

A current limitation which has had an effect on the Bureau's planning is the employment restrictions imposed by Section 201 of the Revenue and Expenditure Control Act of 1968 (P.L. 90–364).

A direct relationship exists between the Bureau's manpower requirements and the positive work program submitted by customer agencies. In the light of the substantial increase in production requirements for fiscal year 1969, it is apparent that the workload cannot realistically be accomplished under the personnel restrictions imposed by the provisions of Public Law 90–364, without resorting to substantial amounts of overtime during the year.

For the past three fiscal years, the Bureau has been required to work unusual amounts of overtime as a result of previous ceilings placed on its level of employment as well as difficulties encountered in hiring the full complement of personnel needed to meet expanding work programs. Continuation of this condition could create a serious morale problem among the employees. On the other hand, failure on the part of the Bureau to meet the requirements of the Nation for currency, stamps, etc, would certainly place the Department, as well as customer agencies and the Government as a whole, in an embarrassing situation.

In submitting the current Program Memoranda and Program and Financial Statements, required under the Planning, Programming and Budgetary System, changes were made to reflect compliance with the law imposing personnel ceilings and Department policy restricting expenditures. However, in making plans for fiscal year 1970 and the following fiscal years, it was assumed that these limitations would be removed.

We trust that these short-time limitations are indeed temporary and will not be a hindrance to continuing improvement of Bureau operations. Cost benefit analyses show that such improvements would save money in the long run for the Government.

OFFICE OF DEBT ANALYSIS: DEBT MANAGEMENT

In the area of debt management policy there are two principal legal constraints which discourage consideration of the full range of alternatives in developing policy to attain public objectives—the 4¼-percent statutory interest rate ceiling on Treasury debt issues of more than 7 years maturity (31 U.S.C. 752), and the statutory limit on the dollar amount of public debt obligations outstanding (31 U.S.C. 757b).

In addition to the operating problems created by these two statutory limitations in the day-to-day management of the public debt, the existence of these limitations may discourage consideration of direct Treasury financing in the planning of some Government programs which can be financed through the sale of obligations in the market by other Federal agencies or through the use of various loan guarantee or insurance techniques.

Since loan guarantees, loan sales, and Federal agency borrowings are not subject to either the statutory debt limit or the 4¼-percent ceiling, there may be a tendency to overuse these techniques especially

at times when the amount of Federal debt is close to the statutory ceiling and at times when market interest rates exceed 4¼ percent and the Treasury is thus limited to relatively short-term financing. For example, because of the 4¼ percent ceiling the Treasury has not been able to issue long-term securities since May of 1965.

This is not to say that direct Treasury financing is always, or generally, more appropriate than the various types of agency market activities or loan guarantee programs. These programs must be assessed in a broader context. The fact remains, however, that in some cases these programs are looked upon as the only feasible means of financing activities which, were it not for the statutory limitations on Treasury debt and interest rates, might be more appropriately financed by the Treasury at a lower cost to the Government.

Similiarly, any statutory limitations on Federal budget expenditures, including expenditures for direct loans, could create pressures for the adoption of financing techniques and loan guarantee programs which might otherwise be financed by the Treasury at less cost to the Government.

STATEMENT OF WILLIAM F. CASSIDY, CHIEF OF ENGINEERS OF THE DEPARTMENT OF THE ARMY

It is my view that by enacting the Water Resources Planning Act of 1965 the Congress has enabled the agencies of the Executive Branch to formulate truly comprehensive plans, and in doing so to give consideration to all alternative means of utilizing, developing and conserving the Nation's water resources. The broad plans formulated under the legislation are to be submitted to Congress and that body may authorize the carrying out of such plans, in whole or in part. From the standpoint of legislation, therefore, there are no longer any insurmountable obstacles in the way of achieving the kind of comprehensive and coordinated river basin plans first urged by the leaders of the Conservation Crusade some 60 years ago. This does not mean, of course, that the Congress will authorized the carrying out of every feature of every comprehensive plan placed before it by the Executive Branch. In a particular basin, or at a particular time, the Congress may be willing to authorize, for example, a flood control reservoir but not a reservoir the primary purpose of which would be to provide recreational opportunities. But this depends upon the attitude of the Congress, and does not indicate a deficiency in the legislation which authorizes the preparation of plans. In other words, there appears to be no need for legislation to authorize consideration of alternatives.

Another important consequence of the enactment of the Water Resources Planning Act was the delegation to the Executive Branch of responsibility for establishing the methods to be used in the economic evaluation of water projects and programs. The Water Resources Council may, under that Act, establish "principles, standards and procedures . . . for the . . . evaluation of Federal water and related land resources projects." The Council is now engaged in reviewing the Presidential standards issued in 1962 and printed as Senate Document No. 97, 87th Congress. At such time as this review has been completed the Council will issue regulations which will take the place of that document. The Council has already proposed a new

standard for determining the discount rate to be used in economic evaluation. It may be said, therefore, that there are no real legislative constraints upon the Executive Branch insofar as sound economic analysis is concerned.

You transmitted with your letter a statement, submitted to the Committee by the Department of Transportation, which classifies as a constraint upon sound economic analysis the effect of Section 7(a) of that Department's organic Act. Strictly speaking Section 7(a) does not, in my opinion constitute a legislative directive prohibiting sound economic analysis. This view is based, in part, upon the fact that economics is not a matter of policy, and hence is no more subject to constraint by legislation than are natural phenomena. It has been the view of the Corps of Engineers that the real purpose of Section 7(a) was to establish a "standard" for use in determining whether a particular waterway does or does not, qualify for inclusion in the National transportation system. Such a standard expresses a policy of Congress. It seems probable, therefore, that what the Department of Transportation intended to convey to the Committee was the fact that application of the standards of section 7(a) will not necessarily assure the selection of that alternative which produces the greatest net economic benefit. This is quite different than saying that section 7(a) constitutes a constraint upon sound economic evaluation by the Executive Branch. This section requires the Executive Branch to inform the Congress whether a contemplated waterway would, or would not, qualify under the standards set by that section. It does not, in my opinion, prevent the Executive Branch from applying any other tests it may consider desirable, and informing the Congress of the results of these tests.

As will be evident from all the foregoing, present laws do not impose any important direct constraints upon the Executive Branch in giving consideration to alternatives in water resource development, or in making sound economic evaluations of water projects and programs. Moreover, it is clear that the Executive Branch has been given adequate authority administratively to make these changes required to bring both planning and evaluation of water resources activities into consonance with generally accepted procedures.

There is, however, one way in which existing laws *indirectly* exercise an undesirable constraining effect upon the efforts of the Executive Branch to achieve the goal of optimum utilization of the Nation's resources. This effect stems from the fact that the various cost-sharing policies which Congress has established over a long period of years are based upon rather widely divergent concepts of the degree to which the Federal Government should assume responsibility for the various aspects of resource development. As a result of these basic divergencies some Federal programs are more attractive to groups of prospective beneficiaries than are others, and this sometimes makes it impossible to gain acceptance of that combination of projects and purposes which would maximize the net benefits resulting from the development of a river basin. The inconsistencies in the doctrines upon which Federal cost-sharing policies have been based over the years have long been recognized, and the Water Resources Council is giving intensive consideration to the development of a consistent system of cost-sharing policies. The results of this work will undoubtedly be made available to

the Congress as a basis for cost-sharing legislation in consonance with the basic concept of the Water Resources Planning Act.

All of the problems touched upon above will, of course, be subjected to further consideration by the recently established National Water Commission, and this body will undoubtedly call to the attention of the Congress any modifications of existing legislation which may be required to bring existing legislation into consonance with any recommendations it may make. Since both the National Water Commission and the Water Resources Council have very broad authority to determine the need for either legislative or administrative changes, and since the latter organization has authority to make administrative changes, there is no need, at the present time, for legislation bearing upon either alternatives or economic analysis.

STATEMENT OF WILLIAM S. GAUD, ADMINISTRATOR, AGENCY FOR INTERNATIONAL DEVELOPMENT, DEPARTMENT OF STATE

Within the broad framework of foreign policy, AID operates primarily on the basis of country programing. We try to focus our analysis on the circumstances of individual countries, and to support aid activities in each country which contribute most to U.S. objectives there. The right program of activities for one country may be quite different from the right program for its neighbor. The constraints listed in the attachment are worrisome essentially because they restrict effective country programing.

The problems raised by these constraints should be kept in perspective. AID country programs can be roughly divided into two categories—those oriented toward long-term development and those that address more immediate political concerns, such as security and stability. In dollar terms, programs emphasizing development account for about three quarters of our total effort in fiscal year 1969. Programs emphasizing security and stability account for the remaining quarter, the great bulk of which is for Vietnam. Economic analysis, which is central to the PPB system, is much more relevant to the former programs than to the latter, although several of the specific items mentioned in the attachment apply to all of our programs.

Moreover, there are intrinsic complications in applying the PPB approach even in development emphasis countries. The objectives at which we aim can be identified in general terms without too much difficulty, but defining the specific "outputs"in measurable terms useful for analysis is not an easy task. Analyzing just how "inputs" of U.S. foreign aid resources—which are marginal in relation to the resources of the recipient country—affect the outputs and objectives is even more complex. There are imponderables, like the host government's political will to develop, which can only be assessed by insight and sound judgment. And of course political circumstances can change rapidly.

Nonetheless, how well we do our job of country programing is hampered by legislative constraints. Many have reasonable purposes, but they also have significant costs in limiting the effectiveness of AID program analysis and the programs themselves.

Listed below is a brief description of a number of legislative constraints to economic analysis and consideration of alternatives that exist in the Foreign Assistance Act and other acts bearing on our programs.

1. *Time Constraints on Planning*—(*a*) *Foreign aid authorizations* are on a one-year basis. This constrains the Agency's ability to plan ahead over a meaningful time frame, which for development purposes should really be 3 to 5 years. In addition, AID typically receives its appropriations 4 to 6 months after the start of the fiscal year, largely because authorizing legislation must be passed before appropriation action may be taken. For AID, the authorization process usually takes the bulk of the Congressional session. It would certainly contribute to effective analysis and sound management if the authorization—at a minimum—covered the two years of each House term. Respite from long-drawn-out yearly fights over the AID authorization would give top Agency management more time to spend on meaningful planning, programing, and operations.

(*b*) *The Foreign Assistance Appropriation Act* requires that all new technical assistance and international organizations projects be justified to the Congress before being implemented. But particularly in foreign assistance, where the contributions and responses of our partners are concerned, it is important to have the flexibility to respond to changes and new opportunities as they occur. AID presents to the Congress an "illustrative budget," although over the years we have moved in the direction of providing specific project data to the Congress in advance. It is a significant constraint, however, to require 100 percent specificity, making it increasingly difficult for AID to respond to new developments. This constraint does not limit analysis *per se*, but prevents prompt implementation of new alternatives.

2. *Constraints Against Actual Provision of Assistance to a Country*—(*a*) *Sections 201(b), 211 and 401* of the FAA limit the number of countries which may receive certain types of aid. In fiscal year 1969, only 20 countries (outside Latin America) may receive development loans; only 40 countries (outside Latin America) may receive technical assistance (other than small amounts for self-help activities); and only 12 countries may receive supporting assistance. While we are in favor of the principle of aid concentration, and have at times not reached the above limits, we believe it is better to use analysis to arrive at the specific number of countries to be aided rather than to set arbitrary limits through legislation.

(*b*) *Section 620* of the FAA contains a series of constraints designed to achieve desirable purposes other than development. We are particularly concerned about requirements that no assistance can be provided, or that consideration must be given to terminating aid, to a country:

If the country furnishes, or permits ships or aircraft under its registry to transport, materials to Cuba or North Vietnam—620(a) and 620(n).

If its government owes a debt to a U.S. citizen—620(c).

If its government has nationalized property of a U.S. citizen without compensation—620(e).

If its government uses U.S. aid to compensate owners for expropriated property—620(g).

If the country fails to prevent, or pay the United States for, mob damage to U.S. property—620(j).

If its government does not sign an investment guaranty agreement—620(l).

If the country seizes or imposes sanctions against a U.S. fishing vessel—620(o).

If its government diverts resources to unnecessary military expenditures to a degree which materially interferes with its development—620(s).

If a country is delinquent with respect to its U.N. obligations—620(u).

However worthy these nondevelopment purposes may be, our experience is that such legislative requirements can disrupt effective development cooperation with host governments, and that the threat to reduce or eliminate U.S. economic assistance is rarely an effective way to attain the desired objectives. From a PPBS standpoint, these contraints generally come from events outside the planning process.

3. *Expenditures for Sophisticated Weapons*—Sections 620(v) of the FAA requires that economic assistance must be withheld in an amount equal to the amount spent by any less developed country for the acquisition of sophisticated weapons from any country unless the President determines that the acquisition is important to the national security of the United States.

This provision raises difficult problems of interpretation, information gathering, and implementation. More fundamentally, it may require reduction (which may, in some cases, be elimination) of assistance of great importance and utility for overall U.S. interests, including technical assistance activities of small cost that have little if any leverage on the actions of the government purchasing the weapons. By assigning overriding priority to punitive deductions, the amendment limits our ability to take actions that may best serve U.S. interests. The legally required deductions may in fact interfere with more effective but less formal means to restrain arms races in less developed countries.

4. *Constraints Against Flexible Use of Total Available Appropriations*—(a) Earmarking funds for a particular purpose, regardless of whether the results of country programming would call for more or less for that purpose, almost inevitably hampers effective management. For example, *Section 292* of the Foreign Assistance Act (FAA) earmarks $50 million of otherwise available funds for use only in programs relating to population growth. While AID strongly supports the intent of this earmarking, the appropriate number and type of population activities, and hence their total cost, varies from one fiscal year to the next. The earmarking of a specific amount can cause overprograming in population with a corresponding deleterious effect on other fields. Eliminating this earmarking could improve project management, without affecting the Agency's top priority for population activities.

b. *Sections 205 and 251(h)* of the FAA permit use of aid appropriations to assist country development by channeling some funds through selected international financial institutions. This alternative channel for providing aid could be very valuable, for example, where

bilateral aid risked too much U.S. political involvement, or where a regional project required the leadership of an international body. Although this authority is provided in the Foreign Assistance Act, the Foreign Assistance Appropriation Act contains a provision denying its use. If this constraint were lifted, AID could consider multilateral as well as bilateral channels in adjusting our assistance to country circumstances.

5. *Constraints on Country Using U.S. Assistance at Least Cost*—(*a*) *Section 901(b)* of the Merchant Marine Act of 1936 requires that at least 50 percent of the value of AID-financed commodities be carried on U.S. ships. This constrains sound economic analysis and limits full consideration of alternatives in a number of ways. The additional shipping cost of using U.S. bottoms, of course, reduces the net benefit of U.S. aid. The varying cost effect on different U.S. commodities also tends to distort the pattern of purchases which would otherwise take place. Low cost bulk exports, for example, such as fertilizers or coke, are discriminated against partly because of the 50–50 freight differential. The 50–50 restriction is one of the factors that tends to make U.S. goods noncompetitive, which in turn raises the costs of additional policies and interferes with future commercial exports. Since the purpose of section 901(b) is to assist the U.S. shipping industry, the charge should really be to that purpose and not to foreign aid.

(*b*) *Section 201(d)* of the FAA provides for minimum interest rates at which development funds may be loaned. In 1961, development loans repayable in dollars had a minimum interest rate of three-fourths percent, with a maximum maturity of 40 years including a maximum grace period of 10 years. Since then, Congress has raised the minimum interest rate four times, and it is now 2 percent during the grace period and 3 percent thereafter. These constraints have been imposed at a time when the debt servicing burdens of developing countries as a group have been growing rapidly, and when many individual countries are clearly in more difficult straits than the average. An across-the-board interest floor is a constraint on the type of aid which a country needs in the light of its debt servicing situation, and thereby limits the consideration of alternatives which may best aid a country's development. Greater flexibility for setting the terms of aid to fit individual country situations could be a significant improvement in our negotiations both with the recipient country and with other donor nations.

(*c*) *Section 620(d)* of the FAA prohibits a development loan for construction of a productive enterprise competing with U.S. enterprise unless the country has agreed to prevent export to the United States of more than 20 percent of the project's annual production during the life of the loan. Effective assistance to a developing country includes the financing of productive industrial enterprises, particularly in fields where the country has a comparative cost advantage and can earn additional foreign exchange through expanding exports. The effect of this provision is to limit our choice in the selection of industries to be aided, hitting fields that are likely to be most important to the recipient country in becoming self-sustaining without aid.

(*d*) *Section 601(d)* of the FAA strongly encourages AID to use U.S. engineering and professional services. This constraint conflicts

with a central goal of our technical assistance programs—to improve local businesses and institutions so that the country can become economically and technically self sufficient.

(*e*) *Section 604* of the FAA requires that commodity procurement should be from U.S. sources. While the U.S. balance-of-payments problem makes this requirement unavoidable, its real costs for the aid program should be recognized.

6. *Constraints on the Use of U.S. Agricultural Commodities*—AID participates with the Department of Agriculture in the planning for and administration of concession sales and grants of U.S. agricultural commodities for use in country development programs. These commodities are provided under the Agricultural Trade Development and Assistance Act (P.L. 480). Net flows of assistance in this form worldwide have in recent years been over half of net flows of AID dollar assistance, and P.L. 480 commodities have been increasingly used to induce critical measures of agricultural self-help in recipient countries. Effective country programing of P.L. 480 sales and grants is therefore extremely important.

A number of the constraints referred to for AID dollar programs also apply to P.L. 480—for example, the 50–50 shipping requirement and the difficulty in fitting credit terms to the debt-servicing capacity of individual countries. In addition, the complex nature of P.L. 480 legislation and difficulties in interpreting specific provisions in sales agreements sometimes delay the decisionmaking process in determining the terms of agreements and the types and amounts of commodities to be provided.

Statement of Glenn T. Seaborg, Chairman of the Atomic Energy Commission

The Commission has long made use of economic analysis, where applicable, in determining courses of action to be recommended or followed. Such analyses have given full consideration to competing alternatives. With the advent of PPB and its emphasis on systems analysis, we have, of course, increased our activities somewhat in this area. This is true particularly with respect to responses to specific requests for analytical studies made by the Bureau of the Budget.

In all of our analyses, however, we do not consider that we have been hindered or constrained as a result of any legally or institutionally imposed limitations. Therefore, we do not feel that any legislative, administrative, or other change need be made, insofar as AEC programs are concerned, in the interest of lessening the influence of constraints on effective economic analysis.

Statement of Lawson B. Knott, Jr. Administrator of the General Services Administration

The six major program categories with GSA are (1) facilities, (2) supply services, (3) other property management and disposal services, (4) transportation and communication services, (5) record services, and (6) agency direction and support services.

The General Services Administration was established to provide for the Government an economical and efficient system for the management of the property, including related services, and records of all Federal agencies. Accordingly, the level of activity by GSA is dependent upon the quantity of supplies, space, equipment, communications and other services required by Federal agencies. With respect to many of these activities funding levels are imposed upon GSA independent of consideration of the quantity of services the agency is called upon to provide for expanding or new Federal programs of other agencies as authorized and funded by Congress. Exceptions to the above include activities financed through revolving funds such as the general supply fund, Federal telecommunications fund, buildings management fund, and automatic data processing fund, established under sections 109, 110, 210, and 111, respectively, of the Federal Property and Administrative Services Act of 1949, as amended (63 Stat. 377). Employment levels are also imposed on GSA activities (including activities financed under revolving funds) independent of the workload created for GSA by activities of other agencies.

GSA has nonetheless performed its assigned statutory responsibilities in a manner generally responsive to the needs of the Federal establishment. However, the lack of adequate funding and satisfactory funding procedures has proven costly over a long term. We recognize, of course, that with respect to adequate funding, appropriations for GSA activities must be considered in terms of the entire Federal program, including the relative priorities assigned to national needs and goals.

The above constraints principally affect the facilities program administered by the Public Buildings Service. GSA has experienced fluctuations of more than $125 million between annual levels of public buildings construction funding since the enactment of the Public Buildings Act of 1959. Due to these wide fluctuations and low funding levels, GSA has encountered great difficulty in carrying out an effective program to meet Federal office space requirements through construction of Federal buildings. Since 1959, one-third of the amount estimated at that time to be needed to meet Federal office space demands has been appropriated. Federal employment and activity, as well as construction costs, have continued to rise, and it is now conservatively estimated that 30 years would be required, at the past annual funding rate, to catch up with the space needs of Federal agencies as currently projected.

A major consequence of our inability to meet space needs through new construction is that it has been necessary to provide space through the more costly leasing of privately owned buildings. Furthermore, commercial buildings seldom provide space in blocks of adequate size to accommodate entire agencies and, as a result, Federal activities have been scattered and potential savings available from consolidations have not been reached. The current method of direct appropriation for public building construction has not only resulted in inadequate funding but in one which impairs proper economic analysis of not only GSA's program but also programs of other agencies. The cost of space significantly affects the cost of the program for which the space is provided; however, this cost is not reflected in program budgets except in the case of leased space and then only for the first full fiscal year of

occupancy. Thus, in subsequent years the program budget does not accurately reflect the total cost of the program, a contradiction to the performance budgetary concept under which total program costs are reflected in the cost accounts of the program agency.

A second major consequence of GSA's inability to meet space needs through the current method of public buildings financing is the growing tendency of other Federal agencies to seek and obtain independent authorization and funds for constructing their own facilities. Any fragmentation of responsibility for building construction impairs the effectiveness of a planned public building program since the appropriation is often made without regard to more urgently needed construction projects which remain unfunded.

We believe that consideration should be given to legislation which would expand the present buildings management fund to provide for the financing of all real property management services rendered by the Public Buildings Service. The fund could include receipts accruing from "user" rates to be charged occupant agencies for all space provided by GSA. Agencies would then be required to obtain appropriations necessary to reimburse GSA for public buildings services performed. In this manner, the costs of facilities would be reflected in the cost accounts of the program agency. Monies covered into the fund would be available for the acquisition, construction, operation, maintenance, and alteration of public buildings. Legislation could be designed to assure continued congressional control over the annual level of the programs by providing for congressional authorization for projects estimated to exceed a minimum dollar amount. While such legislation, without a corresponding increase in appropriations for construction, would not in itself result immediately in an increase in construction of badly needed public buildings, it would provide more consistent level of funding for public building purposes thus assuring sounder economic analysis and planning. In the long term, we believe that it is a more realistic method of financing space needs for public agencies while at the same time assuring proper congressional control over expenditures and programs.

Statement of Leland J. Haworth, Director of the National Science Foundation

In general, the PPB effort has been useful as a formal mechanism to reexamine objectives, the alternatives available to meet these objectives, and their implication for resources. We have found the PPB process useful to crystallize objectives and to elucidate alternative possibilities. However, we have not been able to utilize economic analyses extensively for the reasons outlined below.

Our difficulties with economic analysis are due in part to limitations in the state of the art rather than to legislative or institutional constraints. The central responsibilities of the National Science Foundation are support for scientific research and education in the sciences. Unlike other agencies whose missions relate to categorical problems and whose programs produce more objective outputs, our outputs are primarily: (a) newly gained scientific knowledge; (b) men and women educated in science; and (c) institutions better equipped and

staffed to conduct research and education in the sciences. As indicated, these do not lend themselves easily to economic analysis. Also, it is too early to attempt economic analyses of the effects of any programs entered into to carry out the new applied science authority of the Foundation. Economic analysis requires systematic comparison and alignment of resources with expected outputs, together with appropriate measures of effectiveness. The outputs of research and education, and the associated measures of effectiveness are complex and it has not been possible thus far to quantify them realistically for purposes of such analyses. Outputs are not adequately represented by counting the first knowledge products derived from the research supported or by numbers of science degrees attained. The great importance of quality in research and education, and the difficulty with its measurement, contribute to the complexity of output study. Finally, the major return on the research and science education investment occurs in the follow-on use of the research and in the professional careers of those educated. These facets are not predictable by existing analysis, forecasting or prediction system—again a state of the art weakness rather than external constraints.

Further difficulties and challenge to economic analysis of support for research and science education are presented by the unique role of the Foundation. It, among all Federal agencies, has been given by law the broad responsibility for promotion of science and science education in the United States. Many Federal agencies, with somewhat differing missions and policies, contribute, together with other sectors, to the support of science and related activities. NSF provides only a small fraction of the Federal funds allocated to science—17 percent of Federal academic science obligations and about 12 percent of Federal basic research funds. Thus, the alternatives open to the Foundation, in its efforts to achieve its broad objective of assuring the future health and progress of U.S. science, are very dependent on the actions of other agencies, on the timing of the completion of their program plans in the yearly budget cycle, and on overall budgetary limitations. To keep track of this, as part of its overall responsibility, NSF carries out a program of collection of data on different aspects of manpower in and funding of science, for the general use of all agencies.

At present the guidance and review of programs is based upon maintaining the best knowledge possible of the state of national science activity, of its trends, of its performers and performance and of its support. At present decisions are made primarily on the basis of evaluation of the health and viability of the enterprise rather than on the basis of the predicted economic value of the outputs. As our data and experience build and as they can be correlated with the outcome of past decisions, hopefully new evaluation techniques will emerge through the application of new methodologies now available, such as operations research, systems analysis and computer-based mathematics. Hopefully the new techniques will contain the necessary transformations to align the criteria for research support and science education needs with economic output values. At present it is a possibility, not yet a sure goal. The Foundation will endeavor to enhance the former and pursue the latter. The difficulties present challenges which may inspire advances in the state of the art of economic and policy

analyses for science support decision purposes. The Foundation is anxious to take advantage of any such advances and to encourage advances in the state of the art. As for legal constraints, we feel that none exist at this time which handicap us in our evaluation efforts.

STATEMENT OF BERTRAND M. HARDING, ACTING DIRECTOR OF THE OFFICE OF ECONOMIC OPPORTUNITY

We have from the very beginning enjoyed a substantial amount of freedom to engage in economic analysis without legislative, institutional, or other constraints. Our ability to define problems confronting the poor and to move summarily to develop alternative solutions to these problems can be attributed both to the general lack of constraints encountered in the Economic Opportunity Act and to a deliberate attempt to encourage such analysis, starting with the first Director of OEO. The only significant constraint on this freedom is found in section 201(b) of the Economic Opportunity Act. This section prevents OEO from conducting experimental and demonstration projects that might encourage rural poor to migrate to urban areas. It thereby prevents us from giving full consideration to a major policy alternative. In the present state of knowledge, it seems to us to be premature to foreclose consideration of the possibility that migration to urban areas may be the best hope for some groups among the rural poor.

STATEMENT OF JACK VAUGHN, DIRECTOR OF THE PEACE CORPS

Ideally, economic analysis should provide the answers to programing problems. Since each volunteer goes into the field for approximately two years, a decision must be made every 24 months on whether or not to continue a program. But economic analysis alone would not answer this question. Our criteria for making these decisions cover a broad range of considerations which includes economic analysis only as one step in the decisionmaking process.

Our main criteria for program analysis are built around the premise that an opportunity for Volunteers and Host Country Nationals to come to know and understand one another is a prerequisite to effective job performance. In addition, the program must relate to one or more of the Host Country's most critical needs as identified by the Host Country itself and manifested by a Host Country effort and structure to which the Peace Corps programs can relate and within which the Peace Corps Volunteers can function in close association with Host Country Nationals. In a very real sense, the Peace Corps has no programs of its own; rather Volunteers support Host Country programs. Because Peace Corps Volunteers are integrated into Host Country programs, it is often difficult to separate any unique Peace Corps output of a project. This represents a constraint which is common to all of our programs.

I would note also that our programs have purposes that go beyond pure economic considerations. Accordingly, it is essential that we con-

sider other than strictly economic factors in weighing the merits of the full range of alternative programs that are available to us.

We are, nevertheless, incorporating economic analysis into our consideration of programing alternatives. With the implementation of our PPB system, we are making progress toward the goal of acquiring more cost-effectiveness information for use in program planning. The changes that have been and must still be made to facilitate this effort are administrative in nature rather than legal.

The first of these changes, which is currently being implemented, will provide cost information at the program level for use by decision-makers. Because of the wide variety and large number of projects which we now have overseas, it has become essential to have an automated system for determining project costs. Consequently, we recently signed a contract with Price Waterhouse and Company to assist us in developing efficient and economical methods for acquiring this information. We believe that the information should become available during fiscal year 1970.

The second major change that must me made before cost-effectiveness information can be widely useful is to develop the capability for systematic collection of relevant measurements of output. Our constraints here turn around the problems of defining the relevant outputs of individual projects and conducting measurement procedures at a reasonable cost. In addition, some of our program outputs such as changes in institutions, habits, and attitudes, while measurable, may be so politically sensitive that it would be difficult to carry out the measurement process.

While these are serious and very real limitations, we do not consider them to be insurmountable. We are proceeding with plans to implement procedures which will ultimately lead to the collection of effectiveness data on a broad scale. These procedures involve overseas research as well as careful planning by Host Country Nationals, volunteers, and our staff in the field and in Washington. We have not developed a timetable for the completion of this effort because we must consider separately the problems of each country and each project.

However, we are confident that our efforts will eventually lead to the solutions we desire. Once this is achieved, we will be able to conduct measurements of the effectiveness of our programs, utilizing economic analysis to the extent appropriate, within the context of our policy and criteria for Peace Corps programing.

Statement of William J. Driver, Administrator of the Veterans' Administration

The statutes (now generally codified in title 38, United States Code) specify terms of eligibility, periods of entitlement, and rates of veteran benefit. Technically, the statutes do not constrain the manner and scope of possible economic analysis. However, the implementation of the results of expenditure policy analysis could require legislative changes.

I have no recommendations, from the standpoint of economic analysis, for any change in the mission of the Veterans' Administration or in

the statutes defining and limiting the various benefits of this mission. Changes in eligibility, rates, and periods of entitlement take place over the years but none affect the basic philosophy or legislative intent of the veterans program. As in any benefit program these are continuously reviewed by the Executive Branch and the Congress.

Statement of Rosel H. Hyde, Chairman of the Federal Communications Commission

The Commission, recognizing the value of a Planning-Programing-Budgeting System in managing its resources, has taken several steps toward implementing such a system within the FCC. We have established an output oriented program structure and recently submitted our first Program and Financial Plan to the Bureau of the Budget. Currently, we are taking steps to improve our supporting information systems in order to provide more accurate and useful program information.

We have conducted studies on specific programs, considering alternate ways to implement these programs administratively, at the least cost. However, thus far we have not utilized economic analysis to arrive at decisions on expenditure programs, at least not the type of cost/benefit analysis discussed in much of the current literature on PPB.

Certainly, in a regulatory agency such as the FCC, major policy decisions arrived at through the rule making process may have a substantial economic impact on the public. Consequently, our rule making proceedings may involve extensive analysis of the economic consequences of alternative proposals. In this regard, the appropriation process which determines the amount of resources to be used, is the major constraint on the extent of economic analysis undertaken.

In the future, subject to the availability of manpower, we expect to make greater use of economic analysis in our programming and budgeting process. For example, the need to modernize and expand our field monitoring facilities would be a good subject for this type of analysis since considerable funds (relative to our total budget) might be required if such a program is initiated. We do not feel that there are any legal or institutional constraints to this type of expenditure policy analysis in any of our major programs, and thus do not recommend any legislative, administrative or other changes at this time.

Statement of Lee C. White, Chairman of the Federal Power Commission

The Federal Power Commission, as a regulatory agency, is not involved directly in the disbursement of funds to others for public purposes. It does not administer construction projects, grants, and similar activities financed by public appropriations. There are no legal or institutional constraints analogous to those restricting the Department of Transportation that may hinder or impair the sound economic analysis of our expenditures.

On the contrary, several acts of Congress give the Commission broad authority to investigate, analyze, and evaluate the electric power and natural gas industries within the jurisdiction of the Commission: the Federal Power Act, the Natural Gas Act, and several flood control acts.

For example, under section 4 of the Federal Power Act, as amended, 16 U.S.C. 791a–825r, the Commission is authorized and empowered, among other things, to make investigations and to collect and record data concerning the location, capacity, development costs, relation to markets of power sites, and the fair value of such power, to the extent the Commission may deem necessary or useful for the purposes of the Act.

The Commission, under section 206(a) of the same Act, may determine a just and reasonable rate, charge, classification, rule, regulation, practice, or contract to be thereafter observed and in force, and fix the same by order.

The Water Resources Planning Act, Public Law 89–80, establishing the Water Resources Council of which the Chairman of the Federal Power Commission is a member provides that the Council maintain a continuing study of the adequacies of water supplies to meet the requirements in each water resource region in the United States. It directs further that the Council maintain a continuing study of the relationship between regional or river basin plans and programs and requirements of larger regions of the Nation; and appraise the adequacy of existing and proposed policies and programs to meet such requirements.

The above illustrates that the various legislative acts are rather broad and give the Commission wide latitude in developing the principles, methods, standards, and alternatives to carry out its charge.

STATEMENT OF HOWARD J. SAMUELS, ADMINISTRATOR OF THE SMALL BUSINESS ADMINISTRATION

A careful review of our programs indicates the following constraints affect sound economic analysis:

Several different statutory provisions govern the rate of interest charged on direct loans placed by SBA. Our regular business loans (authorized by section 7(a) carry a maximum interest rate of 5½ percent. Economic Opportunity Loans, Trade Adjustment Loans, and Displaced Business Loans are based on formulae tied to the yield on Treasury obligations. The rate on Economic Opportunity Loans is currently 5⅝ percent. Trade Adjustment Loan rates are set by law as not less than the greater of (1) 4 percent per annum or (2) a rate determined annually by the Secretary of the Treasury based on yields on comparable maturities. Based on the maximum authorized maturity of 25 years the rate would be 5¼ percent for fiscal year 1969; Displaced Business Loans are 4¾ percent; State Development Company (501) Loans carry a variable rate (not less than prime bank rate nor more than 6½ percent); local Development Company (502) Loans carry a maximum rate of 5½ percent. The statute provides that 501 and 502 loans shall carry rates "as fixed by the Administration."

We believe that greater uniformity of rates among the different programs, with a broader range of discretion allowed to the Administrator in setting specific rates, would make for more effective use of resources. We are considering proposing legislation to this effect.

STATEMENT OF AUBREY J. WAGNER, CHAIRMAN OF THE TENNESSEE VALLEY AUTHORITY

We think it is fortunate that in the initial legislation creating TVA and in early amendments of the TVA Act Congress provided TVA a high degree of flexibility which generally has permitted it to consider and choose freely among the various alternative methods of achieving the objectives of the TVA Act. We believe this flexibility has enabled TVA to do an effective job in encouraging and assisting in the development of the resources of the Tennessee Valley region.

We are sure your committee is aware that the greatest deterrent to efficiency in the case of appropriation-financed projects lies in budgetary restrictions resulting in construction stretch-outs. We plan such projects carefully on the basis of the most efficient schedules, which will, in turn, produce the lowest possible cost and largest possible excess of benefits over costs. Budget-produced stretch-outs disrupt these plans and inevitably result in less efficient construction, inflated costs, and a reduction of the ratio of realized benefits to cost.

With respect to specific legislation presenting problems, section 15d of the TVA Act, which was added in 1959 to authorize issuance by TVA of revenue bonds to assist in financing its power program, imposed certain legal constraints which do restrict available alternatives in two respects.

First, section 15d fixes a ceiling on the amount of power revenue bonds that TVA may have outstanding at any time. This ceiling, which was fixed at $750 million in 1959 and raised to $1,750 million by passage of an amendment in 1966, has not so far been a serious handicap. However, long-range plans for increasing the capacity of our power system to meet the growing power needs of the areas served by TVA must be conditioned upon further increases in the ceiling by legislative action at the appropriate times. Any delay in raising the ceiling when required would seriously jeopardize effective planning and could, of course, delay or prevent necessary expansions of our power system capacity and place at least a temporary ceiling on the growth of the area which TVA supplies with power. We believe that TVA's issuance of bonds should properly be limited only by the power needs of the area and that elimination of the ceiling would therefore be desirable.

Second, section 15d provides that TVA may enter into power exchange arrangements only with organizations with which it had such arrangements on July 1, 1957. Reliability of service is of the utmost importance in the operation of any utility business. In maintaining such reliability of service on an economical basis, electric utilities are coming to rely more and more on exchange arrangements with neighboring utilities. The limitation on TVA's power exchange arrangements restricts the choices available to TVA in assuring reliability of

service in its own area of operation; it also diminishes the contribution TVA can make in assuring reliability of service in neighboring areas served by other power systems. We believe this constraint should be removed by deleting the power exchange limitation in subsection (a) of section 15d of the TVA Act.

STATEMENT OF M. CECIL MACKEY, ASSISTANT SECRETARY FOR POLICY DEVELOPMENT OF THE DEPARTMENT OF TRANSPORTATION[1]

The following statement was submitted by the Department of Transportation in answer to questions raised by Mr. Proxmire on August 1, 1968 during the course of public hearings of the Subcommittee on Economy in Government. Since these questions were similar to those of the November 22 letter, no new reply was requested from the agency.

Chairman PROXMIRE. Will you submit for the record an answer to the following question: What are the primary constraints which now hinder the implementation of a sound economic analysis in your Department which could be relaxed by congressional action?

Mr. MACKEY. Yes, sir.

Chairman PROXMIRE. What kinds of legislation would you find most helpful in removing these impediments to sound economic analysis of the Department of Transportation expenditures?

Mr. MACKEY. I shall be glad to.

ANSWERS TO QUESTIONS ASKED BY SENATOR PROXMIRE OF MR. MACKEY

The questions were:

1. What are the primary constraints which now hinder the implementation of sound economic analysis in the DOT which could be relaxed by Congressional action?

2. What kinds of legislation would you find most helpful in removing these impediments to a sound economic analysis of DOT expenditures?

There are three primary features of DOT legislation which may operate as constraints to the implementation of sound economic analysis in the Department. These are:

1. Sections 7(a) and 4(b)(2)(B) of the Department of Transportation Act.
2. Laws relating to the Federal Aid Highway Program.
3. The Federal Airport Act.

Part I contains explanations of how each constraint might affect the implementation of sound economic analysis in DOT and how certain legislation could improve conditions. Part II contains a discussion of administrative and non-economic considerations which might alter the nature of the legislation.

---

[1] U.S. Senate. Subcommittee on Economy in Government. Joint Economic Committee. Hearings. "Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis." 90th Cong., 2d sess., July 30, 31, and August 1, 1968. U.S. Government Printing Office, Washington, D.C., pp. 170–172.

Provisions in the DOT Act

The language of Section 7(a) of the Department of Transportation Act limits the Secretary's role in developing investment criteria for the formulation and economic evaluation of proposals for the investment of Federal funds when those proposals are concerned with certain types of projects.

Section 4(b)(2) of the same Act is also pertinent. It states that "Nothing in this Act shall be construed to authorize, without appropriate action by Congress, the adoption, revision, or implementation of—

(A) any transportation policy, or
(B) any investment standards or criteria."

During the last several years, the Executive Branch, through the Bureau of the Budget has evolved a system (PPBS) for planning Federal expenditures in an economically efficient manner through the setting of what amount to investment criteria. These criteria are applied in most government Departments and agencies as an administrative requirement. There would not appear to be special reasons for imposing particular restrictions such as those in Sections 7(a) and 4(b)(2) on DOT's authority to manage its programs.

Laws Relating to the Federal Aid Highway Program

The Federal Aid Highway Program as it is currently structured under Title 23 would appear to contain the following three principal constraints on the implementation of sound economic analysis.

1. The law provides that Federal Highway expenditures be determined by the amounts collected from motor vehicle user taxes as they are placed in the Highway Trust Fund. Such pay-as-you-go trust fund financing schemes can be restrictive in the following two ways:

i. The trust fund equation of expenditures to amounts received from user taxes may not guarantee an economically efficient total amount of expenditure to the extent that might be possible if the receipts were from a competitive market and the taxes were annually adjusted.

ii. The pay-as-you-go arrangement may build a further inflexibility into the system. In any program in which future receipts accrue as a result of capital expenditures the agency should be allowed to borrow against future earnings at some rate of interest. Under a more flexible system, a program with a rate of growth (and rate of increase of capital stock) higher than the interest rate would spend in a given year more than it collects, while a low growth program would collect more than it spends.

2. Another type of inflexibility, essentially independent of the trust fund concept, may arise from the fact that amounts allocated to specific States and specific programs within States are distributed according to formulas in proportion to areas, populations, and miles of postal route without explicit regard to the relative benefit/cost relationships of projects among or within those States. Furthermore, the funds which States receive for constructing Interstate Highways are apportioned in proportion to the estimated cost of completing the system in each State, which would hardly be expected to provide an incentive for

economic allocation with the State since the State itself contributes only ten percent.

3. Finally, the fact that the funds for highway construction are allocated within each State by the State Highway Department or equivalent agency tends to make it more difficult for DOT to implement economic analysis since these agencies are not under its administrative control.

Broad legislative changes needed to relax the above apparent constraints might be along the following lines:

Money collected from road user taxes might be placed in the Treasury and not in a special fund.

Road user taxes might be revised on a continuing basis according to the long run marginal cost due to the movement of different types of motor vehicles. Special congestion taxes might also be levied where appropriate and feasible.

Grants might be allocated to States based on benefit/cost considerations of capital investments toward which the grants would apply.

The DOT, in cooperation with the States might help develop benefit/cost criteria for the evaluation of highway investment.

The total amount of grants might be determined by the results of benefit/cost studies of capital investments.

The Federal Airport Act

Although the Federal Airport Act contains a method of allocation of funds similar to the highway laws in that most of the fundings is allocated by a fixed formula based on area and population of States, it does not, however, contain the trust fund features of the highway laws and it provides more Federal administrative control over individual projects by means of the National Airport Plan formulation. It also allows the Federal Aviation Administrator to allocate 25 percent of the annual appropriations at his discretion.

In order to provide more flexibility for implementation of economic analysis it might be necessary to allow the Administrator to allocate the full amount of appropriations at his discretion, i.e., according to uniform investment criteria set by DOT.

PART II: ADMINISTRATIVE AND NON-ECONOMIC CONSIDERATIONS

Section 7(a)

As explained in Part I, the amendment of Section 7(a) of the Department of Transportation Act would facilitate implementation of effective economic analysis. There does not appear to be any important administrative or non-economic reason why the Act should remain as it is.

Laws Relating to the Federal Highway Program

The concept of a trust fund related to user charges contains the desirable feature of allowing for long-range programing of highway expenditures. This feature could be retained without much distortion provided that the level and structure of user charges feeding the fund are reviewed at relatively frequent intervals with an eye to revision

in the light of changes in National transportation demand. The pay-as-you-go nature of the fund could possibly be revised, however, so that bonds might be issued against future user charges revenues, and the total amount of annual construction might be determined by economic studies of transportation demand. The pay-as-you-go formula is currently mitigated to some extent by the provision that Federal funds can be used to repay State bonds for highway construction.

Perhaps the best feature of the distribution formulas is that they provide a generally predictable source of funds for each State for its planning purposes. But a portion of these funds might still be allocated at the discretion of the Secretary using guidelines which give priority to States with projects having high benefit/cost ratios. Also the distribution of expenditure as among Interstate, primary, secondary, and urban extension roads within a State could be somewhat more flexible than provided by the current law (which provides a separate set of funds for Interstate Highways, and which allocates 45 percent of another set of funds to primary, 30 percent to secondary, and 25 percent to urban extensions and does not allow either of these three apportionments to increase or decrease by more than one-fourth within a State through transfers).

As to the restrictions caused by lack of direct administrative control of State Highway Departments, there appears to be logic in retaining this feature of decentralized planning since a large portion of projects, particularly in urban areas, are of essentially local concern.

The Federal Aid to Airports Act

The distribution formula serves the same general purposes here as in the highway case. However, the current 25 percent allocated at the discretion of the Federal Aviation Administrator might likewise be increased to allow further flexibility without eliminating the automatic allocation completely.

Grant Programs in General

Eventually, grant programs regarding urban transportation and ultimately intercity transportation should be put on a common basis. A decision-making body at the local level which is choosing between alternatives, one of which involves 90 percent Federal financing (an Interstate highway), and the other of which involves competition with other areas for a small amount of grant money (mass transit), is likely to be biased in its choice. Legislation of a scope broad enough to change the current situation should be designed to improve the incentive structure for grant recipients to induce them to implement economically efficient systems. This kind of incentive structure would reinforce the effect of investment criteria established by DOT.

Section E

The Lack of Knowledge and Data as a Constraint on Effective Policy Analysis

THE ABSENCE OF KNOWLEDGE OF DISTRIBUTIONAL IMPACTS: AN OBSTACLE TO EFFECTIVE PUBLIC PROGRAM ANALYSIS AND DECISIONS

BY JAMES T. BONNEN*

James T. Bonnen is Professor of Agricultural Economics at Michigan State University.

In this paper, Professor Bonnen discusses the importance of understanding the distributional impact of government programs. He cites numerous issues of current interest which raise major equity questions, and notes the extent to which our present knowledge of the distribution of program benefits and costs falls short of the information needed to make appropriate equity decisions. Professor Bonnen sets forth a list of questions about the size and incidence of the benefits and costs of government programs and the characteristics of the beneficiaries. He suggests that "When one can answer . . . these questions with some degree of clarity and when this information can be reasonably well integrated, it should be possible to evaluate the distributional impacts of the public program involved."

After explaining how the source of funding and the design of programs can have a major effect on the incidence of program benefits and costs, Professor Bonnen examines the distributional consequences of three government programs—higher education, water resource development, and the farm program. He asserts that "In some cases, the disbutional characteristics of a program . . . have become, over time, quite perverse when measured against the original purpose. . . . More commonly, the distributional consequences of the programs may have been perverse, but we have not really known this with any certainty because it was not obvious and we never bothered to measure it." Professor Bonnen concludes that strong efforts must be made by both government and academic analysts to improve the data and analysis related to the distributional impacts of government programs. Moreover, he asserts "Until Congress concerns itself with these matters more systematically, programs are likely to develop dynamics of their own, running on in This should be a source of concern to the Congress. Distributional data are not currently demanded in the administration of programs. If Congress and the President do not demand them, they will never exist."

Introduction

The distributional impacts of both public and private decision making increasingly are being questioned. Despite our society's equity commitments, many public programs are administered with little attention to their distributional impact. It should not surprise us then when these programs exhibit perverse distributional consequences. Yet surprised we are at nearly every turn because we almost always fail to collect information on distributional impacts of programs. We do not even really understand the process by which distributional impacts work their way through the society. Without such knowledge we often work at cross purposes, we waste resources, and we fail to attain program objectives.

*I wish to acknowledge the helpful suggestions of William Capron, Lynn Daft, William Easter, Robert Haveman, Gail Steg, and Gene Wunderlich.

This paper has four parts. First is a discussion of the importance of distributional impacts in society. Second is a brief exploration of the endemic lack of data on distributional impacts. Third is a longer discussion of the type of information necessary to measure and evaluate distributional impacts. Fourth are recent analyses of the somewhat surprising distributional consequences of three public programs, presented as examples of the kind of distributional analyses that are possible in some cases even with limited data.

The objective of this paper is to develop some understanding of the need in public decision making for greater concern over distributional effects and their measurement.

The Importance of Distributional Impacts*

The distributional imacts of public programs are important, and are becoming more important. There are several reasons for this.

First is the great growth in wealth of this nation. As a society moves from relatively modest levels of per capita and family income to unprecedented levels of wealth, its social and economic priorities shift from an early dominating need for economic growth to some greater concern with how that growth is distributed. Since 1929 U.S. real national income has more than tripled and per capita income has doubled. Median U.S. family income in 1968 exceeded $8,000 per family [14]. Much of this growth has occurred in the period since World War II. As median income grows, the absolute dollar difference between the top and the bottom incomes increase even more dramatically. As a consequence, the capacity for vast differences in human welfore and life styles is increased. With great growth in the wealth of a society, the concept of poverty changes and the society's notions of the minimum acceptable level of physial welfare increase. Finally, also with economic growth, transportation and communication technologies improve and the economic and social space of society shrink greatly. With television in almost every American home, consciousness of the disparities in wealth grows, even more rapidly than the disparities themselves. Thus, it is to be expected that as a society's wealth grows its priorities will shift to include a greater concern for equity goals.

Increased concern in society over matters of equity creates a need to know more about existing distributions and the distributional impacts of, for example, public programs. Distributional information is usually necessary before meaningful determinations of equity can be made. Equity itself is a value judgment made by society about what is fair or equitable. This judgment normally is legitimized by the political process usually through legislation that establishes some standard or criteria of equity. The idea of equity is not to be confused with that of equality. An equal distribution may or may not be considered equitable depending on society's or the individual's criteria for what is fair or equitable.

There is a social organizational or political reason that also explains the greater importance of distributional issues today. The question of equity is tied to the concept of community. Between groups where there is no community, there can be no meaningful notion of equity.

*Further discussion of this issue is found in the papers by Weisbrod, Schmid, Freeman and McGuire, in this volume, and Feldman in vol. 3 of this collection.

When we begin to redefine community, politically and socially, the meaning of equity also is redefined. We are today in the middle of just such a redefinition of community. Thus, our sense of equity is being transformed and what was acceptable by yesterday's standards of equity may not be acceptable today. For this reason, many older public programs are increasingly questioned today.

There is, in addition, a related or underlying technological cause for the increasing concern over equity. In the last thirty to forty years, we have built into the dynamics of our society a high, sustained rate of technological change. Our industrial revolution has moved from an age of invention into a scientific age in which technical advance is generated by the consciously planned application of science. Our scientific research and development capacity is today much heralded. Through the early stages of this transportation we have lived with a political and social mythology that assumed technological change was neutral in its distributional impact, i.e., that it did not change the existing distributional patterns. In truth, this is rarely the case. Typically, some groups gain and others lose in any technological change.

The naive belief that all technical change is neutral in distributional impact is today questioned both by social scientists and by citizens. In fact, unrestrained technological change has created so many undesired environmental effects in man's biological and social order that the necessity for humane management of our technology is now generally viewed as urgent. Unfortunately, we know very little about the processes by which technological change works its distributional effects and we seem to be greatly deficient in institutional arrangements to contend with the problem.

All of this results in a pervasive concern about society's performance in terms of equity. This greater concern focuses primarily on the equity performance of the economic engine of society as well as on various major institutions, the equity consequences of the political and social organizational rules of society, and the equity impact of our public program decisions. Every major institution in the society, the universities, corporations, foundations and government agencies, is under increasing pressure to re-evalue it behavior as a part of our social system and to measure itself more conscientiously against society's evolving standards of morality and equity. No matter the legitimacy of the pressures or one's reaction to them, they are real. They exist and we must in some fashion contend with them in the political process. To do this for public programs we must have information on the distribution of program benefits and costs.

This paper is concerned with information on distributional impacts of public programs. However, the neeed for and the importance or value of this information derives primarily from the equity problems and concerns of society. Without judging the values involved let us describe some the equity issues that suggest the types of distributional information that are needed today.

ISSUES REQUIRING DISTRIBUTIONAL KNOWLEDGE

The equity issues currently raised are many—some old, some new. Let us describe a few that in recent years have been voiced in some form in the Congress. There have been many questions and much

conflict over the apparent lack of racial and minority group access to public programs as, for example, in education, food and nutrition, health and employment. In these as in other issues, much of the conflict is over the question of what the distributions of program benefits actually are. Adequate distributional information is rarely available upon which to judge the equity issues.

At the heart of the debate on the price and wage guidepost policy of the last administration is the equity issue of how increased productivity is to be shared. Besides the immediate interest of consumers, labor and management, other major groups are affected if the overall consequence of price and wage decisions leads to inflationary redistributions of income and wealth. Such *ad hoc* arrangements as the guideposts are a monument to our failure to build adequate institutional arrangements for deciding such issues in a politically acceptable and equitable manner.

The debate on the need to share and the advisability of sharing Federal revenues with State and local government is generated by the very great disparities between the States in willingness and ability to meet human needs.* The pressure toward changing the rules on public assistance grants to States is also grounded in a similar concern with Aid to Dependent Children (ADC) payments, which, for example, in New York are many times those of Mississippi. Some believe that program differences of this sort create undesirable migration and compound other problems.

The perennial efforts to reform the tax structure are primarily equity matters. Besides the general concern for tax loopholes, there is a rising current concern over the alleged abuse of tax exempt status by creation of some spurious foundations and the maladministration of others. There is, as well, the longstanding dispute over the oil, gas, and mineral depletion allowance.

The public debate over the selective service or draft law primarily involved disputes on the equity of who was being drafted.

The economic growth consequences and equity implications of the prevailing geographic and regional distribution of public expenditures on defense contracts, public facilities, and the federal support of higher education have all come up for debate.

Equity considerations arise in the distribution of research and development (R. & D.) funds since R. & D. investment is believed to have a great deal to do with the distribution of future economic growth by area as well as functionally in the society. An additional equity concern of increasing importance arises because the benefits and costs of technological change are not by some standards equitably distributed. Indeed, as indicated earlier, the society increasingly questions the distributional consequences of technological change in a most fundamental manner. There is an obvious need today to do the best we can in analyzing both of these kinds of distributional consequences from public investments in R. & D.

There is a Pandora's box of related equity issues. Historically, our society has tended to allow market forces to determine the answers to the question of who should benefit from a new idea or technology.

---

*Further discussion of this issue is found in the papers by Schultze, and Mushkin & Cotton in this volume.

However, rising levels of conflict suggest that the distributional consequences of this approach are less and less acceptable to the society. Yet we are not now actively developing adequate alternative institutions to determine and legitimize who receives the benefits from new technology or other innovations in society. A case in point is the current conflict over who is to control Community Antenna Television (CATV). Many question whether this uninhibited clash of self-interest will well serve society and the public interest except as an accident. This scene is replicated practically every time there is a major innovation in our technologies. Its most fundamental implication for society lies in the fact that this is the process by which we change many of our social rules. It is a process that permits a nearly random rearranging of the fabric of our society with all that implies for equity. Many social observers argue that an inadequate consideration of equity lies at the origin of much of the corrosion of community that we see in social behavior today.

Law and order is the subject of much debate today in the face of a not inconsiderable amount of disorder. Some argue that much of the disorder has its origin in the manner in which law itself is enforced. Leaving aside the current most controversial examples, take the limited case of the enforcement of property rights in this society. The benefit impact from strict enforcement of property rights has a distributional consequence rarely appreciated. Those who benefit directly are naturally limited to those who have property and the more property you have the greater tend to be your benefits. Since property also is highly correlated with the legitimate access to the exercise of police power this tends to be a closely linked, directly reinforced, distributional system. This, in turn, leads directly to another issue of considerable import, the performance of the instrumentalities of law when property and human rights come into conflict. For example, recourse to civil or criminal suit to protect human rights is not equally accessible to every member in the society. Indeed, access to justice is highly correlated with the control of sufficient assets to command good legal talent.

The right of access to many public services is a property right that presumably comes with citizenship. However, the distribution of the right of access to public services in this society has come to be a matter of major contention. Yet, we know almost nothing about how to protect or allocate these rights. We really do not know how legitimately to change the rules of society without incurring great costs. In fact, social scientists do not understand the process by which such property rights are related to the other strategic factors of the social order or of the economic engine of society. Thus, at this point we know almost nothing about how to develop institutions that legitimately may determine the distributional performance of the society or its economic sector.

One might go on almost endlessly, but this list should suffice to make concrete the fact that equity problems are important, have major political implications, and are in some cases both longstanding and of increasing importance. Different groups do not reflect the same concerns but in the various problems and stresses reflected in the political and social processes today, equity questions are high on many agenda for social action. If one is to face these issues intelligently good data on distributional effects are necessary.

The Lack of Knowledge and Data on Distributional Impacts

As one reviews Congressional hearings and agency program materials, one is impressed by the lack of knowledge or even raw data on program impact including the distribution of program benefits. It is as if these were impolite if not impolitic questions to raise.

A recent experience of the author is instructive. At the invitation of Charles L. Schultze, then on the senior staff of the Brookings Institution, and more recently U.S. Budget Director, I attempted to measure for a Brookings conference the distribution of benefits of certain Federal subsidy programs. It was decided that these should be the Bureau of Reclamation's irrigation water development program, the Maritime Administration's ship construction and operating subsidies, and the Federal Aviation Administration's subsidy programs. Despite eight months of consistent work, it was not possible to put any really useful numbers together. The Maritime Administration and the Federal Aviation Administration provide very little public information. They put minimal effort into systematic data collection or analysis. The apparent reason is that the demands made upon the administration of these programs do not require much data. The number of direct beneficiaries of the Maritime subsidies run only to about twelve or thirteen firms in any one year. One gets the impression that as far as quantitative records are concerned these programs are managed from the lower left-hand drawer of the administrator's desk.

The Bureau of Reclamation is another matter. Here there is a wealth of data, but readily available only in forms that tend to defeat any comprehensive analysis of the programs distributional impacts. The total reclamation program data that the Bureau reports is aggregated from reclamation project reports but the project level detail is not given. Thus, it is not really possible to connect the statistics that are reported for the projects with the statistics that are reported for the U.S. Both sets of data must be integrated for any comprehensive analysis of the program's distributional impacts. It almost seems as if these data were consciously arranged to avoid the possibility of any distributional analysis.

For the purposes of the Brookings paper I was forced to abandon my original ideas and to use some data that were available from the 1964 agricultural price support program for cotton. The U.S. Department of Agriculture's Agricultural Stabilization and Conservation Service (ASCS) produces for its own administrative purposes strikingly detailed statistics on allotment acreage and number of farmers by allotment size group. From this and certain of the pricing and other value data is was possible to produce an estimate of the distribution of benefits [3].

ASCS deals with about two and one half million beneficiaries for whom detailed records must be kept as to their eligibility for allotments and the size of those allotments and their compliance. Due to the administrative complexity of the program a substantial amount of internal data collection is necessary. Over time much of this has come into the public domain. It is important to note, however, that a lot of the data in the public domain today is there only because Congressmen or Senators initially asked for it and saw that it was published.

After reviewing a number of Federal program areas one must conclude that we are unlikely to improve the situation, with respect to data availability for this type of analysis, until Congress asks some of the more important distributional questions of those administering the programs. Congress is going to have to be willing to put additional resources into more sophisticated data retrieval systems than presently exist. These data are not currently demanded in the administration of the programs. If Congress and the President do not demand them they will never exist.

The facts are that we do not even fully exploit presently existing data. Economists have not done the analysis of distributional issues that is now possible in those policy areas where some data are available. The PPB system has already helped to some extent to bring distributional questions into focus and to establish the relevance of equity issues and the urgency of many distributional questions. However, in a general sense, economists are still shirking the responsibility to move beyond efficiency criterion in their analytic interests.

THE INFORMATION NECESSARY TO APPRAISE DISTRIBUTIONAL IMPACTS

What de we need to know if we are to estimate the distributional impacts of public programs and judge their desirability? Look first at the ideal situation. We need to be able to answer the following questions in something like the order in which they are stated.

For benefits:

1. What is the purpose or objective of the public program or legislation, part of which is the question, *who should benefit?*
2. *Who actually benefits*, what groups? It is sometimes not easy to identify beneficiary groups clearly.
3. How much are the *total benefits* of the program? Placing a value on the benefits of many programs is also not an easy analytical proposition.
4. What is the *distribution of program benefits* among beneficiaries?
5. What is the *current distribution of incomes and assets* or other relevant dimensions of welfare among (a) actual beneficiaries and (b) intended or potential beneficiaries?

For costs:

6. *Who should pay the program costs?* Sometimes the nature of the program contains strong implications as to who the burdened should be; other times this is almost an unanswerable question.
7. *Who actually does pay* the cost of the program? Identification of the burdened groups should consider not only the tax structure, but direct price and income effects and the indirect effects of major factor and product substitution caused by the program.
8. What are the *total program costs?* Many times this includes, as it does in question 7, economic and social costs not reflected in federal budget expenditures but market and non-market costs generated through the operation of the program itself. Thus, these are not simple questions.
9. *How are program costs distributed* among the burdened groups?
10. What is the *current distribution of incomes and assets* among

(a) the actual burdened groups and (b) the intended or potential burdened groups?

Finally:

11. Integrating the above information, what are *the alternatives* in achieving the same program objective and which alternative is most efficient; that is, attains the same desired distributional (or other) impact but at a lower cost?

When one can answer all of these questions with some degree of clarity and when this information can be reasonably well integrated, it should be possible to evaluate the distributional impacts of the public program involved. I know of no case in which such an ideal has been obtained to date. Major theoretical as well as data gaps must be filled first. Until then, decision makers will operate with far less information. It should be pointed out, too, that economists as economists cannot answer all of these questions. The normative matter of who should benefit or who should pay involves value judgments which economists may not make as scientists. Of course, if some general statement of norms or objectives can be provided, the economist like any logician can develop certain conditionally normative deductions upon which to base a decision as to who should benefit and who should pay.

It is often not possible to answer with any clarity questions one and six about who should benefit and who should be burdened. The evidence from the legislation or from the program's history may not be that clear. The economist, however, may make a substantial clarifying contribution if he can answer questions two and seven; that is, who actually does benefit and who actually is burdened by the costs of these programs. By injecting positive information of this nature into the policy process one often helps produce a clarification of the objectives.

The economic theory literature most related to these questions is focused on the classical theory of factor shares and national income distribution.[1] Little of it is of value for our needs and even on its own grounds it is an intellectually underdeveloped area today. That which does serve to some extent is the eclectic benefit-cost and public finance literature. We are in need of theoretical and analytical constructs that will allow us to interrelate in a causal manner changes in macro-economic variables, public program impacts, and changes in the economic and social rules of society with the distribution of income and assets, not only for the Nation but for regions, various sociodemographic subsets, and interest groups. This is a tall order, but the needs are more and more urgent and economists must pursue this task.

Economists have spent most of their conceptual and empirical energy on questions three and eight, that is, in attempting to measure the total benefits and the total costs of particular programs. The empirical part of the benefit-cost measurement literature focuses almost exclusively on these questions.

Answering questions four and nine on the distribution of benefits and costs is rarely a simple problem for it varies profoundly with differences in the nature of the program and the variety of funding techniques used. Both funding and program influences are discussed in some detail in the following two sections.

---

[1] See the review of literature in [22, 23, 24] and the volume on the theory of income distribution by Weintraub [30]. Also look through the A.E.A. Index of Economic Journals [1]. Note that the classification scheme of the Index contains no major distribution category except for factor shares.

Questions five and ten, about the prevailing income and asset distributions among actual or intended beneficiaries and burdened groups, are not independent of the first two questions about who the beneficiaries or burdened groups are or should be. Indeed, collecting data on such very specific incidence groups can be a difficult analytical and data collection problem. Not only are asset distributions generally nonexistent but getting income distributions for specific incidence groups can be extremely difficult, if not impossible, in many cases today. However, it is absolutely necessary to proceed through this stage, for even after you have measured the distribution of program benefits and burdens, it is not possible to judge the welfare or equity impact of the programs until you know the welfare and equity situation of those affected. Weisbrod and Hansen have a recent interesting treatment of this problem [31].

In the case of none of the programs that I am aware of, are we in a position today to mesh the information on the distribution of program benefits with really adequate knowledge of the prevailing welfare conditions of the target population. In fact, about the best we are able to do in most of these programs is to identify who the beneficiaries are. In few cases are we able to say how much of the benefits individuals with various characteristics receive. In even fewer cases can this then be matched with a little data on income characteristics of beneficiaries. In almost no case can we match benefits with current asset distributions for the target population. If you do not know what a program is currently doing to a target population you cannot possibly work intelligently and systematically to improve the direction and character of impact. Generally, we do not really know today what we are doing to ourselves in equity. Our state of knowledge is frankly pitiful.

Finally, answers to question 11 on the cost of alternatives to attain a program objective are necessary to provide information on the trade-off between alternative distributional (as well as other) features of a program.

THE EFFECT OF FUNDING SOURCE

The funding of the social insurance programs out of trust fund sources presents a fundamentally different problem from that of a welfare program transferring money income to dependent children out of general funds collected directly from taxpayers. Measuring the distribution of burden presents a different problem if the program is funded from an income tax rather than a sales or property tax. In the situation where the program is funded out of general revenues, one has to assume a cross section of the revenue sources in structuring one's analysis of the distribution of burden.

In many cases where a program involves a product that is eventually sold to consumers (as in the wheat price support program) part of the cost of the transfer of benefits (to farmers) is funded not by taxpayers but by consumers through some device, the cost of which is passed on directly in the price to the consumer. In the case of the wheat program, a certificate payment system produces this consequence. Wheat price supports also result in higher than current free market prices and thus create a direct transfer from consumer to producer. In other cases public laws or programs may affect real wealth transfers di-

rectly between groups in the society without use of the market. Other programs are funded in a manner that transfers the incidence of burden over time, classically through the sale of debt. There are other issues that involve substantial spatial differences in the distribution of burden so that one region, in net effect, may be subsidizing another even though both are eligible for the program. Other questions of burden may require measurement of the impact on different income groups and various economic interests. In any case, it is readily seen that the problem of measuring the distribution of burden of program costs can be quite a different analytical problem from one program to another.

The real incidence of burden is not always obvious. Factor and product substitution effects can be very substantial and are rarely included in calculations of program costs. Structural factors intervene. The incidence literature in taxation identifies many situations in which incidence is shifted. Analogous situations arise in measuring the distribution of burden of program costs.

THE EFFECT OF PROGRAM DIFFERENCES

The nature of the individual program also has a profound effect on the analytical problem. Expenditure programs are of many different forms. Capital investment programs, particularly of the natural resource development type, have been the primary focus of most of the benefit-cost type of analyses. Some of the most common Federal expenditure programs, however, are those that transfer real income. These programs can be quite varied in nature. Some transfer income directly, other provide social services and even others transfer real income in kind, as in the food distribution programs. While equity considerations are not given very high priority in most capital investment programs they tend to be a prime objective of most public expenditure programs that transfer real income directly to some group in society. In any case, some equity considerations are eventually imposed on most direct income transfer programs even if they were not an original consideration in the intent of Congress or of the political interest groups that generated the support to create the program.

Government research and development (R. & D.) expenditures constitute a third type of public expenditure program that has distinctive implications for data needs and analysis. While the immediate beneficiaries of a federally subsidized R. & D. investment may be the initial recipients of the funds, the long run consequences of these programs transform the technologies of society and through technology many of society's social rules. These programs are of singular importance for they cast long shadows into all parts of the society.

Closely related are the regulatory programs and legislation that change the social organizational rules of society. Some of these, such as the antitrust laws, attempt direct manipulation of markets by imposing rules on the structure of markets or on market behavior. Much legislation deals directly with the rules by which society lives. Many of these rules, as for example, the patent and copyright laws, specify criteria that attempt to affect who shall get the initial direct benefit of a new idea or technology.

Finally, there are some additional matters. It is worth noting that the distributional impacts of program changes are likely to be different in the situation where the change in the program involves an expansion than where expenditures are cut. In short, there is a need to analyze distributional impacts at the margin when evaluating programs. The experience of tight budgets over the last several fiscal years has underlined this rather dramatically in some cases. Every economist understands that most changes in prices not only have a price effect but an income effect on both household and firm behavior. Similarly, most income changes have a price effect as well as an income effect. It is well to keep in mind in analyzing public programs that most changes in programs will have both a resource use effect as well as an income redistribution impact. Both will have incentive and disincentive consequences.* In short, our responsibility in analyzing public programs does not end with the goal of efficiency and its measurement in a benefit-cost ratio. Indeed, efficiency and equity are frequency intertwined analytically as well as in social policy.**

ADDITIONAL CONSIDERATIONS

Besides gaining command over the distributional impacts of individual programs there is another level of knowledge that is necessary for intelligent and comprehensive policy planning. Congress and the executive branch must know something more about the aggregated impact of Federal programs by regions and for the United States as a whole. That is, it is well that we know the regional effects of individual programs and thus their differential regional impacts. It would be most useful to be able to aggregate the natural resource development investments of various resource development programs into regional aggregates with distributional knowledge within each region and between regions. The same would be true for the whole of our educational input or of the full range of our health programs. There is a need to gain an overview of the plethora of individual programs which presumably add up to some comprehensive whole but for which at this point in time we have very little in the way of knowledge about how they actually add together, if they do. Nor do we have any real notion of the spatial implications for these general program areas. There are indications in our development literature and in studies of the social pathologies of this society, such as poverty, that some of our difficulties today arise as a result of the systematic long term differentials in the distribution of investments in education, in natural resource development, etc. Before we can avoid these kinds of problems we need to know they exist.

There is, in addition, a need to develop a measurement of the distributional impacts of total Federal program expenditures. In other words, we need an overview of the national picture of public expenditures and funding with their impact by various distributional characteristics—income, age, sex, family status, degree of urbanization and race, for example. Again some subcategorization such as educational programs, health, manpower, etc., would be useful as well.

---

*Further discussion of this issue is found in the paper by Schultze in this volume.

**Further discussion of this issue is found in the paper by Weisbrod in this volume.

An interesting, though given the state of the arts a necessarily rather assumption-ridden, approach to this problem has been produced by Gillespie. His general finding is rather striking. His results indicate that for total Government expenditures and funding:

> The Federal pattern of fiscal incidence generally favors low incomes, burdens high income, and is mainly neutral over a wide middle income range. The State and local pattern also favors low income but is essentially neutral over both the middle and upper income ranges. A comparison of the two patterns controverts the conventional view that the State and local's tax structure is regressive and the Federal tax structure progressive; indeed, just the reverse is true, for in net terms the State-local structure grants larger benefits to the lower income range than does the Federal [17].

Gillespie goes on to conclude:

> More basically, the question arises whether equity . . . should be considered only with reference to tax burden distribution, or whether the relevant criterion should be defined in terms of net benefit (or burden) distribution. If the latter view is taken, tax reform and expenditure reform can no longer be considered independent problems [17].

There are several reasons why one might question Gillespie's results but his attempt at empirical measurement of the overall impact of public programs on the distribution of income underlines dramatically the very significant conceptual and data needs we face before genuinely satisfactory measurement will be possible.[3] Indeed, one of the conclusions that should be drawn from Gillespie's effort is that this kind of empirical measurement of public program performance can best be done by the Federal Government itself or in very close cooperation with the Bureau of the Budget and other immediate sources of internal program data.

Perhaps we need again to repeat the model that was followed in the creation of our national income statistics. Much of the intellectual investment in the primary design of concepts and system was executed by the National Bureau of Economic Research in close conjunction with Federal statistics agencies. It was then implemented by the Government as a Federal system. Certainly the conceptual and data problems are on the same order of difficulty and we are in a similar early stage in developing these numbers for program analysis purposes. Such a system, of course, would be designed to collect data for the full range of program analysis needs including distributional questions.

Distributional Consequences of Three Government Programs

This Nation has a long history of creating public programs at least in major part for equity reasons. Among older examples are the farm programs, Corps of Engineers and Bureau of Reclamation programs, and many of the public assistance or welfare programs; examples from recent years include the poverty programs, the food distribution programs, many of the manpower and some of the education (title I of

[3] For an earlier interesting study of total Federal fiscal distributional impact see Alfred H. Conrad's study in [27].

ESEA) programs. Much of the current aid to urban areas has a major equity element among its purposes.

Yet the distributional consequences of our equity programs often turn out to be perverse. We seem to manage these programs without much continuing thought to their equity objectives. Equity objectives or standards are rarely spelled out in the legislation. Once enacted equity concerns seem to recede. Measurement of goal attainment is often reduced to gross rules of thumb, smearing and confusing the original legislative intent. The efficiency of administrative organization is substituted for program efficency as an overriding congressional concern once clientele support develops. In execution this seems eventually to lead to a fuzziness if not outright drift in the purpose of some programs. The PPB system should lead to improvement but until Congress concerns itself with these matters more systematically, programs are likely to develop dynamics of their own, running on in a self-serving contemplation of the use of public wealth for private ends. This should be a source of concern to the Congress.

We know that, in some cases, the distributional characteristics of a program, while perhaps not to begin with, have become over time quite perverse when measured against current equity standards or even the program's original purpose. This would appear to be the case in the instance of the farm programs which were created because farmers as a group were receiving far less for their labor and resources than nonfarm people. Today the program is under criticism because the distribution of program benefits *among* farmers is highly concentrated with a large number of farmers receiving very little and a small number receiving very large payments and price-support benefits.

More commonly, the distributional consequences of a program may have been perverse but we have not really known this with any certainty because it was not obvious and we never bothered to measure it. Investment in higher education, for example, may not, as we tend to assume, provide general upward mobility on the basis of ability rather than wealth.

There is evidence, too, that some natural resource investments such as those made in the West by the Bureau of Reclamation have very perverse, and, one assumes, unintended distributional consequences on the growth of other regions.

Let us look briefly at these three cases as examples of the kind of analysis that can be done in the context of public program analysis for policy decisionmaking. By some standards at least, all exhibit distributional characteristics that are not now being considered adequately in program decisions and design.

THE DISTRIBUTION OF BENEFITS AND COSTS OF HIGHER EDUCATION

What do we know about the distributional consequences of society's investment in public higher education? The belief is fairly general that most youth have access to this system on the basis of ability, that education provides an avenue for upward mobility in this society based on ability rather than wealth. Certainly this society holds normative beliefs that suggest that the direct benefits of higher education should be distributed in this manner. What are the facts about the distributional impact of this investment in human resources? W. Lee

Hansen and Burton A. Weisbrod have analyzed this question for the State of California [18, 19]. Their work is instructive for its ingenuity, for the kind of distributional analysis that it suggests is possible in human investment programs, and for the kind of surprises that distributional analyses can provide.

The California system of higher education has the general reputation of being one of the most open and accessible and as well financed and with as low tuition as any in the Nation. Yet Hansen and Weisbrod's data suggest strongly that the effect of the system is most regressive. It can be seen in table 1 that the subsidy as a percent of income rises from 12 percent at $8,800 of income to 41 percent at the $12,000 level of income. The authors find that while the tax structure is slightly regressive the benefit distribution is quite regressive.

This results from several factors. The total subsidy to a student's education averages $4,870 at the University of California, $3,810 in the State college system, and only $1,500 in the junior college system. Since the average family income rises progressively from $7,900 for those who have no children in California public higher education to $12,000 per family with children at the University of California, the effect is to provide the largest educational subsidy to the wealthiest. This is partly the consequence of academic entry requirements since 80 percent of the high school graduates are not eligible to enter the University of California. But that is not the end of it. They go on to state:

> Even more interesting is the fact that the percentage of all students qualifying for the University of California [table 2, column 1] rises quite dramatically by family income level—from about 10 percent in the lowest bracket (under $4,000) to 40 percent in the highest (over $25,000). Thus, the correlation between high school achievement and family income—and all that it reflects—is startling indeed. This pattern persists as we widen our view to include those eligible for both the university and State colleges [table 2, column 3]. But a close examination of the differences between the two columns shows that the percentage of those eligible only for the State college system is roughly constant at all income levels; thus, university eligibility requirements account largely for the unequal distribution of opportunity [18].

TABLE 1.—AVERAGE FAMILY INCOMES AND AVERAGE HIGHER EDUCATION SUBSIDIES RECEIVED BY FAMILIES BY TYPE OF INSTITUTION CHILDREN ATTEND, CALIFORNIA, 1964

	California families without children in California public higher education	California families with children in California public higher education, children in—		
		Junior college	State college	University of California
Average income	$7,900	$8,800	$10,000	$12,000
Average subsidy	0	1,050	3,810	4,870
Subsidy as percent of income	0	12	31	41

Source: W. Lee Hansen and Burton A. Weisbrod, "Benefits, Costs and Finance of Public Higher Education," (Chicago: Markham Publishing Co., 1969), ch. IV, table 6.

TABLE 2.—DISTRIBUTION OF HIGH SCHOOL GRADUATES BY ELIGIBILITY FOR PUBLIC HIGHER EDUCATION IN CALIFORNIA, BY TYPE OF EDUCATION AND FAMILY INCOME

[In percent]

Family income	Percentage distribution of high school graduates by eligibility for—		
	University of California (1)	State colleges (2)	University of California and State colleges (3)
0 to $3,999	10.7	17.3	28.0
$4,000 to $5,999	11.5	14.8	26.3
$6,000 to $7,999	11.9	18.6	30.5
$8,000 to $9,999	16.2	17.0	33.2
$10,000 to $12,499	19.4	17.7	37.1
$12,500 to $14,999	22.5	17.3	39.8
$15,000 to $17,499	27.9	17.5	45.4
$17,500 to $19,999	29.5	15.6	45.1
$20,000 to $24,999	33.3	12.8	46.1
$25,000 and over	40.1	14.2	54.3
Not reported	13.3	14.7	28.0
All	19.6	16.7	36.3

Source: W. Lee Hansen and Burton A. Weisbrod, "Benefits, Costs, and Finance of Public Higher Education" (Chicago: Markham Publishing Co., 1969), ch. IV, table 7.

Hansen and Weisbrod also estimated for California that "the combined State and local tax structure is regressive below $8,000 and is essentially proportional above that level."[4] Thus, they conclude that:

> The general nature of the redistributive effects of the current method of financing public higher education in California is clear. Some low-income people have benefited handsomely from the availability of publicly subsidized higher education. But on the whole, the effect of these subsidies is to promote greater rather than less inequality among people of various social and economic background, by making available substantial subsidies that lower income families are either not eligible for or cannot make use of because of other conditions and constraints associated with their income position. To overcome the effects of the present system would require a substantial overhaul of the pricing system in public higher education, a realignment of the tax structure, and/or a broadening of the eligibility base for public expenditure programs. With respect to the latter alternative eligibility for public subsidies to young people might well be expanded so as to embrace all young people—not only those who go on to college but those who opt for alternative ways of expanding their earning power, such as apprenticeship or on-the-job training, or even investments in businesses. In any case, it is clear that whatever the degree to which our current higher education programs are rooted in the search for equality of opportunity, the results still leave much to be desired [18].

THE DISTRIBUTIONAL IMPACTS OF WATER RESOURCE DEVELOPMENT

A considerable literature has developed around the problems of constructing benefit-cost ratios for natural resource investment projects. These have mostly concerned water and power development, conservation, and flood prevention projects. In all of this activity little effort

[4] Note this is different from Gillespie's results for the U.S. [17].

has gone into comprehensive attempts to measure the distributional impacts.

There are some recent major exceptions I should like to point out. There is first of all the pioneering study by Robert Haveman which develops and applies a procedure for adjusting the benefit-cost calculation for the income redistribution consequences of the benefit distribution as well as the distribution of the cost of funding Corps of Engineers' projects in the 10 Southern States [20]. This adjustment results in a number of (though by no means all) projects with otherwise negative B/C ratios to exhibit positive values. Haveman's analysis suggests (as do some of the Corps hearings) that welfare and equity considerations entered into the political decisions to invest in these projects. The substantial adjustments that result also suggest that distributional analyses should be done for all major natural resource investment projects.

Subsequently, A. Myrick Freeman has applied a similar methodology to six Bureau of Reclamation projects in the West with interesting results on the income distribution impacts [16].*

Even more recently in a very different kind of study K. William Easter and Charles W. Howe have explored the interregional redistributive consequences of Bureau of Reclamation water development investment in the West [21]. Their emphasis is upon the regional transfers that take place. This fascinating study is soon to be published by Resources for the Future.

The issues are complex. Western farmers pay an annual price of from $30 to $135 per acre below the cost of supplying water [28]. And the consequent price of irrigation water is below that which "other potential users are willing to pay although at present not all of the irrigation water could be put to these higher valued uses" [21]. The value added per acre foot of water varies greatly by use, from well under a hundred dollars per acre foot in agriculture, through several hundred dollars in various recreational uses, to thousand of dollars an acre foot in municipal and industrial uses [32, 33]. Most of the water developed in Western reclamation projects goes into agricultural uses despite its lower value product. For economic, legal, and political reasons it is difficult to transfer to other uses later. Thus, major malallocations of resources are developing within agriculture and between agriculture and the nonagriculture sector in the West.

But this is not where the problem ends. Since there exists a chronic overcapacity to produce in agriculture, the issue arises as to whether the development of new irrigation lands has any positive benefit nationally whatever its local value. In addition to which the creation of highly productive irrigated land in the West must (particularly in the face of the existence of excess agricultural capacity and acreage allotments) create, through production substitutions, income losses in other regions. The question is where and what are the welfare implications. Easter and Howe have developed an empirically based three region farm production model for several commodities along the general lines suggested by Tolley [29]. From this they are able to estimate the substitution effects in production and the approximate impact on regional farm production and income. Finally, they also attempt to estimate the

*Further discussion of this issue is found in the paper by Freeman in this volume.

cost which reclamation projects add to the annual cost of agricultural programs pointing out the inconsistency between federal programs in which tax monies are used in one case to increase production (reclamation) and in another case to decrease production (farm price supports and acreage allotments).

With the permission of the authors, I have reproduced below tabular materials on cotton production from their three-sector model. The model also provides similar detailed data on potatoes, vegetables, fruits and nuts, dry edible beans, rice, sugar beets, feed grains, and alfalfa hay. It can be seen in Tables 3 and 4 that under the pressures of excess capacity and a consequent reduction in U.S. acreage allotments, cotton acreage has declined in every area, except the reclamation West where it rose, between 1949 and 1966. Nowhere is the decline as precipitous as in the South where 1964 acreage was less than half that of 1949. In Tables 5 and 6 it can be seen that differential changes in yields mute these shifts considerably. Nevertheless, with respect to cotton the authors conclude that:

> If cotton were not produced on reclamation land, then 978 thousand bales could be produced in other areas by increasing acreage allotments. To meet this production allotments could be increased by about one million acres since the land coming back into cotton production would be less productive and would yield on the average about one bale per acre while the reclamation land taken out yields almost two bales per acre. Assuming a uniform increase in the allotment based on the 1964 distribution of acreage, the South would harvest about 550 thousand more acres and the North 20 thousand, while in the West (including Texas and Oklahoma) acreage would decline by about 85 thousand acres. The West would be a net loser but the non-reclamation West would have a gain in acreage allotments of approximately 430 thousand acres.
>
> In terms of annual net income for farmers, the South would gain roughly $27.5 million and the North $1 million while the reclamation West would lose $50 million and the non-reclamation West would gain $21.5 million [21].

The authors go on to conclude from their model that:

> There can be little doubt from the data analyzed that reclamation irrigation has had a significant effect on U.S. agriculture. Increased production on reclamation served land has increased USDA payments, stimulated regional production shifts and reduced non-reclamation farmers' income. Annual commodity program payments were estimated to increase from $20 to $464 per acre of land irrigated by reclamation, varying with the commodity and assumptions concerning relative productivity and the effectiveness of commodity programs. The reclamation impact on crops not under the annual commodity programs ranged from the $9 million to $20 million shift in gross farm income for pears to the $350 million estimated shift in farm income for potato farmers. Finally, from 5 million to 17 million acres of cropland have been displaced by the 1944–64 increase in reclamation irrigation. In terms of income foregone this would be a minimum of $50 to $170 million annually.

TABLE 3.—U.S. COTTON ACREAGE HARVESTED BY REGIONS

[In thousands of acres]

Regions [1]	1964	1959	1954	1949	1944
California-Arizona	1,128	1,184	1,318	1,234	401
New Mexico-Nevada	188	192	199	286	104
Total, West [2]	1,316	1,376	1,517	1,520	505
Texas-Oklahoma	6,174	6,728	8,417	11,802	8,070
South Central	4,504	4,542	6,115	9,010	6,915
South Atlantic	1,570	1,598	2,383	3,673	3,074
Total, South [2]	12,248	12,868	16,914	24,485	18,060
North	351	405	427	594	397
Total [2]	13,915	14,649	18,858	26,599	18,962

[1] The South Central includes the States of Alabama, Arkansas, Kentucky, Louisiana, Mississippi, and Tennessee. The South Atlantic includes the States of Florida, Georgia, Maryland, North Carolina, South Carolina, and Virginia. The North includes the States of Illinois and Missouri.
[2] May not add to totals due to rounding.

Source: Charles W. Howe and K. William Easter, "Interbasin Water Transfers—Economic Issues and Impacts," Resources for the Future, Johns Hopkins Press, 1969, ch. IV.

TABLE 4.—COTTON ACREAGE HARVESTED IN 17 WESTERN STATES

[In thousands of acres]

Region	1964	1959	1954	1949	1944
California-Nevada-Arizona	1,131	1,187	1,320	1,235	401
New Mexico-Texas-Oklahoma	6,359	6,917	8,614	12,087	8,174
Total	7,490	8,104	9,934	13,322	8,575
Reclamation portion [1]	514	504	514	381	145
Nonreclamation	6,976	7,600	9,420	12,941	8,430

[1] In 1964, 360 acres were in California, Arizona, and Nevada while 154 acres were in New Mexico, Texas, and Oklahoma.

Source: See table 3.

TABLE 5.—U.S. COTTON PRODUCTION BY REGIONS

[Thousands of bales]

Region [1]	1964	1959	1954	1949	1944
California-Arizona	2,524	2,488	2,301	1,761	449
New Mexico-Nevada	259	306	293	262	113
West [2]	(2,783)	(2,794)	(2,595)	(2,023)	(562)
Texas-Oklahoma	4,260	4,521	3,825	6,117	3,157
South Central	5,760	4,838	4,605	5,142	5,330
South Atlantic	1,529	1,277	1,489	1,663	2,381
South [2]	(11,549)	(10,637)	(9,919)	(12,922)	(10,869)
North	401	483	407	474	407
Total [2]	14,733	13,914	12,921	15,419	11,838

[1] See table 3.
[2] May not add to totals due to rounding.

Source: See table 3.

TABLE 6.—COTTON PRODUCTION IN 17 WESTERN STATES

[Thousands of bales]

Region	1964	1959	1954	1949	1944
California-Arizona-Nevada	2,529	2,494	2,304	1,762	449
New Mexico-Texas-Oklahoma	4,514	4,820	4,116	6,379	3,270
Total [1]	7,043	7,315	6,420	8,141	3,719
Reclamation portion [2]	[3] 978	984	898	487	164
Nonreclamation	6,065	6,331	5,522	7,654	3,555

[1] May not add to total due to rounding.
[2] In 1964, 775 bales were in California, Arizona, and Nevada while 203 bales came from New Mexico, Texas, and Oklahoma. Also 567,000 bales were from land receiving full reclamation irrigation service, 403,000 bales from land with supplemental service and 8,000 bales from land with temporary service.
[3] Only 6 percent o r64,000, bales was American-Egyptian cotton.

Source: See table 3.

> These impacts of reclamation projects have not been adequately considered in evaluating project feasibility. Therefore, additional research indicating possible procedures for including these costs in the cost-benefit analysis would seem very appropriate. Even if the United States were in a situation of no excess capacity in agriculture (which the numerous land retirement programs help remind us we are not) it does not follow that additional Federal irrigation is needed. If additional capacity is desired, the least cost means of increasing capacity should be selected. Ruttan's study indicates that the returns from additional investments in irrigation itself would be greater in regions outside the West. In addition, one should consider investments in drainage, land clearing, the development of new or low cost inputs before deciding on a particular mix of programs to increase agriculture's productive capacity [21].

Both the Hansen-Weisbrod and the Howe-Easter studies are thoughtful and imaginative analyses of distributional questions. While neither produce all of the information implied by the ideal set out earlier here, both provide impressive evidence of perverse distributional impacts in major public programs. Their contribution is substantial for little quantitative knowledge of such impacts previously existed.

THE DISTRIBUTION OF BENEFITS FROM THE FARM PROGRAM*

The final case involves some analysis the author has done on the distribution of benefits from the farm program. In a general way we have known for some time that this program has a concentrated distribution of benefits. Many agricultural economists, most prominently T.W. Schultz, have long drawn attention to this fact. However, there have not previously been any systematic quantitative measures of the degree of that concentration. Little has been known about distributional differences between commodities and between regions and States within commodities [2]. Nor has there been any comparison of the differential distributional impacts of price supports and the direct payments that have entered into the program as a major factor since 1962.

*Further discussion of this issue is found in the paper by Ruttan in vol. 3 of this collection.

In the United States in the middle of the Nineteenth Century, rural life meant farm life. There was little rural economic activity besides that generated by farming. The preponderance of U.S. population lived in rural areas and practically all of that was on farms [8]. Today less than 30 percent of our population is rural, and less than 5 percent is on farms [6, 7, 9]. While some of that rural population living in small rural communities are dependent on the economic activity generated in agriculture, it is a *prima facie* matter that farm programs in the nature of things simply cannot and do not serve all rural people today.

The farm commodity programs, like any government subsidy program, have two quite different general effects and potential purposes. The first is to redistribute income from nonfarmers to farmers; the second is to affect resource allocation—to attract to an economic activity more resources than the market mechanism would otherwise pull in. The farm commodity programs were developed during the late 1920's and the 1930's. The primary political support that brought them into existence had as its clear objective the redistribution of income premised on the proposition that farmers were generally poor and that the resources farmers commanded earned far less than they could have earned in the nonfarm sector. Because of the unique characteristics of this industry it was obvious that if anything was to be done about this income situation it would have to come through public policy. It is interesting to note that even in the early history of these programs most economists, however, focused on the resource allocation effects and practically all of the analysis that has been done by economists has concerned itself with the resource allocation problems of the industry. However, a concern for equity explains the origins of these programs even in the case of the resource allocation argument where it was contended that agriculture as an industry suffered certain inherent disadvantages in the market which resulted in a chronic low rate of return on investment. This usually carried the implied social consequence that growth in productivity would be slowed relative to the nonfarm sector with consequent impact on the overall rate of U.S. economic development.

The criticisms of the farm program today seem to be generated by two factors. The program has succeeded beyond anyone's dreams in attaining its objective of supporting total farm income. Secondly, the program's distributional characteristics have been ignored.

A 1967 study by the Department of Agriculture for the Senate Agricultural Committee provides a direct insight into the equity considerations that relate to the resource allocation argument [25]. Four different standards were used for computing a parity income or parity return to resources in farming.[6] Farm programs enacted during the Kennedy and Johnson administration years had, by 1966, provided parity or higher income returns (by all four standards) for the 16 percent of all farmers who had gross sales of $20,000 or more. This group produced more than two-thirds of all agricultural products in 1966. The 16 percent of all farmers in the $10,000 to $20,000 gross

[6] A "landlord standard" assumed that the resources were rented and the farmer's labor paid for at nonfarm rates for similar age, sex, and skill levels. A "stockholders standard" assumed sale of farm resources and investment in a cross section of the Standard and Poor's 500 stock average. Both of these standards were then estimated with and without capital gains.

sales category in 1966 who produced 18 percent of all output earned 81 to 98 percent of a parity return depending upon the standard used. While there is a wide dispersion around these mean figures, these programs have now generated a parity return or better to the most efficient farmers who produce the bulk of our farm products.

The entire structure of the industry has been transformed from the traditional subsistence farming economy of the Nineteenth and early Twentieth Centuries into a highly specialized industrial enterprise in which command over resources, output, and income in agriculture have become far more highly concentrated. Today less than 10 percent of all farmers produce over half of all agricultural output; about half of the farmers produce 95 percent of all the U.S. agricultural output [15]. With few exceptions, farm programs, whether they are credit, conservation, or commodity programs, are designed today so that a farmer's access to them is directly related to the size of assets he controls, the amount of land he operates and his volume of output. Under these conditions it perhaps is not surprising that the farm programs are increasingly questioned with respect to the equity of their distributional consequences.

Recent work by the author throws considerable light on the distributional characteristics of farm commodity programs [4]. For our purposes this presentation will be limited to a summary of the results and a sample of the statistical evidence for the reader's own inspection. The data presented are in the form of Lorenz curves and Gini concentration ratios. The Lorenz curve tells one what percentage of the farmers received what percentage of the benefits. The Gini concentration ratio measures the degree of concentration of the distribution of benefits—that is, it measures how far a given distribution departs from a completely equal distribution of benefits between all beneficiaries.[7]

From table 7 it can be seen that all of the commodity programs are fairly highly concentrated, some of them greatly so. Looking down this table you will see that the 40 percent of the smallest farmers (allotment holders) receive much less than a proportionate share of the program benefits even in the case of the programs with the least concentrated distribution of benefits. In a typical program such as peanuts, table 7 suggests that it would be necessary to generate about $10 of program benefits for every dollar going to the lowest 40 percent of smaller peanut farmers. Variation across all the commodity programs would range from about $6 to $34 of total benefits for every dollar going to the lowest 40 percent of the farmers. Looking at the very bottom of the distribution, this same table shows that it would be necessary to generate from $20 to $100 of benefits for each dollar going to the lowest 20 percent of farmers.

Even after considering all the qualifications that go with these numbers, the data suggest very clearly that the farm programs are not efficient means for affecting an income redistribution to the smaller low income farmers. Since it is clear from the parity income study that the more efficient farmers are already receiving near parity

[7] A Gini ratio of 0 indicates a completely equal distribution between all beneficiaries. A Gini concentration ratio of 1 indicates that one beneficiary received all of the benefits and the other potential beneficiaries none. Thus, as the concentration ratio rises from 0 toward 1 it reflects a greater and greater concentration in the distribution of benefits. For those interested in the nature of the data sources, the methodology, and the limitations of the estimating procedure see pages 461–465 of [4].

income or better, it seems an inescapable conclusion that any attempt to solve the low income small farmer problem via price supports would generate huge windfall profits to the more efficient larger scale operators.

TABLE 7.—DISTRIBUTION OF FARM INCOME AND VARIOUS PROGRAM BENEFITS: PROPORTION OF INCOME OR BENEFITS RECEIVED BY VARIOUS PERCENTILES OF FARMER BENEFICIARIES [1]

	Percent of benefits received by the—						
	Lower 20 percent of farmers	Lower 40 percent of farmers	Lower 60 percent of farmers	Top 40 percent of farmers	Top 20 percent of farmers	Top 5 percent of farmers	Gini concentration ratio
Sugar cane, 1965 [2]	1.0	2.9	6.3	93.7	83.1	63.2	0.799
Cotton, 1964 [2]	1.8	6.6	15.1	84.9	69.2	41.2	.653
Rice, 1963 [2]	1.0	5.5	15.1	84.9	65.3	34.6	.632
Wheat, 1964:							
Price supports	3.4	8.3	20.7	79.3	62.3	30.5	.566
Diversion payments	6.9	14.2	26.4	73.6	57.3	27.9	.480
Total benefits [4]	3.3	8.1	20.4	79.6	62.4	30.5	.569
Feed grains, 1964:							
Price supports	0.5	3.2	15.3	84.7	57.3	24.4	.588
Diversion payments	4.4	16.1	31.8	68.2	46.8	20.7	.405
Total benefits [4]	1.0	4.9	17.3	82.7	56.1	23.9	.565
Peanuts, 1964 [3]	3.8	10.9	23.7	76.3	57.2	28.5	.522
Tobacco, 1965 [3]	3.9	13.2	26.5	73.5	52.8	24.9	.476
Farmer and farm manager total money income, 1963 [5]	3.2	11.7	26.4	73.6	50.5	20.8	.468
Sugar beets, 1965 [2]	5.0	14.3	27.0	73.0	50.5	24.4	.456
Agriculture conservation program, 1964: [6]							
All eligibles	7.9	15.8	34.7	65.3	39.2	(7)	.343
Recipients	10.5	22.8	40.3	59.7	36.6	13.8	.271

Source: Except as noted all figures are from a 1968 study by Bonnen [4].

This table presents portions of 2 Lorenz curves relating the cumulated percentage distribution of benefits to the cumulated percent of farmers receiving those benefits. Cols. 1 through 3 summarize this relationship cumulated up from the lower (benefit per farmer) end of the curve, and cols. 4 through 6 summarize the relationship cumulated down from the top (highest benefit per recipient) end of the curve.

[2] For price support benefits plus Government payments.

[3] For price support benefits.

[4] Includes price support payments and wheat certificate payments as well.

[5] David H. Boyne, "Changes in the Income Distribution in Agriculture," Journal of Farm Economics, vol. 47, No. 5, December 1965, pp. 1221–1222.

[6] For total program payments. Computed from data in "Frequency Distribution of Farms and Farmland, Agricultural Conservation Program, 1964," ASCS, U.S. Department of Agriculture, January 1966, tables 3 and 8.

[7] Not available.

The welfare impact of these distributions cannot actually be ascertained fully unless we have a measure of the current distribution of income and assets among beneficiaries for comparison with the distributions of program benefits. Ideally, one would desire to have this not only for the U.S. as a whole for all farmers but one should also have the income and asset distribution of cotton producers for comparison with cotton program benefits, and similarly for all the rest of the programs. Such estimates are not available. The best that is immediately available for this purpose is a measure of the Lorenz curve of the net money incomes of farmers and farm managers estimated for 1963 by Boyne [5] and for 1964 by Coffey [11].[8] Boyne's figures are arrayed in Table 7 where they can be compared with similar Lorenz distributions for the various programs. At the level of the lowest 40 percent of farmers only tobacco, sugar beet, and ACP programs have the effect of adding proportionately more program benefits to a farmer's income than he commands as a share of farm

[8] One may also compare the program distributions of Table 7 with several income comparisons that have been computed in Table 10 using very different and less satisfactory data.

income. That is, this group of farmers receives 11.7 percent of farm income, but a higher percentage than this of tobacco, sugar beet, and ACP program benefits. The lowest 20 percent of farmers receives 3.2 percent of net farm money income. But they receive more than 3.2 percent of the benefits of the wheat, peanut, tobacco, sugar beet and ACP program. Rice, feed grains, cotton and sugar cane all provide to this lowest group of allotment holder less of a share of program benefits than they average as a share of farm income. One is tempted to say that these latter programs are regressive in their income impact in farming, but this is not proved by this crude though relevant comparison. Nor can we argue conclusively that the ACP, sugar beet and tobacco (and possibly peanuts and feed grains) programs have a progressive income impact—even though our data seem to suggest this.

There are too many logical difficulties to bridge. Cotton benefit distributions should be compared with cotton farmer income distributions—not all farm income. Also some regionalization of the farm income distribution is needed to allow for geographic and industry mix differences associated with differences in productivity. One cannot always assume that one is necessarily dealing with the same general set of low incomes, or indeed with low income at all, when one speaks of the lower end of the distribution of benefits from a program. While it may be fairly reasonable to assume that a small cotton allotment represents a small farmer, such an assumption is not necessarily reasonable in the case of wheat or feed grains. Farmers do grow more than one commodity typically, and a small allotment may sometimes represent a minor enterprise in a substantial operation. It also should be noted that farm income accounts for a relatively small part of the total income of quite a number of smaller prdoucers [15]. This partially accounts for the very low concentration of total income of farmers from all sources as compared to the high concentration of gross and net incomes from farming operations that can be seen in table 10.

Thus, the net effect of these programs may be less regressive than the data suggest—or possibly more regressive—but the pattern is clear.

There are a few other conclusions that can be drawn from the distributional data. The programs exhibiting the greatest concentration in the distribution of benefits are indeed extremely concentrated. The U.S. sugar cane program exhibits a Gini ratio of 0.8 but individual states have concentration ratios as high as 0.9 (Hawaii). Perhaps even more remarkable is that the sugar program which is administered as a single program for both cane and beets demonstrates both the highest concentration of benefits among farm programs in sugar cane (.8) and the lowest concentration of program benefits in sugar beets (.456). All other commodity concentrations fall in between. This speaks worlds about the different technological structures of these two sugar industries and their differing political behavior over time in deciding how to allocate increases in acreage allotment.

Another conclusion of interest is the evidence of a time trend toward greater concentration. This can be seen in Table 7 for those commodities in which data for different years is available.

The Agricultural Conservation Program (ACP) is an anomaly among programs that generate farm income. In table 7 it can be seen that its benefits are far more equally distributed than any commodity program. ACP subsidizes various conservation practices, from putting lime on the soil to tilling the land and building ponds. It is obviously administered in a manner to assure the largest number of participants. This distributional outcome would seem to have more to do with ACP's political role among the farm programs than with any organizational or technological imperatives of conservation.

It is also worth remarking that the variation in concentration within a particular commodity program from state to state and region to region can be rather large. These reflect primarily great variations in productivity and yield, and rather substantial differences in the median allotments as well as the relative variation around those medians. See the example of cotton in Tables 8 and 9.

Tables 8 and 9 are included to provide an example of the kind of detailed data that can be computed and are available by state and region for each commodity benefit stream listed in Table 7. These may all be seen in another study by the author [4].

The available data also allow one to compute the distribution of benefits by size of allotment (see Table 8). Indeed, this is an intermediate step to producing the Lorenz curves. The availability of such a distribution with an absolute dimension that is highly correlated with productivity allows one, particularly in the case of cotton, to gain some insight into the potential for conflict between efficiency and equity objectives. Let me just quote the conclusions from the author's earlier study of cotton benefit distributions [3].

> The older eastern cotton-growing regions are higher cost areas. In fact, cost per unit of output on efficiently managed cotton farms is reputed to decline sharply, possibly by as much an one-half in some major production areas, as one moves from the Southeast to the West [12]. Historically, since 1935 the Delta and the Southwest have maintained their share of national cotton output (about one-third each), while the share of the Southeast had declined from one-third to less than one-fifth, and the West's share has increased from 4 percent to about 20 percent.
>
> Thus, in effect, cotton acreage has been shifting out of the Southeast and to the West. In efficiency terms this is a movement of allotment rights from an area where yields are about one bale per acre to one where average yields are over two bales. In equity terms it is a movement of allotment rights from smaller, lower-income farmers to farmers operating on a large scale and receiving much higher incomes. An acre of allotment rights in California is worth almost three times (before considering production costs) what it is worth in Georgia. Yet almost 40 percent of all cotton farmers earn their living in the Southeast as compared to about 3 percent in the West. The logic of this suggests that the disparity is so great today that it is probably impossible to attain the welfare objectives in the older eastern cotton areas with price supports of the 1964 program type, except at unacceptably high levels of government-generated intramarginal rents in the West and other producing areas of high productivity.

TABLE 8.—DISTRIBUTION OF 1964 UPLAND COTTON PRICE-SUPPORT BENEFITS: PROPORTION OF UNITED STATES, REGIONAL, AND STATE BENEFITS ACCRUING TO FARMERS WITH ACREAGE ALLOTMENTS UNDER OR OVER VARIOUS SPECIFIED SIZES

State or Region	Percent or benefits accruing to allotments										
	Under 5 acres	Under 10.1 acres	Under 15 acres	Under 30 acres	Under 50 acres	Under 100 acres	Under 200 acres	200 acres and over	350 acres and over	500 acres and over	1,000 acres and over
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Alabama	6.5	24	37	57	69	82	92	8.4	3.7	1.8	0.3
Florida	9.8	34	49	73	84	94	99	1.3	0	0	0
Georgia	2.1	12	22	45	63	81	93	6.9	3.0	1.3	0
North Carolina	12.1	30	41	60	72	86	95	5.4	1.8	1.0	.3
South Carolina	5.6	18	28	47	61	79	92	8.3	3.1	1.6	.4
Virginia	37.7	66	77	88	94	98	100	0	0	0	0
Southeast	6.3	21	32	53	66	82	93	7.5	3.1	1.5	.3
Arkansas	0.9	5	11	24	34	50	65	34.5	21.9	15.1	6.6
Illinois	3.5	15	19	31	48	67	90	10.4	0	0	0
Kentucky	4.9	14	21	37	58	78	96	3.9	0	0	0
Louisiana	1.6	9	17	32	44	60	76	24.3	12.0	6.8	1.7
Mississippi	2.4	10	17	29	37	48	63	37.3	22.7	14.1	5.2
Missouri	2.1	7	15	34	50	71	85	14.5	8.7	5.1	1.7
Tennessee	7.3	26	40	61	74	88	95	4.6	1.6	.6	0
Delta	2.5	11	18	32	43	56	71	29.4	17.7	11.3	4.3
Oklahoma	.5	3	8	28	52	82	95	4.9	1.8	.7	.2
Texas	.2	1	2	9	21	49	76	24.3	11.0	6.6	2.7
Southwest	.2	1	3	10	23	51	77	23.1	10.4	6.2	2.6
Arizona	.1	1	2	5	9	21	40	60.0	39.4	30.5	14.5
California	.2	1	4	9	15	29	48	52.4	39.3	33.2	21.5
New Mexico	1.0	4	8	21	38	65	83	16.8	7.6	5.5	1.5
West	.2	1	3	9	16	30	49	51.0	36.3	29.8	17.7
United States	2.0	8	13	25	36	54	71	28.6	17.1	12.1	5.9

Source: "1964 Upland Cotton: Final Planted Acres and Number of Farms Planting Cotton by Size of Effective Allotment," USDA, ASCS, Policy and Program Appraisal Division, mimeo, Nov. 6, 1964 (2 pages).

TABLE 9.—DISTRIBUTION OF 1964 UPLAND COTTON PRICE SUPPORT BENEFITS: PROPORTION OF U.S., REGIONAL, AND STATE BENEFITS RECEIVED BY VARIOUS PERCENTILES OF FARMER BENEFICIARIES [1]

State	Percent of total benefits received by the—									
	Lower 10 percent of farmers	Lower 20 percent of farmers	Lower 33 percent of farmers	Lower 50 percent of farmers	Top 50 percent of farmers	Top 33 percent of farmers	Top 20 percent of farmers	Top 10 percent of farmers	Top 1 percent of farmers	Gini concentration ratio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Alabama	2.1	4.3	8.0	17	83	73	60	45	15	.546
Florida	2.7	5.3	8.8	19	81	69	54	37	10	.483
Georgia	1.2	3.0	8.0	16	84	71	58	42	11	.531
North Carolina	2.5	4.9	8.2	13	87	76	64	47	15	.577
South Carolina	1.7	3.3	5.6	13	87	77	63	48	13	.594
Virginia	5.2	10.4	17.3	26	74	65	52	37	11	.401
Southeast	1.9	3.7	6.2	15	85	75	61	47	14	.571
Arkansas	0.8	2.6	5.2	11	89	80	70	56	20	.652
Illinois	1.2	2.4	4.9	11	89	83	71	53	12	.650
Kentucky	1.5	3.0	5.0	11	89	80	66	47	10	.613
Louisiana	1.0	2.8	6.4	12	88	79	69	54	16	.628
Mississippi	1.0	2.1	4.9	9	91	84	75	64	23	.701
Missouri	1.3	3.0	6.5	14	86	74	61	44	14	.565
Tennessee	2.4	4.8	9.0	18	82	72	58	42	13	.515
Delta	1.2	2.3	5.9	11	89	81	70	58	21	.657
Oklahoma	1.1	3.7	9.6	21	79	65	50	31	7	.446
Texas	.4	2.0	6.4	15	85	71	56	37	10	.530
Southwest	.5	2.0	6.3	14	86	73	56	39	11	.542
Arizona	.5	1.5	4.1	10	90	80	65	47	15	.628
California	.7	1.9	4.2	8	92	84	72	57	25	.686
New Mexico	.8	2.4	5.7	14	86	75	60	42	11	.565
West	.5	1.6	3.9	8	92	84	72	56	22	.682
United States	.9	1.8	4.9	10	90	80	69	53	21	.653

[1] This table presents portions of 2 Lorenz curves relating the cumulated percentage distribution of benefits to the cumulated percent of farmers receiving those benefits. Cols. 1 through 4 summarizes this relationship cumulated up from the lower (benefit per farmer) end of the curve, and cols. 5 through 9 summarizes the relationship cumulated down from the top (highest benefit per recipient) end of the curve.

Sources: (a) "1964 Upland Cotton: Final Planted Acres and Number of Farms Planting Cotton by Size of Effective Allotment," USDA, ASCS/Policy and Program Appraisal Division, mimeo, Nov. 6, 1964 (2 pages); (b) "Agricultural Statistics, 1966," USDA, 1966, p. 62. Prices from this source were used in computing State value of production figures for use as weights in combining the distributional data from source (a); (c) "Crop Production, 1965 Annual Summary," USDA, SRS, Dec. 20,1965, p. 84. Yield and acreage data from this source were used in computing State value of production figures for use as weights in combining the distributional data from source (a).

In the case of wheat and feed grains there was also an opportunity to observe the consequence of the recent addition of direct payments to the price support system allowing price supports to be dropped approximately to world prices. Economists have long argued that:

> If the avowed purpose is to redistribute income from those with higher incomes to those with lower incomes, then a program that directly provides income or services to the intended beneficiaries is more efficient and more effective than one that raises the demand for some product they, and others, produce [10].

Agricultural economists have argued for years that this is the better way to approach the income problem in agriculture. However, it obviously does not happen automatically when one shifts to a direct payment scheme. As one can see in Table 7, in the case of the feed grain program in 1964, direct payments did indeed reduce slightly the overall concentration of program benefits from that which would have prevailed if there had been only a price support program. However, in the case of wheat the direct diversion payment actually increased the Gini ratio on the total benefit distribution above that for price supports—though not significantly. It is obvious that if you design a direct payment distributional system in much the same manner as the price support program, the distributional consequences will be approximately similar. In agriculture today the capacity of the direct payment instrument for redistribution is clearly not being used. Rather, the distributional ground rules of the price support system are being imposed also on the direct payments.

Since the distribution of government payments are both very visible (unlike price support benefits) and highly concentrated, efforts have been made in Congress in recent years to place a limit on the size of the total payment that a single farmer may receive.

Recently available data on the distribution of total payments in 1967 make it possible to examine the concentration of total government payments and to test the distributional impact of a payment limitation [13]. A total of $3.1 billion in government payments were made in 1967. Eighty percent of this total went to three crops: $932 million to cotton, $865 million to feed grains and $731 million to wheat. The only other commodities with direct payments were sugar and wool which received $70 million and $29 million respectively. All the remaining $439 million went into conservation (ACP, $225 million) and land withdrawal [15].

In Table 10 it can be seen that the distribution of total payments in 1967 was highly concentrated exhibiting a concentration ratio of 0.671 This exceeds the concentration in every commodity benefit distribution in Table 7 except sugar cane. It falls between the concentration exhibited in Table 10 by gross and net farm income in 1967—which is what one might logically expect.

TABLE 10.—DISTRIBUTION OF 1967 GOVERNMENT PAYMENTS AND FARM INCOME: PROPORTION RECEIVED BY VARIOUS PERCENTILES OF FARMERS[1]

1967	Percent of income received by the—						
	Lower 20 percent of farmers	Lower 40 percent of farmers	Lower 60 percent of farmers	Top 40 percent of farmers	Top 20 percent of farmers	Top 5 percent of farmers	Gini concentration ratio
1967 total government payments:							
With no limitation on size of total payment[2]	1.1	5.7	13.3	86.7	69.0	42.4	0.671
Assuming $25,000 limitation[3]	1.1	6.0	14.1	85.9	67.2	39.0	.652
Assuming $10,000 limitation[3]	1.2	6.5	15.3	84.7	64.4	33.8	.623
Various measures of farmer income in 1967:							
Gross receipts from farming[4]	1.6	3.3	10.1	89.9	72.3	40.4	.693
Realized net farm income[5]	4.5	9.0	19.3	80.7	50.0	26.2	.541
Nonfarm income of farmers	25.5	51.0	70.1	29.9	15.6	5.6	.125
Total income of farmers	14.9	29.8	44.5	55.5	37.0	16.0	.211

[1] This table presents portions of 2 Lorenz curves relating the cumulated percentage distribution of benefits to the cumulated percent of farmers receiving those benefits. Columns 1 through 3 summarize this relationship cumulated up from the lower (benefit per farmer) end of the curve, and columns 4 through 6 summarize the relationship cumulated down from the top (highest benefit per recipient) end of the curve.

[2] Government payments to farmers as actually distributed in 1967. Total payments were $3,100,000,000.

[3] Assumes all 1967 beneficiaries continue to participate in programs and are eligible for payments. Under the $25,000 limit payments would total $2,800,000,000 and under the $10,000 limit $2,600,000,000.

[4] Including Government payments and imputed nonmoney income from farm products consumed at home and from the rental value of the farm dwelling.

[5] Net of farm production expenses and changes in farm inventories of livestock and crops.

Sources: Computed from data in "Farm Income Situation," USDA, FIS-211, July 1968, pp. 68-69, except direct payment data which are from the Congressional Record, July 31, 1968.

If one assumes a $25,000 limit were imposed on total payments to a single beneficiary in 1967, the concentration of the overall distribution of benefits declines only from a Gini ratio of 0.671 to 0.652 and total payments decline from $3.1 to $2.8 billion. If one assumes a $10,000 limitation, the concentration ratio declines only to 0.623 and total payments decline to $2.6 billion. This suggests that even quite low payment limitations are unlikely to have any revolutionary impact on the actual concentration of total payments made under present direct payment eligibility rules. It suggests as a consequence that any decision to limit payments should be based primarily on (1) the politics rather than the economics of equity between farmers, (2) the politics and economics of equity between large payment beneficiaries and taxpayers (the $10,000 limit creates a potential transfer of $0.5 billion), and (3) consideration of the impact of a payment limitation on increased federal budget costs for production control that could result from a decline in program participation by large producers.

Changes in program design such as have occurred since 1961, shifting emphasis from price supports towards direct payments, have considerable effects upon the distribution of the costs of the program also. While no quantitative analysis is attempted here it is clear from the logic of the situation that the cost of the price support operation is borne through the taxes necessary to sustain the storage and control operation, and also through consumers who pay some of the cost through higher market prices. The shift toward direct payments shifts this relative burden away from the consumer and toward the taxpayer.

One exception to this is the wheat program where direct payments are generated by a certificate system that is paid eventually by the consumer through the market. The price support approach, in which a higher proportion of the costs are borne by the consumer in the form of food costs, has a far greater impact on the low income consumer than would be the case of direct payments in which the cost is borne primarily by a progressive income tax. Thus, price support-acreage diversion programs can be described as doubly regressive—that is, a major share of the cost is borne by consumers with below average incomes and a major share of the benefit is received by farm producers with above average incomes. The shift toward direct payments, while not significantly affecting the distribution of benefits among farmers, has reduced the regressiveness of the distribution costs.

Conclusion

This paper has argued that distributional consequences of public decisions have major and increasingly important impacts in our society. We have a long history of public commitment to equity purposes. Yet we know very little about the social processes by which distributional impacts are institutionalized and are filtered through the society. We know even less about how to redesign distributional systems without incurring excessive political and social cost. In fact, we do not at this point even have good descriptions of the distributional impacts or characteristics of our public programs. What little we do know suggests, as in the three cases just presented, that there are many surprising and apparently perverse distributional impacts.

Both the integrity of our many public commitments in equity and our efficiency in the use of tax monies to attain public ends require far greater effort to collect data for program analysis of the distributional impacts of public decisions.

The three cases—higher education expenditures, Bureau of Reclamation investments, and the farm program—contain several lessons. If we do not systematically attempt to collect data and assess distributional impacts, we shall always be surprised by the many unintended consequences of our public decisions. Even if we accept the original objectives of older programs, we fail to attain our ends because we have failed to keep the equity objectives clearly before us. The use of indirect means, such as price supports, is a tricky and uncertain way of attaining equity objectives.

Finally, past decisions made without adequate distributional knowledge now appear (given our objectives) often to lack in economic and social wisdom. But even more importantly, these past decisions cannot now be easily or cheaply reversed. This is particularly true in light of the irrevocable specialized capital investments in the reclamation program and the past farm program benefits that are now capitalized into the cost structure of agriculture. The cost of our past ignorance of distributional impacts is clearly high. There is no need to persist in such error. But we must now collect the data and do the analysis of the distributional impacts that are needed for today's decisions.

LITERATURE CITED

[1] American Economic Association, *Index of Economic Articles*, vols. 1–7, Homewood, Ill., Irwin, 1961–67.

[2] Back, W. B. and John E. Waldrop, Jr., ed., *Regional Studies of Income Distribution*, Louisiana Agricultural Experiment Station and U.S. Department of Agriculture, Baton Rouge, June, 1966.

[3] Bonnen, James T., "The Distribution of Benefits from Cotton Price Supports" in Samuel B. Chase Jr., ed., *Problems in Public Expenditure Analysis*, Washington, D.C., The Brookings Institution, 1968, pp. 223–48.

[4] Bonnen, James T., "The Distribution of Benefits from Selected U.S. Farm Programs", *Rural Poverty in the United States: A Report of the President's National Advisory Commission on Rural Poverty*, Washington, D.C., GPO, 1968, pp. 461–505.

[5] Boyne, David H., "Changes in the Income Distribution in Agriculture", *Journal of Farm Economics*, vol. 47, No. 5, December 1965, pp. 1213–24.

[6] Bureau of the Census—Economic Research Service, "Preliminary Estimates of the Fram Population of the United States: 1967," *Current Population Reports. Farm Population Series P–27*, No. 38, April 2, 1968.

[7] Bureau of the Census, U.S. Department of Commerce, "Estimates of the Population of the United States to December 1, 1968", *Current Population Reports: Population Estimates, Series P–25*, No. 413, January 17, 1969.

[8] Bureau of the Census, U.S. Department of Commerce, *Historical Statistics of the United States: Colonial Times to 1957*, Washington, D.C., 1962, p. 9.

[9] Bureau of the Census, U.S. Department of Commerce, *Historical Statistics of the United States, Colonial Times to 1957: Continuation to 1962 and Revisions*, Washington, D.C , 1965, p. 1.

[10] Capron, William M. "Comments" in Samuel B. Chase, Jr., ed., *Problems in Public Expenditure Analysis*, Washington, D.C., The Brookings Institution, 1968, p. 250.

[11] Coffey, Joseph D. "Personal Distribution of Farmers' Income by Source and Region," *American Journal of Agricultural Economics*, vol. 50, No. 5, December 1968, pp. 1383–1396.

[12] *Congressional Record*, vol. 109, pt. 15, 88th Congress, first session (1963), pp. 19796–19798.

[13] *Congressional Record*, vol. 114, No. 135, 90th Congress, second session (1968), p. H7928.

[14] Council of Economic Advisers, *Annual Report of the Council of Economic Advisers*, Washington, D.C., January 1969, pp. 244, 250.

[15] Economic Research Service, U.S. Department of Agriculture, *Farm Income Situation* (FIS–211), July 1968, pp. 64, 68–9.

[16] Freeman, A. Myrick III, "Six Federal Reclamation Projects and the Distribution of Income," *Water Resources Research*, 1967, vol. 3, No. 2.

[17] Gillespie, W. Irwin. "The Effect of Public Expenditures on the Distribution of Income," in Richard A. Musgrave ed. *Essays in Fiscal Federalism*, The Brookings Institution, 1965, pp. 112–183.

[18] Hansen, W. Lee and Burton A. Weisbrod, *Benefits, Costs and Finance of Public Higher Education*, Chicago, Markham Publishing Company, 1969 (in process).

[19] Hansen, W. Lee and Burton A. Weisbrod, "The Distribution of Costs and Direct Benefits of Public Higher Education: The Case of California," *Journal of Human Resources*, Spring 1969, pp. 176–191.

[20] Haveman, Robert H. *Water Resource Investment and the Public Interest*, Nashville, Vanderbilt University Press, 1965.

[21] Howe, Charles W. and K. William Easter, *Interbasin Water Transfers—Economic Issues and Impacts*, Resources for the Future, Johns Hopkins University Press, Chapter IV, (publication planned for 1969).

[22] Kravis, I. B., *The Structure of Income: Some Quantitative Essays*, Philadelphia, 1962.

[23] Mishan, E. J., "A Survey of Welfare Economics, 1939–1959," in Mishan, ed., *Welfare Economics*, New York, Random House, 1964.

[24] Morgan, James N., Martin H. David, Wilber J. Cohen, and Harvey E Brazer, *Income and Welfare in the United States*, New York, McGraw Hill, 1962, pp. 15–24.

[25] *Parity Returns Position of Farmers*, Senate Document 44, 90th Congress, first session, August 10, 1967, p. 22.

[26] Pfaff, Martin, *The Transfer Economy: The Relationship Between Unilateral and Bilateral Transfers in the National and World Economy*, The Brookings Institution Economic Studies Seminar (mimeo), Washington, D.C., October 28, 1968, pp. 40–42.

[27] Peacock, Alan T., ed., *Income Redistribution and Social Policy*, London, Jonathan Cape, 1954, pp. 178–267.

[28] Ruttan, Vernon W., *The Economic Demand for Irrigation Acreage*, Baltimore : The Johns Hopkins Press, 1965, p. 49.

[29] Tolley, George S., "Reclamation Influence on the Rest of Agriculture," *Land Economics*, vol. 35, May 1959, pp. 176–80.

[30] Weintraub, Sidney, *An Approach to the Theory of Income Distribution*, Philadelphia, Chilton, 1958.

[31] Weisbrod, Burton A. and W. Lee Hansen, "An Income-Net Worth Approach to Measuring Economic Welfare," *American Economic Review*, vol. 58, No. 5, Part 1, December 1968.

[32] Wollman, Nathanial, *The Value of Water in Alternative Uses with Special Application to Water Use in the San Juan and Rio Grande Basins of New Mexico*, Albuquerque, University of New Mexico Press, 1962, p. 31.

[33] Young, Robert A. and William E. Martin, "The Economics of Arizona's Water Problem," *Arizona Review*, vol. 16, No. 3, March 1967, p. 10.

THE ABSENCE OF PROGRAM EVALUATION AS AN OBSTACLE TO EFFECTIVE PUBLIC EXPENDITURE POLICY: A CASE STUDY OF CHILD HEALTH CARE PROGRAMS

BY JOSEPH S. WHOLEY [1]

Joseph S. Wholey is Director of Program Evaluation Studies at the Urban Institute, and the former Staff Director of the Program Analysis Group on Child Health Care at the Department of Health, Education, and Welfare.

The failure of many government agencies to perform appropriate analysis of their past activities seriously inhibits the effective design of new programs or new program directions. This is especially true in those programs which are of an avowedly experimental nature. In his paper, Mr. Wholey discusses the findings of the HEW Program Analysis Group in its effort to apply PPB analysis to child health-care programs. In particular, he stresses "the impact (on program recomendations and program decisions) of our lack of knoweldge of the effects of existing maternal and child health programs. * * * In particular, little information was available on the effectiveness of existing Federally supported maternal and child health programs. * * * As a result * * *, decisions on possible expansion of these programs were bound to depend almost entirely on 'professional judgment.'" The general lack of analytical information on health problems and programs and the lack of program evaluation in particular hinder the process of new program design and reduce the persuasiveness of arguments for program changes.

Mr. Wholey concludes that "the whole structure of the Federal Government's Planning-Programing-Budgeting System suffers from a crucial weakness: the lack of a system of program evaluation which would provide measures of effectiveness related to program objectives. In the absence of knowledge of the effects of past programs, analytical studies will often fail to influence Federal decisionmaking, or may succeed in having such influence when they shouldn't." To correct this situation, he offers some recommendations for improving the economic evaluation of government's social programs.

I. *Introduction and Summary*

By *program evaluation* we mean assessment of the impact of past and present programs, projects, and project components ("treatments"). Program evaluation attempts to determine what the government and people have been getting for their money: what works? what doesn't work? what have been the effects of past and present programs and projects? what factors are associated with success or failure of programs and projects?

A large number of major social programs ought to be viewed as quasi-experiments. They were designed on the assumption that certain courses of action would improve health, raise employment, improve housing, or the like.* Generally, however, the Federal Government has

[1] This paper is based on the author's work as staff director of the HEW Program Analysis Group on Child Health Care. The author thanks Drs. Ruth Covell, Arthur J. Lesser, Arthur Levin, David Seidman, and George A. Silver, Mr. Donald Kummerfeld, and Mr. J. Michael Stern, who reviewed earlier drafts of this paper and made a number of helpful suggestions. Responsibility for any errors of fact or interpretation rests with the author.

*Further discussion of this issue is found in the papers by Rivlin, Levine, Grosse, and Brandl in vol. 3 of this collection.

made no real attempt to evaluate the effectiveness of its social programs or of local projects within these programs.[1] Too often relevant "output measures," in terms of which programs could be evaluated in a meaningful manner, remain unidentified. Often "evaluation" has meant only internal program monitoring or the preparation of self-justifying progress reports.

Evaluating the effectiveness of social programs is a complex problem. It is usually time-consuming and expensive to find out the effects of a particular project or project component ("treatment"), and to separate the effects of that project or component from the effects of other forces in the environment. Such efforts are needed, however, to allow us to expand effective programs and projects and weed out ineffective ones, to predict the likely consequences of alternative decisions on future programs, and to assist implementation of more rational decision on future programs.

This paper, a case study of one of the early attempts to apply the planning-programing-budgeting system to social programs, is based on the author's experience as staff director of the Department of Health, Education, and Welfare Program Analysis Group on Child Health Care: an analytical group established in 1966 to help set priorities for the Administration's 1967 legislative proposals and fiscal year 1968 budget proposals for child health programs. The paper discusses HEW analyses of proposed maternal and child health programs in 1966, documenting some of the policy implications of the Federal Government's failure to evaluate the effectiveness of its maternal and child health programs prior to 1966, and recounts decisions on legislative proposals and budget requests for HEW maternal and child health care programs over the last three years. The paper concludes with recommendations on ways to increase the impact of program evaluation studies on Federal decision-making.

The paper is organized as follows: in section II, we discuss the work of the HEW program analysis group on child health care in formulating child health objectives and estimating national needs, reviewing existing maternal and child health programs, developing proposals for new and expanded maternal and child health programs, estimating the costs and effects of proposed programs, and developing conclusions and recommendations on child health programs. Sections III and IV trace executive and Congressional decisions on maternal and child health programs over the past three years and draw conclusions on the implications of the lack of program evaluation prior to 1966. In the final section, we suggest how more attention to program evaluation can improve the effectiveness and efficiency of existing Federal programs and improve decision-making on possible new programs.

II. HEW's Program Analysis Group on Child Health Care

This section discusses the genesis, operation, and output of the Program Analysis Group on Child Health Care, established by the Secre-

[1] Of late, however, there has been some limited progress in this area. For example, the Office of Economic Opportunity has devoted substantial funds to program and project evaluation; the Department of Labor has funded some good work in manpower program evaluation; the Department of Health, Education, and Welfare has funded some significant evaluations of compensatory education programs; and the Department of Housing and Urban Development has planned and is implementing a significant set of evaluation studies of the Model Cities Program.

tary of Health, Education, and Welfare in May 1966 to define explicit objectives for the health of children, examine the effectiveness of current programs in meeting these objectives, and estimate the costs of meeting these objectives through new or expanded programs.[2]

This program analysis group was one of the few analytical efforts that produced observable impact on the Administration's legislative program and proposed budget in the first year of the initiation of the Planning-Programming-Budgeting system throughout the executive branch of the Federal Government. The work of the Program Analysis Group on Child Health Care was among the four program analyses discussed by the President in his fiscal year 1968 budget message: "This analysis led to the legislative program focused on early identification and treatment of needy handicapped children and experimental projects aimed at improving delivery of medical care to children.[3]

In this section we will discuss briefly the establishment of the Program Analysis Group on Child Health Care; the work of the group in proposing national objectives in child health; the availability/lack of information on the effectiveness of existing maternal and child health programs; and the work of the program analysis group in analyzing the costs and effectiveness of proposed new (or expanded) maternal and child health programs. We will be particularly interested, in this section and in the next, in the impact (on program recommendations and program decisions) of our lack of knowledge of the effects of existing maternal and child health programs.

A. ESTABLISHMENT OF THE PROGRAM ANALYSIS GROUP

In 1966 President Johnson had publicly committed himself to include recommendations for new or expanded child health programs in his 1967 legislative proposals to Congress. To assist in formulating recommendations for what was to become the Child Health Act of 1967, the Secretary of Health, Education, and Welfare established a Program Analysis Group on Child Health Care. This program analysis group, one of several established in HEW's first year of operation with the new PPB system, consisted of staff members of the Office of the Secretary, the Children's Bureau, and the Public Health Service in HEW, as well as a number of physicians, economists, and systems analysts from within and outside of the government. The group was established on the recommendation of the HEW Assistant Secretary for Planning and Evaluation, who provided analytical support for the program analysis group.

The program analysis group was selected by the Secretary's office in a conscious attempt to apply analysis to establishing priorities among proposed maternal and child health care programs for fiscal year 1968. The program analysis group, which would report to Secretary Gardner, was asked to review existing federally supported child health programs and was given a broad mandate to propose new or expanded maternal and child health programs within a budget increment of $100 million for the next fiscal year.

---

[2] For another view on the study by the Program Analysis Group on Child Health care, see Levin, Arthur L. "Cost Effectiveness in Maternal and Child Health: Implications for Program Planning and Evaluation." *New England Journal of Medicine*, 278: 1041–1047. 1968.

[3] *The Budget of the United States*, 1968. U.S. Government Printing Office, Washington, D.C., p. 37.

Primarily addressing itself to the costs and effects of medical care programs, the program analysis group did not analyze important related questions of the effects on child health of changes in income level, environmental health programs, improved nutrition, communicable disease control, or accident prevention programs; nor did it go into the problems of the unequal geographical distribution of physicians, the need for strengthening obstetric and pediatric departments, or the status of State laws governing medical and nursing practice.

B. CHILD HEALTH OBJECTIVES AND NATIONAL NEEDS

At the time of the work of the Program Analysis Group on Child Health Care, there were no agreed-on national objectives relating to the health of children. Part of the work of the program analysis group was therefore devoted to development of some reasonable proposed national objectives for maternal and child health care programs.

The program analysis group determined that priority should be given to "health-depressed" areas, areas characterized by high infant mortality rates, high proportions of families in poverty, and high proportions of substandard housing. For simplicity, the program analysis group used the traditional best single indicator of community health status and defined a "health-depressed" area as an area within which the infant mortality rate is high. In urban areas, the program analysis group suggested, health-depressed areas could be defined as census tracts or combinations of contiguous census tracts; in rural areas, health-depressed areas would be counties with high infant mortality rates.

Lack of knowledge of the effects of existing child health programs was explicity recognized by the program analysis group even in its development of proposed child health objectives:

> This section proposes one general and three specific maternal and child health objectives * * *.
>
> 1. Make needed maternal and child health services available and accessible to all, in particular, to all expectant mothers and children in health-depressed areas.
>
> There is no universal index of good or bad health among children. Therefore, in looking at the problem of assuring needed health care, we necessarily primarily concern ourselves with some particular health problems which are highly prevalent, which are highly adverse, and which can be mitigated or even avoided given proper health care. Three such health problems are mortality (infant mortality, in particular), chronic handicapping conditions, and bad teeth. Health care directed toward these problems would also yield important benefits to the general health of the mothers and their children.
>
> Progress toward the above major objectives can therefore be measured, to some extent, by progress toward accomplishment of the following subordinate objectives:
>
> 2. Reduce numbers of chronic handicapping conditions, in particular reduce the incidence of preventable handicapping conditions and the prevalence of uncorrected handicapping conditions: in particular, congenital defects; mental retardation;

vision, hearing, and speech defects; and mental and emotional disorders.

3. Reduce infant mortality rates, particularly in "health-depressed" areas.

4. Reduce unmet dental needs, particularly in "health-depressed" areas.[4]

The major objective of maternal and child health care programs was, therefore, stated by the program analysis group in terms of increased services. Much of the analysis was carried on, however, in terms of the three subordinate objectives which were stated in terms of the effects to be achieved by improved services.

1. *Estimating the Universe of Need*

When it came time to estimate the numbers of children needing health care, lack of data on the size of populations in "health-depressed" areas (i.e., areas having high infant mortality rates) made it much more practical to use estimates of the poverty population as estimates of the universe of need for federally-supported child health care programs, inasmuch as areas having high infant mortality usually coincide with poverty areas. The program analysis group estimated, for example, that at least 800,000 mothers living in poverty required comprehensive prenatal services each year and that some 15 million children under age 18 were in the poverty group (poor children making up 21 percent of the child population and 43 percent of the population living in poverty). It was much less clear, however, what proportion of these mothers and children were receiving adequate health care services or any health care services.

On the basis of fragmentary data, the program analysis group estimated that approximately a third of the million women received care in maternity clinics under existing maternal and child health programs; the level of services provided to those women could not be determined. The program analysis group also had available reports that, in most of the major cities, one-third to one-half of the women delivered at city hospitals had received virtually no prenatal care at all.

On the basis of fiscal year 1964 data on State and local maternal and child health programs, the program analysis group estimated that, of the four-to-five million children under five years of age in the poverty group, only about a million and a half received services supported by State and local maternal and child health care programs. The most useful estimates of level of services provided came through rough calculations of the number of dollars expended in relation to the number of children receiving services under these programs (see section C below).

2. *Mortality*

The program analysis group relied mainly on 1964 data on infant mortality for the United States and other countries. Infant mortality rates by race, divided into deaths in the first month and deaths between one month and one year, were available for each county in the

[4] U.S. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Program Coordination. *Program Analysis: Maternal and Child Health Care Programs.* Washington, D.C., Government Printing Office, 1966, pp. II.1–II.3. (This document is hereafter referred to as the *Report of the Program Analysis Group on Child Health Care.*)

United States. Disparities between the United States infant mortality rate (especially the United States nonwhite infant mortality rate) and rates for other countries, as well as disparities among infant mortality rates within the United States, provided indications of unmet needs for maternity and infant care.

3. *Chronic Conditions*

The program analysis group concluded that, "Except for data from studies of congenital malformations, little hard data exist on the incidence and prevalence of chronic conditions in childhood. Populations examined (and conditions counted) in the major studies are usually not comparable. Neither are definitive data available on the extent of disability caused by chronic illness." [5] Based on the major studies undertaken over the past decade, the program analysis group estimated that between 20% and 40% of children suffer from one or more chronic conditions and that only 40% of such conditions are under treatment in low-income areas. Consistent with the latter estimate were Children's Bureau data that 19% of those examined in school health programs were referred for further examination or treatment but only 8% of those examined completed referral.

The program analysis group made estimates of numbers of children with congenital malformations, vision problems, hearing problems, psychiatric problems, and other medical problems based on a wide variety of published and unpublished studies and statistics from public health programs. The difficulty of this task of estimating the prevalence of chronic handicapping conditions can be gathered from the following statement in the program analysis group report: "Amblyopia (blindness in one eye) now occurs in 2–3% of the population (*estimates actually range from 0.5% to 8%*) * * *" [emphasis added].[6]

A good deal of the analysis in the program analysis group report was based on the finding that approximately 15% of 18-year old non-college-bound youths are chronically handicapped to the extent that they fail the Selective Service medical examination.[7]

4. *Unmet Dental Needs*

Since almost all children require some dental care each year to clean and fill teeth and for necessary care of gums, National Health Survey data (by race, income, and region) on the percentage of children with no visit to the dentist within a year furnished adequate information for the subsequent analysis. Approximately 45% of children aged 5 to 14 had no visit to a dentist within the year June 1963 to June 1964. The percentages were much higher for children from the South (60%) and for children from families with incomes under $2,000 per year (79%).

C. EXISTING MATERNAL AND CHILD HEALTH CARE PROGRAMS

At the time that the Program Analysis Group on Child Health Care was in existence, the Federal Government had increased its expenditures for maternal and child health care from approximately

---

[5] *Report of the Program Analysis Group on Child Care*, p. II.7.

[6] *Ibid.*, p. II–10.

[7] Preliminary data based on rejection rates in special Selective Service examinations of 18-year-old non-college-bound youth (July 1964–December 1965) under the "Conservation of Manpower" program (Source: Dr. Bernard Karpinos, Office of the Surgeon General, Department of the Army).

$150 million in 1965 to more than $400 million in fiscal year 1967, mainly under the title XIX (medicaid) program, the Office of Economic Opportunity's Neighborhood Health Centers program, the HEW Children's Bureau grant-in-aid programs to States for maternal and child health care and for crippled children's services, and under the new Children's Bureau project grant programs for maternity and infant care and for comprehensive health care for school and preschool children. The program analysis group estimated total Federal, State, and local expenditures for maternal and child health care at more than $480 million for fiscal year 1965, but was unable to provide more current estimates of total expenditures.

This section describes the extent to which the program analysis group was able to determine what services were provided and what were the effects of these services under existing federally supported maternal and child health programs.

1. *Services provided*

The program analysis group reviewed the sixteen major ongoing federally supported programs providing material and child health care, including four maternal and child health programs administered by HEW's Children's Bureau; the Title XIX (Medicaid) program of HEW's Welfare Administration;[8] State and local school health programs; the Office of Education's program under Title I of the Elementary and Secondary Education Act; the Office of Economic Opportunity's Head Start and Neighborhood Health Centers programs; the Public Health Service's Indian Health program, immunization program, and referral program for Selective Service medical rejectees; and HEW's Vocational Rehabilitation program. Working in early fiscal year 1967, the program analysis group was able to get fairly accurate estimates of Federal expenditures for these programs in fiscal years 1965, 1966, and 1967. Total Federal, State, and local expenditures for these programs could be estimated for all of these programs for fiscal year 1965 but, because of lags in State reporting, could be estimated for only some of the programs for subsequent years. It was much more difficult to get reports or current estimates of the numbers of mothers and children being served in these programs, and usually impossible to get detailed data on the services provided under the programs. The program analysis group was forced to rely on fiscal year 1964 data on the numbers of mothers and children receiving various types of services in the State and local maternal and child health programs and in the State Crippled Children's programs (both of which receive substantial Federal matching support).

Although the problem of estimating what numbers of children were receiving what services under Federal, State, and local maternal and child health programs proved to be formidable, the rough estimates presented were useful for the development of program recommendations.

The program analysis group arrayed available expenditure data against the latest available data on services provided and estimated numbers of mothers and children served under these programs, and derived *average yearly expenditure per person served* as a rough indicator of the extent of services provided. It turned out, for example,

[8] Now the HEW Social and Rehabilitation Service.

that State and local *maternal and* child health programs spent approximately $2 per year for each child under age 18 (approximately $10 per poor child under age 18), State and local school health programs spent approximately $3 per child of school age, medical vendor payments for care of AFDC children had been approximately $18 per year per child but would certainly be much higher under the new Title XIX (Medicaid) program, while per capita costs of the new OEO Neighborhood Health Centers ranged from $100 to $135 per person (all ages).

The available information on services provided, numbers of children served, and costs of services provided made it clear that not many poor children (or expectant mothers) were receiving standard (much less, comprehensive) health care and that there were extensive failures to provide follow-up diagnostic services and treatment even after potential or actual health problems had been identified in children. The most extensive services for poor children were concentrated in the first year of the child's life: 15% of infants, but only 4% of children aged 1–4, were examined in well-child conferences under State and local maternal and child health programs. The program analysis group estimated, for example, that—

> Currently, approximately one-half of the 6 and 9 year-olds receive vision screening through school health programs, and hearing screening is received by one-half of the 6 year-olds and one-third of the 9 year-olds. Most of these programs do not have effective referral mechanisms, nor are funds available for paying for treatment. At the present time, 3 year-olds are virtually unreached * * *. Almost 10% of preschoolers have vision problems (most of them preventable or correctable); approximately 3% have hearing problems, half of them correctable or preventable.[9]

The program analysis group concluded that—

> Taken together, present programs fall short of providing adequate health care for those mothers and children who live in poverty or for those children who are physically handicapped: the majority of low-income mothers fail to receive adequate maternity care; most children of low-income families go without adequate preventive or remedial health care (even after handicapping conditions have been identified in screening programs).[10]

2. *Program Effectiveness in Improving Child Health*

The health care system does not generally measure the effects of services provided. In particular, little information was available on the effectiveness of existing federally supported maternal and child health programs while the program analysis group was at work in 1966. Reporting systems for the State grant-in-aid programs yielded some summary statistics on services provided but no information on the effects of those services in improving the health of children. For many of the new Federal project grant programs, current information on services provided was scanty; information on program effectiveness was not yet available.

As a result of the lack of data on effectiveness of existing maternal and child health programs, decisions on possible expansion of these

[9] *Report of the Program Analysis Group on Child Health Care*, p. III.25.
[10] *Ibid.*, p. A.2.

programs were bound to depend almost entirely on "professional judgment."

Little information was available either through routine reporting or through program evaluation studies, on the degree to which existing federally supported maternal and child health programs (individually or collectively (reduced number of chronic handicapping conditions, reduced infant mortality rates, or reduced unmet dental needs. There were essentially no data available on the effectiveness of the existing maternal and child health care services even in those programs where such effectiveness data might reasonably have been anticipated, for example, in the Children's Bureau's new maternity and infant care projects (reduction of infant mortality) or in the State Crippled Children's programs (reduction of chronic handicapping conditions).

It is possible that, because of the difficulty of establishing relationships between health care services and improved health status, evaluation of program effectiveness was considered beyond the province of Federal health care program administrators. For whatever reasons, evaluation of program or project effectiveness was low on the list of priorities of Federal program administration. The Federal agencies involved had no overall plan for evaluation of child health programs nor for evaluation of the effectiveness of different projects, project components ("treatments"), and combinations of components.

One possible exception to this bleak picture *might* have been the Children's Bureau program of maternity and infant care projects. The Children's Bureau maternity and infant care projects represented the largest-scale maternal and infant health care program attempting to provide comprehensive care in 1966. "The objectives of these programs are the reduction in maternal and infant mortality and morbidity and taking steps which will assist communities in so organizing their maternity and infant care services as to increase the accessibility of care, to improve the quality of care and to make use of the best available resources in providing comprehensive maternity and infant care for low-income high-risk patients." [11] The maternity and infant care projects had been established in the spring of 1964, and approximately $70 million was spent on these projects in fiscal years 1965 through 1967.

The Children's Bureau had established a reporting system for the new maternity and infant care program. Maternity and infant care projects completed both the standard accounting reports and research reports designed ultimately to yield information on the effectiveness of the maternity and infant care projects. The maternity and infant care program was still new and reporting requirements had not been enforced, however; hence, little information was available in 1966 on the effectiveness of the maternity and infant care projects in reducing infant mortality and mental retardation.

The program analysis group concluded, "Past and present maternal and child health care programs have lacked systematic procedures for evaluation of effectiveness. Data are lacking, for example, on the effec-

[11] *Comprehensive Obstetric Care Program and the Child Health Act of 1967*, by Arthur J. Lesser, M.D., Deputy Chief, Children's Bureau, Department of Health, Education, and Welfare.

tiveness of these programs in reducing mortality and in preventing disability." [12]

D. PROPOSALS FOR NEW AND EXPANDED MATERNAL AND CHILD HEALTH PROGRAMS

The initial discussions of the Program Analysis Group on Child Health Care were general and wide ranging, as are the discussions of most task forces. Early and later discussions focused attention on a number of new programs which were analyzed and finally recommended in the program analysis group report and in the Administration's proposed Child Health Act of 1967. Among such programs were comprehensive maternal and child health care programs (including innovative programs using physician assistants and other new types of health aides), programs to screen children for birth defects and other chronic handicapping conditions and provide the treatment required for these defects, and expansion of family planning programs.

This section and the next review the analyses completed and the resulting program recommendations developed by the program analysis group. Section III discusses decisions on national and child health programs by the Department of Health, Education, and Welfare, the President, and Congress, upon completion of the program analysis group report and over the following two years.

1. *Programs Considered*

The program analysis group estimated the costs, manpower requirements, and effects of fourteen proposed maternal and child health care programs, grouped under seven major headings: (1) comprehensive maternal and child health care programs, (2) programs to provide early case-finding and treatment of congenital and other chronic handicapping conditions in children, (3) a program to provide early case-finding and treatment of vision and hearing defects, (4) programs to reduce unmet dental needs, (5) intensive care units for high-risk premature and other infants, (6) a program to provide treatment for Selective Service medical rejectees, and (7) a program of support for expanded family planning services. To allow direct comparison of program costs and effects, whenever possible, the estimated costs and effects of the fourteen proposed programs were examined in a hypothetical "health-depressed" (low-income) urban or rural community of 50,000 persons, including 1,000 expectant mothers, 1,000 infants, 1,000 one-year-olds, * * *, and 1,000 18-year-olds.

2. *Estimating Costs and Manpower Requirements*

This paper will not give extensive attention to the difficult task of estimating the costs and the manpower required to provide services under each of the fourteen programs considered. This task was done in far greater depth than is done by the usual task force. As a result, the Office of the Secretary was in a far better than usual position to defend the HEW budget requests and legislative proposals.

[12] Since 1966, infant mortality rates have been dropping both for the nation as a whole and in low income areas. As Dr. Lesser states, "While the reasons for this dramatic change in infant mortality have not been established in a cause and effect relationship, the only new contributory factors which have been identified are the rapid increase in family planning among the poor and comprehensive maternity programs focused on the most vulnerable population." (Lesser, *op. cit.*, p. 12.)

One of the major conclusions of the program analysis group report was the following conclusion on manpower requirements: "With current methods of delivery of health care, comprehensive health care for the 15 million poor children would require the services of more doctors than the total number of pediatricians practicing in the country today. With existing, or anticipated, numbers of physicians, comprehensive child health care will not be feasible on a wide scale unless far more efficient use of physicians is achieved." [13]

A key manpower question that could not be resolved on the basis of available data was the question of the extent to which physicians' assistants (new types of health professionals) could replace (some of) the services of physicians and hence allow a physician to furnish high quality medical care to greater numbers of patients. One of the major program analysis group recommendations was, therefore, the recommendation that a number of comprehensive maternal and child health care pilot (demonstration) projects ought to be established, to test the feasibility, costs, and effectiveness of new ways of providing children with comprehensive health services and new ways of training and using skilled physician assistants.

3. *Cost-Effectiveness Analysis of Proposed Programs*

To provide a common analytical framework, the program analysis group compared the estimated costs, estimated manpower requirements, and estimated effects of each of the fourteen proposed programs in a hypothetical low-income urban or rural community ("health-depressed" area) of 50,000 people and then compared the cost-effectiveness of these programs in accomplishing each of the child health objectives discussed in section II.B. As the program analysis group stated, "All of the estimates of program effectiveness were made on the bases of imperfect knowledge (often only informed estimates) of the proportion of the population needing care, the proportion of the population not getting adequate care, and the effects of adequate care on health (in particular, the effects on infant mortality and on handicapping conditions) * * *. The interrelationships among the effects of environment, education, and medical care make it very difficult to predict the improvements in health which would result from improvements in the delivery of health care. The results of health programs are usually not ascertainable for some years." [14]

Lacking information on national objectives for maternal and child health care, on the extent of the unmet need for such care, and on the impact of past programs of maternal and child health care (to greater or lesser degree), the program analysis group proceeded to structure the problem, assume some reasonable objectives toward which progress could be measured, and use the medical literature and professional judgment of a small number of physicians on the staff of the program analysis group to estimate the effectiveness of expansion of existing programs and mounting of certain new maternal and child health programs. The report made a contribution to defining national objectives for maternal and child health care, structuring the problem of analyzing alternative child health care programs, developing a common framework of cost data for the programs

[13] *Report of the Program Analysis Group on Child Health Care*, p. V.2.
[14] *Ibid.*, pp. III.2 and III.1.

considered, and making some estimates of the effects of the programs considered in meeting the objectives assumed by the program analysis group.

The program analysis group was able to establish that some of the programs suggested were infeasible because of personnel requirements: There simply were not enough pediatricians in the country, for example, to furnish comprehensive health care to all of the children in the country, not even enough to furnish comprehensive health care to all of the poor children in the country. The report took explicit account of the uncertainty in much of the data presented on program effectiveness when it recommended in the area of comprehensive health care that a number of experiments be undertaken to learn more about the feasibility, costs, and effectiveness of providing comprehensive maternal and child health care through use of obstetric assistants, pediatric assistants, and other types of health aides.

Even the fragmentary data available for the program analysis group report were sufficient, however, to support a strong recommendation for a major program of early case-finding and treatment of chronic handicapping conditions. Reasonable estimates of costs and effectiveness of comprehensive health care programs and the program of periodic examinations, screening, and follow-up treatment for potentially or actually handicapping conditions, together with information on the lack of sufficient physicians to mount wide-scale programs of comprehensive health care for children, formed the basis for convincing arguments for a "gap filler" program of screening and follow-up treatment for handicapping conditions in children. In this case, it seems unlikely that refinements in the data on program costs or effectiveness would have changed policy recommendations.

Similarly, the recommendation for family planning programs was clear. The imprecision in cost estimates, and in estimates of program effectiveness had no important effect on the strength of the policy recommendation to expand family planning programs.

The best effectiveness data were probably the data on the effectiveness of intensive care units in preserving lives of low birth weight, newborn infants and other high-risk infants. It is fairly clear that cost and effectiveness data on this program helped strengthen the argument against the program. Although the program appeared to be the most cost-effective in preventing infant deaths, the high cost of the program made it unattractive.

Again, the costs of a comprehensive dental care program made it unattractive in comparison with an equal-cost program of screening for, and follow-up treatment of, chronic handicapping conditions.

E. PROGRAM ANALYSIS GROUP RECOMMENDATIONS

The work of the program analysis group on child health care became the basis of recommendations by the Secretary of Health, Education, and Welfare to the President for maternal and child health care programs for fiscal year 1968. In the spirit of the planning-programming-budgeting system, the Secretary's recommendations included both proposals for new legislation and proposals for budget levels for existing maternal and child health programs. Given funding limitations in fiscal year 1968, and because the programs could not be fully operational in fiscal year 1968 as a result of the need to enact

legislation, fund the program, and carry out the steps required to reach full operational status, the Secretary indicated that he would actually need less than $100 million for the first year of the new and expanded maternal and child health care programs.

Secretary Gardner's major recommendations for new programs were the following:

1. Comprehensive maternal and child health care demonstration projects, to test the feasibility, costs, and effectiveness of training and using new types of health personnel (obstetric assistants, pediatric assistants, and dental auxiliaries) to assist doctors and dentists in providing mothers and children with comprehensive health care. The Secretary proposed 10 such experimental projects to be established in fiscal year 1968. He noted that, if the demonstrations were successful, larger numbers of these centers would be proposed for subsequent years, usually as part of neighborhood health centers serving all of the population in a given area and serving middle-income as well as low-income residents.

2. A program of early case-finding and treatment of chronic conditions, to detect and assure treatment of potentially handicapping conditions in children. The Secretary noted that this program would remove one of the defects in present school health programs which tend to find the same defects in some children year after year, there being no existing mechanism for assuring follow-up and treatment. To cover the cost and to provide a mechanism for furnishing services for this program, the Secretary stated that the program would make use of crippled children's services when available. He recommended that the title XIX medicaid program be extended to include payment for diagnostic examinations several times in the first 10 years of a child's life and treatment of medical problems discovered in these examinations, including provision of necessary appliances such as eyeglasses.

3. Family planning programs aimed at high-risk groups such as teenage girls and low-income women with three or more children. The Secretary recommended a program of project grants to public or private nonprofit agencies to develop or improve family planning programs as an integral part of maternal and other health services. He noted that these programs should lighten the financial burden imposed by large numbers of unwanted children in many families and stated that such programs have the potential to reduce infant mortality in the high risk groups dramatically.

4. Program of fluoridation of community water supplies. The Secretary recommended a program of grants for equipment and operating costs to communities which undertake to fluoridate their community water supplies. The Secretary stated that his program was designed to reduce dental disease by more than 50% at minimum expense, thus reducing either unmet dental needs or costs of dental treatment in subsequent years. He noted that fluoridation is effective up to age 12 while the teeth are being formed. (Although this program of fluoridation of community water supplies was included in the Secretary's recommendations to the President for maternal and child health care programs, this program was not included in the proposed Child Health Act of 1967.)

5. Treatment of Selective Service medical rejectees. The Secretary also recommended expenditure of funds for treatment of Selective Service medical rejectees, to fill in the gaps in present programs and to make use of Selective Service medical examinations to gain extremely useful information on the health of our youth. (This fifth program, although included in the Secretary's recommendations to the President, did not appear in the proposed Child Health Act of 1967 nor in the fiscal year 1968 budget request.)

6. Program evaluation. In his program memorandum to the President, "Recommended Maternal and Child Health Care Programs for Fiscal Year 1968," the Secretary stated, "Lack of data on the effects of current programs has been a handicap to evaluation and planning. I am, therefore, directing that the Welfare Administration and the Public Health Service develop and implement improved reporting systems and intensified health interview and health examination surveys to improve our knowledge of the need for and the effectiveness of maternal and child health care programs. I am directing a continuing evaluation of the effectiveness of all maternal and child health care programs in meeting specific objectives such as the reduction of infant mortality and chronic handicapping conditions."

On reflection, it appears that for the major policy questions examined by the Program Analysis Group on Child Health Care, the available data, though poor, were strong enough to support clear policy recommendations. Better data on program costs and effectiveness would have allowed more confident predictions of the cost-effectiveness of the programs recommended, but each of the policy recommendations was fairly clear once the available evidence was organized. The problems arose later, however, when it came time to try to implement the program analysis group recommendations.

III. Decisions on Child Health Programs

This section discusses both the immediate impact of the child health study and decisions within HEW and Congress over the subsequent two and one-half years. In this section, as in the preceding discussion of the operations of the program analysis group, we are particularly interested in problems resulting from lack of knowledge of the effectiveness of past programs.

A. THE SECRETARY'S DECISION

The program analysis group report was prepared in the form of a draft program memorandum from the Secretary of Health, Education, and Welfare to the President. The program analysis group recommendations, a number of which represented departures from Children's Bureau program planning, were not received with universal enthusiasm. Nevertheless, after a briefing on the programs analysis group conclusions and recommendations by the program analysis group chairman and its staff director, the Secretary accepted

the recommendations of the program analysis group. Because of the time required for enacting legislation, funding the programs, and carrying out the steps required for new program to reach full operational status, the Secretary stated in his memorandum to the President that the fiscal year 1968 budgetary requirements would be less than the $100 million increase recommended by the program analysis group.

B. DRAFTING THE 1967 LEGISLATIVE PROPOSALS AND THE FISCAL YEAR 1968 BUDGET PROPOSALS

After the Secretary's decision, the Secretary's staff and the Children's Bureau staff turned to preparation of a proposed "Child Health Act of 1967" and a proposed maternal and child health budget for fiscal year 1968. Early in this process, two of the recommendations of the program analysis group were dropped: the proposed fluoridation program (which, though highly cost-effective, would have been politically unacceptable) and the proposed program to provide needed medical treatment to Selective Service medical rejectees (which depended on another program whose continued existence was itself under examination at that time).

The proposed new child health programs were to be included in the amendments to Title V and Title XIX of the Social Security Act. On the basis of the Secretary's decision, the authors of the program analysis group report were able to get the major program analysis group recommendations included in the Department's legislative and budget proposals:

- Provision was made in the proposed Child Health Act of 1967 to expand the research authority of the Children's Bureau to include demonstration programs to test the training and use of physician assistants and other types of health aides.
- The Title XIX (Medicaid) program was to be expanded to include *provision of periodic screening* ("*early case-finding*") *and follow-up treatment of handicapping conditions in children* among the required services in State Title XIX programs.
- A decision was made, perhaps not a wise one from the point of view of the program analysis group report authors, to organize delivery of services under the early case-finding and treatment program through the State Crippled Children's programs.
- Children's Bureau legislation was amended in the proposed Child Health Act to make clear the authority of the Children's Bureau to support family planning projects.
- The proposed Child Health Act included authority for the Secretary to reserve up to one-half of one percent of maternal and child health appropriations for program evaluation.

(On the other hand, a demonstration dental health care program, *not* recommended by the program analysis group, was added to the proposed Child Health Act.)

- The administration included $38 million in the proposed 1968 budget for maternal and child health programs under the new

legislation. Tentative decisions in HEW in January 1967 were to distribute the $38 million as follows: $15 million for early case finding and treatment of handicapping conditions, $10 million for demonstration projects to test the use of physician assistants, $3 million increase for training programs, $5 million increase for maternity and infant care projects, and $5 million for dental health care projects.

C. CONGRESS' LEGISLATIVE AND BUDGET DECISIONS IN 1967

The House Ways and Means Committee, which considers amendments to the Social Security Act, was responsible for passing on the proposed child health legislation as it considered other proposed changes in social security and public assistance programs. In this arena, the HEW Office of the Secretary, armed with much better than usual backup data on the program requests, did rather well. Congress reorganized and simplified the administration's proposed child health legislation but included the major elements recommended by the President in his proposed Child Health Act:

1. Broadened authority for the Children's Bureau to mount demonstration programs, to test the training and use of physician assistants and other health aides in furnishing maternal and child health care.
2. Changes in both the crippled children's program and the title XIX medicaid program, to include in both the requirements that the States furnish programs of periodic screening of children for handicapping conditions and follow-up treatment of chronic conditions found in these screenings.
3. Authority for the Children's Bureau to support family planning projects. Here Congress went beyond the administration's request and earmarked at least six percent of all maternal and child health appropriations for family planning.
4. Authority for the Secretary of Health, Education, and Welfare to reserve up to one-half of one percent of maternal and child health appropriations for evaluation of the impact of these programs.

The new legislation was not supported by increased fiscal year 1968 funding, however. In a very tight budget year, Congress took a general position of not funding new programs, in particular, not adding funds for the new maternal and child health programs. The Children's Bureau maternal and child health budget increased by $10 million over fiscal year 1968, from $168 million to $178 million, but most of the additional funds were required for "uncontrollable" increases in existing programs (see table I).

TABLE I.—COMPARISON OF PROPOSED AND ACTUAL CHILDREN'S BUREAU MATERNAL AND CHILD HEALTH BUDGETS, FISCAL YEARS 1966–70

[In millions of dollars]

	1966 actual	1967 actual	1968 Proposed	1968 Actual	1969 Proposed	1969 Estimated	1970 proposed
Total [1]	132	168	[2]219	178	237	209	231
Maternal and child health services	45	50	50	50	50	50	50
Crippled children's services	44	50	[2] 65	50	65	57	58
Special project grants for maternity and infant care	24	28	[2] 35	30	55	48	63
Special project grants for health of school and preschool children	15	32	40	37	42	39	41
Special project grants for dental heath of children			[2] 5		1		
Training		4	[2] 10	7	10	9	11
Research	4	5	[2] 14	6	13	6	8
Evaluation					1		

[1] Columns may not add to totals, due to rounding.

[2] The 1968 proposed budget included a proposed supplemental appropriation of $38,000,000. The figures shown here include the planned distribution of the $38,000,000 supplemental, as of January 1967:

	In millions of dollars
Crippled children's services (early case-finding and treatment of handicapping conditions)	15
Maternity and infant care	5
Dental health	5
Training	3
Research (pilot projects to test use of physician assistants)	10

D. DECISIONS IN SUBSEQUENT YEARS

1. *The fiscal year 1969 budget*

Again in fiscal year 1969 (another very tight budget year), the Department of Health, Education, and Welfare requested $15 million from Congress for the early case finding and treatment program under the crippled children's program and $7 million for demonstration programs to test the use of physician assistants in furnishing maternal and child health care in new and more efficient ways. To support expanded family planning programs and expanded maternity and infant care, HEW proposed a $25 million increase in project grants for maternity and infant care. The administration also proposed that $1 million be appropriated for program evaluation, as authorized under the 1967 amendments (see table I).

In another tight budget year, Congress increased the 1969 maternal and child health budget by $32 million over fiscal year 1968, from $178 million to $209 million, approving half the new funds requested for early case finding and treatment of handicapping conditions and about three-fourths of the funds requested for maternity and infant care (including family planning) but *not* approving funds to test more efficient ways of delivery of maternal and child health services or funds for program evaluation.

2. *The fiscal year 1970 budget*

By fiscal year 1970, except for continued emphasis on family planning programs, the new thrusts in the Child Health Act of 1967 had largely disappeared from the administration's program requests (see table I). The President's fiscal year 1970 budget proposed essentially no increase for the crippled children's program and (therefore) for

early case finding and treatment of handicapping conditions, nor for the research and demonstration program to test the use of physician assistants in delivery of maternal and child health care. On the contrary, the major thrust of the maternal and child health budget request had, by 1970, entirely shifted to family planning and maternity and infant care projects which, in conjunction with family planning, would attempt to further decrease infant mortality rates in low-income areas within the United States.

IV. Conclusions

This paper has documented some of the problems encountered in translating the conclusions of analytical studies into program policy changes when the studies are not based on knowledge of the impacts of past programs. The child health study is especially notable in this respect, since its findings influenced both administration and congressional policy decisions and were translated into legislation in 1967 but (except for the recommended family planning programs which had powerful political support) still have not led to significant program changes.

In 1966, the Federal Government had no system for managing evaluation of the effectiveness of its programs. In particular, no agreed-on objectives had been established for federally-assisted maternal and child health programs; management information systems for Federal maternal and child health programs did not make available information on the effects of services provided under these programs; there was no system for planning what research and evaluation studies were required to measure the overall effectiveness of Federal maternal and child health programs or the relative effectiveness of different projects within these programs; and there was no system for review and use of program evaluation and project evaluation studies in reaching policy decisions on maternal and child health programs.

In this environment, definitive analysis of child health program proposals was impossible and effective program planning, difficult. Without knowledge of the effectiveness of past and present programs, the analyst could not confidently recommend reallocations within the existing maternal and child health budget, nor could the analyst provide definitive conclusions on the relative costs and effectiveness of alternative proposals for new programs. In the absence of knowledge about program effectiveness, what could have been compelling arguments for program changes were reduced to semi-persuasive arguments, more convincing to the analyst than to HEW decision-makers committed to their own programs and their own legislative proposals.

As a result, decisions taken on needed new thrusts in child health care became unstuck and were reversed over the next two years as policy makers came and went in the Office of the Secretary of Health, Education, and Welfare.

In the absence of knowledge of the impact of past programs, no one can say for certain who was right or wrong about policy decisions made or unmade. Fashions and legislative proposals will come and go. In the absence of knowledge of program effectiveness, development of more effective expenditure policy is possible but not likely. Without evaluation of the effectiveness of past programs and projects, it

seems more likely that history will repeat itself: the traditional forces within and outside the Federal Government will continue to shape and reshape Federal programs without benefit of the guidance that well-founded analyses could give.

The whole structure of the Federal Government's planning-programming-budgeting system suffers from a crucial weakness: the lack of a system of program evaluation which would provide measures of effectiveness related to program objectives. In the absence of knowledge of the effects of past programs, analytical studies will often fail to influence Federal decision-making, or may succeed in having such influence when they shouldn't.

V. Recommendations for Improving Evaluation of Social Programs

With so many unmet needs in health, education, and other areas of national concern, and especially in the face of limited resources, it is imperative that we spend our resources wisely and improve the effectiveness of our Federal, State, and local programs. In the past three years, the Federal Government has taken a number of steps to improve its capacity to evaluate the effectiveness of ongoing programs. The executive branch of the government has allocated limited staffs and limited financial resources to program and project evaluation. Congress has authorized or earmarked funds for evaluation in several pieces of legislation, including the Child Health Act of 1967, the Elementary and Secondary Education Amendments of 1967, and the Partnership for Health Amendments of 1967. Some agencies are developing systems for planning the overall direction and use of program evaluation studies to assist future decision-making. The picture has improved but, in the absence of substantial commitments of staff and money, has not improved enough.

In any case, increasing the funds available for evaluation is not the entire answer.

We also need in our Federal agencies and in Congress the burning desire and determination Adlai Stevenson spoke of in 1952, the determination to get the best services for the public at least possible cost—the determination to be both effective and efficient. At a time when the nation may be pausing in its enactment of one piece of social legislation after another and at a time of change of Administrations, it may just be possible to develop the will to attempt to learn which Federal programs are effective and which are the most efficient ways to achieve national objectives.

The child health study and a number of the other early PPB studies were valuable for the rough estimates they provided on the value of alternative program proposals. Repeated PPB studies without input of program effectiveness data are likely to become wheelspinning exercises, however.

Too often, moreover, evaluation funds have been spent on crash efforts, on occasions to serve as after-the-fact justification for program decisions. As a result, evaluation money often was spent inefficiently, and the conclusions of evaluation studies often turned out to be of ques-

tionable validity. Such experiences contribute to erosion of the credibility and desirability of program evaluation.

This is not the place for development and justification of a detailed set of recommendations on how to improve Federal program evaluation. The following suggestions on Federal program evaluation are offered to help promote discussion among those interested in improving the effectiveness of Federal programs:

1. *Resources.*—In all of our major social programs, we are recognizing the need for evaluation of the effects of the programs so that the programs can be improved and program objectives reached. Congress is now asking for evaluation of program effectiveness and is authorizing use of program funds for evaluation and more effective management of Federal programs.

In 1967, the Administration recommended and Congress agreed that up to one-half of 1 percent of the funds appropriated for maternal and child health programs should be reserved to the Secretary of Health, Education, and Welfare for evaluation of the effects of maternal and child health programs, and that up to 1 percent of the funds appropriated for "Partnership for Health" programs should be reserved to the Secretary of Health, Education, and Welfare for evaluation of the effects of Partnership for Health programs. Similar provisions were included in a number of pieces of legislation enacted in 1968. The need for program evaluation is at least as great in education, welfare, and training programs, as well as in the extremely complex community action and Model Cities programs.

Following the precedent of the Partnership for Health and the Child Health Amendments of 1967, the Administration should recommend that Congress allow agency heads to reserve for program evaluation a small percentage of the appropriations under each piece of major social legislation. Such funds should then be allocated to evaluation at policy planning and at operating levels, as is suggested in No. 4 and No. 5 below.

2. *Agreement on program objectives.*—Federal agency staffs should, on a continuing basis, address the question of defining program objectives in terms that will allow later assessment of the programs and their local projects as more or less successful in achieving agreed-on outputs. Further, some serious efforts should be made to rank program objectives so that evaluators will work with an agreed-on frame of reference.

3. *Routine reporting on program effectiveness.*—Serious efforts must be made to develop reporting systems that will generate data on program effectiveness, perhaps through follow-up surveys on samples of program recipients and members of appropriate control groups or comparison groups.

4. *Resources for evaluation studies.*—One of the major obstacles to program evaluation is the simple lack of staff members whose job is program evaluation. The Administration and Congress can do a great deal to improve the management of Federal programs simply by designation of responsibility and allocation of staff and financial resources for program evaluation at departmental, agency, bureau, and operating program levels. Consideration should also be given to the

desirability of program or multi-program evaluation studies by the Bureau of the Budget, by the Congress and the General Accounting Office, and by organizations not funded by the Federal Government.

5. *Planning for evaluation.*—At least for the more important social programs, a process should be established for short-term and multi-year planning for evaluation studies and information systems required to answer basic questions about the program's effectiveness and about the relative effectiveness of different projects within the program. Evaluation plans should take into account needs for both long-term and short-term evaluation efforts, requirements for continuity of funding of evaluation efforts, personnel as well as funding requirements, division of evaluation responsibilities among departmental, agency, bureau, and operating levels, and development of suitable methodologies and measures of effectiveness. This process should include provision for review of the designs for proposed evaluation studies (at departmental, agency, bureau, and operating levels), to ensure that proposed studies address the important questions on program effectiveness and that their research designs are technically sound and appropriate to the programs to be studied.

6. *Review and dissemination of evaluation studies.*—Consideration should be given to development of a system for review of all relevant reports and evaluation studies when significant decisions are being made on project grants, on allocation of funds within a Federal agency budget request, and on allocation of funds among Federal agency budgets.

THE ROLE OF SOCIAL INDICATORS AND SOCIAL REPORTING IN PUBLIC EXPENDITURE DECISIONS [1]

BY ISABEL V. SAWHILL

Isabel V. Sawhill, formerly a member of the Social Indicators Staff at the Department of Health, Education, and Welfare, is now a consultant at the Bureau of the Budget.

Sound public expenditure policy requires that the outputs produced by expenditures be defined and evaluated. Because the outputs of many public expenditures consist of improvements in some aspect of the quality of life such as health, education, or the physical environment, public decisionmakers and analysts require knowledge of the level of, and changes in, these social characteristics. This is necessary for both setting objectives and evaluating the results of expenditures.

The recent effort by the Department of Health, Education, and Welfare to develop a set of social indicators is described in this paper. Both the methodology of social reporting and the progress in developing social indicators are discussed. Mrs. Sawhill indicates that the establishment of a Council of Social Advisers may be helpful in establishing the link between social indicators and more effective public expenditure decisions.

Introduction

Social indicators are quantitative measures of social conditions designed to guide choices at several levels of decision-making. In one case, social indicators might be involved in a decision to redirect resources from the War in Vietnam in order to escalate the War against Poverty. In another case, social indicators might be used to allocate expenditures among our various broadly-defined social programs such as health, education, and the control of pollution. In still another case, social indicators might guide the choice of a specific program to improve the health of the nation.

In general, the higher the level of decision-making (the broader the choices involved) the more difficult it is to develop satisfactory indicators of performance. This is particularly true in the case of social goals and programs. In fact, at the higher levels of decision-making, it is not now possible, and probably never will be, to construct quantitative measures of progress. Only the judgment of the voters and their political representatives can determine whether an extra dollar spent on schools and parks increases the well-being of the nation more than an extra dollar spent on guns or butter. However, when the choice is narrowed somewhat, social indicators can play an important role. Indeed, in an indirect way, they may serve to reorder our fundamental national priorities by drawing greater attention to social problems.

[1] Many of the ideas in this paper are the result of discussions I have had with other members of the Social Indicators staff at the Department of Health, Education, and Welfare. I am particularly indebted to Mancur Olson, Jr., former Deputy Assistant Secretary (Social Indicators), who has done extensive work in this area.

Recent History of Social Indicators

The recent interest in quantifying social phenomena, or in developing social indicators, is in part an outgrowth of the space program. A few years ago NASA commissioned the American Academy of Arts and Sciences to study the social implications of space exploration. Scientists were aware that important technological advances often have indirect effects on social institutions. The development of the automobile, for example, profoundly altered the American way of life and there is substantial reason to suppose that the program to send men and machines into outer space may also change the social environment. The result of this concern was the publication of a volume entitled, *Social Indicators*,[2] to which a number of scholars contributed, and which, to date, is one of the most comprehensive treatments of this subject. It was recognized at the outset of this work that it is impossible to separate the effects of the space program (or any other specific program) from the effects of the many other influences which are operating simultaneously to produce social change. The primary concern then was to develop quantitative measures of social conditions. Once this was accomplished, further work could be done to show how changes in these conditions relate to changes in specific programs.

Since these early beginnings growing out of the space effort, the interest in social measurement has increased enormously. In 1966, the President directed the Secretary of Health, Education, and Welfare to develop a series of social statistics and indicators in order that we might "better measure the distance we have come and plan for the way ahead." In response to this directive a "Panel on Social Indicators," consisting of a group of leading social scientists, was assembled to gather and develop what information they could. With the help of a staff within the Department, headed by Mancur Olson, this resulted in the publication of a document entitled, *Toward A Social Report*, possibly the first step in the development of regular social reporting.[3] In the meantime, Senator Walter Mondale and a group of other senators have introduced a "Full Opportunity and Social Accounting Act" in both the ninetieth and ninety-first Congresses. The bill calls for the establishment of a Council of Social Advisors who would be responsible for reporting to the President and the Nation about social conditions in the same way that the Council of Economic Advisors reports on economic progress.

Other governments have also been interested in developing new approaches in the field of social measurement. In Great Britain, for example, the Central Statistical Office plans to publish a new periodical called *Social Trends* which will bring together a variety of time series on social conditions. The academic community has also responded to the need for social indicators [4] and the recent publication of the HEW report will probably stimulate even wider discussion of some of the questions which these preliminary efforts have raised.

---

[2] Raymond A. Bauer, ed. *Social Indicators*. Cambridge Mass.: M.I.T. Press, 1966.

[3] Some of the material in this paper is taken from the Appendix to this Report.

[4] John Oliver Wilson of Yale University, for example, has written an interesting paper (unpublished) on "Inequality of Racial Opportunity—An Excursion into the New Frontiers of Socioeconomic Indicators." Other literature in this field includes "Social Goals and Indicators for American Society," *Annals of the American Academy of Political and Social Science*, vol. I and II (May and September, 1967) and Eleanor Sheldon and Wilbert Moore, *Indicators of Social Change*, Russell Sage Foundation, New York, 1968.

Rationale for Social Reporting

Can social indicators really improve our ability to make wise policy choices? The hope is that they can. In part, this hope stems from the contribution which economic indicators have made to better economic policy. Without our measures of GNP, unemployment, and price levels, economists could not monitor the performance of the economy and make the required shifts in policy. Presumably similar measures of social performance—such as carefully constructed indices of health, learning, crime, or pollution—will contribute to better social policy. Regular publication of such indicators would at least focus attention on these problems and would allow the informed citizen as well as the policymaker to see what was happening in several broad areas of social concern.

It would be a mistake, however, to carry the analogy between economic indicators and social indicators too far. In the first place, there is no unified body of social theory comparable to what exists in the field of economics. Economic theory permits economists to trace the effects of a change in tax rates or a shift in monetary policy on such variables as income, prices, and employment with some accuracy. It is much more difficult to predict the effect of, say, a change in police practices on crime rates. Secondly, there is little agreement about goals where social policy is concerned whereas the goals of economic policy are clearly set forth in the Employment Act of 1946. However, the fact that there are gaps in social theory and little consensus about social goals does not imply that better measurement of social conditions can serve no purpose. In fact, one of the benefits of such measurement would be to assist both the development of better theory and the establishment of national goals or priorities in the social field.

Even more important perhaps is the fact that the American people are bombarded with reports and messages urging them to take steps to get their clothes cleaner, their kitchen floors shinier, and their cigarettes longer but they are seldom reminded of the benefits of alternative uses of their time, money, and interest, such as reducing the pollution of the air they breathe, lengthening their lives, or making their cities more habitable.

Moreover, for better or worse, we are all somewhat impressed by numbers. The advertising industry has been using statistics to dazzle their audiences for quite some time; they even have their own brand of PPB. We are told that if we will use toothpaste A instead of toothpaste B we will have 14 percent fewer cavities. Hopefully, we would be equally impressed if we were told that income maintenance scheme A would lead to 14 percent less poverty than income maintenance scheme B. (The toothpaste industry does not always tell us whether brand A costs more than brand B; those in charge of program evaluation would presumably supply this information.)

What Kind of Social Reporting?

Although there is probably not too much debate about the ultimate usefulness of social reporting, there is more uncertainty about just what a social report should include.

First of all, a decision has to be made on what areas of our social life to measure. Here, there are no established criteria for choosing one area over another. There is perhaps some agreement that we should limit our concern to those social problems which are legitimate targets for public policy. This means, for example, that it might be inappropriate to report on the state of religion, sexual mores, or childrearing practices since these are areas where civil liberties are important and where, as a result, governmental action is not generally accepted. (Even here, however, Government programs might have unintended, indirect consequences which we would want to know about.) Beyond this, most criteria for the collection of social statistics are rather vague, involving, for example, the requirement that they be of general interest or pertaining to goals widely valued in the society. There may also be some desire to be reasonably comprehensive but it is impossible to be completely comprehensive by everyone's standards. Thus, if a group of ten people were each asked separately to list those social problems or conditions of greatest importance to the Nation, one would undoubtedly get ten different lists (albeit, with some overlap). In the recent HEW report, the topics covered included health and illness; social mobility; the condition of the physical environment; income and poverty; public order and safety; and learning, science, and art. There was also a chapter on participation and alienation, but a lack of data necessitated that the chapter simply raise some important questions. There was no specific justification for this set of topics other than it was "an attempt, on the part of social scientists, to look at several very important areas and digest what is known about progress toward generally-accepted goals." [5]

There is, at present, no easy solution to the problem of how to choose which social conditions to measure. Many people may not even feel that this is a problem and may take the view that we should simply produce as many indicators as we can without worrying about their relative value or usefulness. But this could lead to volumes of reports and statistics which no no could profitably digest and to a confusing picture of the social state of the Nation. Other people may feel that the contents of a social report should be dictated by the political mood of the country, the interests of Congressmen, or the judgment of public administrators. This approach, while probably the most realistic one, may not provide a very stable basis for making cumulative progress toward the development of better social measurement and better social theory. Thus, unless there is a developing consensus about social goals (and they may change from year to year) and unless these goals are embodied in law, this uncertainty about the direction of social reporting will probably persist.

Those responsible for producing a social report face another difficulty—they must find some way to handle the overlap which exists between most social problems. Any attempt to systematically assess the social state of the nation should give explicit recognition to the intricate interplay of social variables. Education, for example, contributes to well-being not only because it improves the intellectual quality of our lives but also because it raises both individual and na-

[5] *Toward A Social Report,* Department of Health, Education, and Welfare, January 1969, p. 6.

tional income, tends to reduce prejudice and discrimination, and helps to preserve our cultural heritage and democratic institutions. Similarly, a high income not only provides individuals with adequate shelter, food, clothing, and the enjoyment of numerous other goods and services for sale in the market place but also helps indirectly in the attainment of good health, and the acquisition of education, and brings greater insulation from the dangers of crime and the unpleasantness of the physical environment. In addition, the progress which can be made in one area may depend on progress made in another area. For example, without a highly-educated population, various kinds of preventive health programs might prove quite ineffective. Without control of pollution, and other environmental problems, higher incomes might simply be dissipated into expenditures for goods (clean air and water, land, a pretty view) which were once virtually costless.

It is the recognition of one particular kind of overlap—that between social problems on the one hand and economic problems on the other—that has led some people to resist the proposal to establish a separate Council of Social Advisers.[6] They argue that because of this interdependence it would be unwise to have two separate sets of presidential advisors—one dealing with economic problems and the other with social problems. Instead, there might be a combined Council of Economic and Social Advisors,[7] and this combined Council might issue a joint Social and Economic Report. Economists, of course, have been making extended forays into the social field in recent years—treating such subjects as health, education, and job training, for example, as "investments in human capital." A perusal of any recent *Economic Report* indicates that the Council has been analyzing a much wider range of topics than the level of production, employment, and prices. Thus, it would be consistent with present trends to broaden the scope and change the title of the present Economic Report to include an assessment of social conditions.

On the other hand, there are arguments for creating a separate Council of Social Advisers which would produce its own report. First of all, some have argued that social reporting need not be as frequent as economic reporting since social conditions presumably do not change as rapidly as economic conditions.[8] Secondly, an autonomous Council of Social Advisers might more readily capture the attention of the Nation (including social scientists from the non-economic disciplines) than an enlarged CEA.

The Development of Social Indicators

Once we have decided what aspects of life to measure, and how to handle the complexity of social life, the next step is to provide social indicators or quantitative measures of social progress. A social indicator has been defined to be "a statistic of direct normative interest which facilitates concise, comprehensive, and balanced judgments about the condition of major aspects of a society. It is in all cases

---

[6] See, for example, the statement made by William Gorham before the Senate Subcommittee on Government Research, Committee on Government Operations, July 26, 1967.

[7] This suggestion was made by Secretary Cohen at a press conference held January 20, 1969.

[8] See, for example, the editorial in *The New York Times*, January 20, 1969, and the dissenting view expressed in a letter from Senator Mondale, February 24, 1969.

a direct measure of welfare and is subject to the interpretation that, if it changes in the 'right' direction, while other things remain equal, things have gotten better, or people are 'better off.'" [9] Thus, statistics on the number of teachers employed in the school system would not be a good social indicator because teachers are only an input to the educational system and there could be an increase in the number of teachers with no corresponding increase in the quality or quantity of what children learn. In other words, there is no one to one relationship between the amount of money spent on social programs and the benefits to society. Sometimes they even change in opposite directions as the result of changes in the efficiency with which resources are used.

The requirement that a social indicator be a measure of welfare is a sound one since it separates those statistics which tell us something about social progress in a given area from those which do not. It must be remembered, however, that a statistic which is a direct measure of welfare in one context may have no normative content at all in another context. To see this, assume that the grandest and most universal social indicator is some measure of happiness, satisfaction, or gross social product. Such a statistical construct does not, and probably never will, exist. But theoretically, at this level, all other statistics would merely be inputs or sources of welfare. From a more limited vantage point, however, such things as health or education might be of a direct normative interest and statistics which summarized the state of health or learning would then be social indicators. But, if this is the case, it is only because we believe that these statistics have normative content and this belief, in turn, is based on the assumption that more health and knowledge imply greater happiness or some other vague ideal. By the same kind of reasoning, at a somewhat lower level of analysis, even the number of teachers employed by the school system could be a social indicator or measure of the desirable output of some program if it had previously been established that more teachers mean more learning. In other words, what is a measure of input or a source of welfare at one level can easily become a measure of output or a social indicator at another level. It is, of course, important to start with broadly defined social goals and indicators and to work down to lower levels of analysis in a systematic way to insure that we are optimizing the right variables at the lower levels. It should be added that the forging of causal links between one level of analysis and another is likely to improve our knowledge of important social relationships.

Very little work has been done on the actual construction of new social indicators in areas where they are now lacking. Yet the possibilities are almost unlimited. From a technical standpoint, social phenomena are not as difficult to quantify as has often been assumed. If we can define social goals fairly carefully, we can usually measure social conditions. The principal barrier to quantification, in the long run at least, is not a lack of meaningful data but a failure to define what is meaningful, to have a clear idea of what we want in various areas of social concern, or to give operational content to our ideals.

[9] *Toward A Social Report*, U.S. Department of Health, Education, and Welfare, January 1969, p. 188.

For example, there is nothing difficult about measuring poverty once there is some agreement about what poverty is. Can it be defined in terms of income alone or is there a culture of poverty which is partially independent of an individual's financial condition? If income can be used as an indicator of poverty, where should we draw the line between poor and nonpoor? And should this line be changed from time to time to reflect not only changes in the cost of living but also changes in the general level of affluence within the Nation?

These are questions which have already been asked and for which some partial answers exist. One should not insist that these answers be incontrovertible. Our definitions of unemployment, for example, are not beyond debate yet they have been used for a long time now to count the unemployed, and any conceptual error which may be hidden in such definitions is far outweighed by the convenience of having these statistics. Although great care must be exercised when the definitions and guidelines for new social statistics are formulated, we should not let the mumblings and grumblings of the experts about relatively minor problems of measurement deter us from benefitting from the potential usefulness of these statistics.

Table 1 presents a list of some of the statistics which were used to measure social conditions in *Toward A Social Report.* They are not in every case the best social indicators that could be used. New data and/or different definitions of social goals could lead to the use of a different set of indicators. However, they will serve to illustrate some of the possibilities and the problems involved in social measurement.

TABLE 1

Some Social Indicators Used in *Toward A Social Report* to Measure Major Social Conditions

Aspect of Life To Be Measured	*Some Social Indicators* [1]
Social Mobility	Correlation coefficient between father's occupational status and son's occupational status
Health	Expectancy of healthy life (free of bed disability and institutionalization)
The Physical Environment	Ratio of a city's actual level of pollution to an acceptable standard
	Proportion of housing that is substandard or overcrowded
Income and Poverty	Personal income per capita
	Number in poverty
Public Order and Security	FBI Uniform Crime Reports Index of Crimes
	Value of property involved in theft per $1,000 of appropriable property
Learning, Science, and Art	Performance on selected achievement tests
	Technological balance of payments
	Attendance at theaters, operas, ballet

[1] These are not necessarily the best indicators but rather what was currently available.

Basically, there seem to be three methods by which to develop a social indicator. One method is to construct a weighted index number. Another way is to use a surrogate or proxy variable to represent a social condition for which direct measures are lacking. Finally, "critical-valued" social indicators can be used. Each of these three types of indicators will be discussed in turn.

Index Numbers as Social Indicators—An index number is a weighted average which summarizes a large number of observed changes into one number, giving the greatest weight to those changes which are deemed to be most important. (There are unweighted index numbers as well but they are not very common.) One of the best known index numbers is the Consumer Price Index. Basically, it is an average of price changes which gives the greatest weight to the prices of those items which loom large in the typical consumer's budget.

Similar indexes can be constructed to measure various social conditions. A crime index, for example, might serve as an indicator of the general level of crime in the nation, but one would have to find some way to weight the more serious crimes like murder more heavily than the less serious crimes like purse snatching. There are a number of ways to obtain such weights. They might be based on the average length of prison sentences for various crimes or on the results of a survey in which people ranked various crimes according to their relative seriousness. Or, the monetary losses associated with different crimes might be used, although this would be less satisfactory since it is fallacious to measure the value of human life or even personal injury in financial terms alone.[10]

Proxy Variables as Social Indicators—In a sense, all social and economic indicators are proxy variables in that they are not direct measures of welfare. We are, in this section dealing with measures that are even farther removed from what we would ideally like to measure.

An example of this use of one variable to represent another is the indicator of social mobility cited in Table 1. To construct this indicator it is assumed that a man's socioeconomic position in our society is closely associated with the kind of occupation he is in. Accepting this assumption makes it possible to measure the extent of social mobility or equality of opportunity since there are data on occupational status but not on socioeconomic status.

Critical-Valued Social Indicators—A critical value is a number which defines the acceptable level which a social statistic can reach before there is cause for concern. Social indicators are created by simply looking at the number of cases (people, houses, cities, etc.) which fall on one side or the other of the critical value. The best example of such a critical-valued social indicator, and one which has already been discussed, is the incidence of poverty defined as all those families with less than a specified income (the particular income level depending on such things as residence and size of family). Other examples include measures of pollution, where a ratio of the actual level of pollution to some acceptable level is computed, and of the quality of housing, where the number of units which fail to meet certain standards with respect to structural soundness, plumbing facilities, and persons per room are counted.[11]

The above discussion with its emphasis on index numbers, proxy variables, and critical values may seem like an attempt to complicate what are in fact rather ordinary statistics. Its purpose is to facilitate

[10] Such calculations have been made, however, with the value of human life being set equal to the projected earnings lost because of injury or death. This makes the value of the life of a child or of a non-working wife very small indeed, and is consequently one of the drawbacks of this procedure.

[11] Sometimes a "critical value" is not a number at all but a verbal definition or criterion which divides a population into two or more groups. Unemployment statistics are an example.

and encourage the construction of other social indicators by outlining the rather simple structure which lies behind most of our present indicators. By using these same principles, any number of other indicators could be produced.

A few other comments are in order. First, most social indicators will be more useful if they are disaggregated by race, sex, age, residence, and other pertinent characteristics since social conditions may be very different for different groups. Thus, the expectancy of healthy life is greater for females than for males, and for whites than for nonwhites. This kind of disaggregation also makes possible some rather simple kinds of calculations which are useful in analyzing social problems. For one thing it is possible to show the effect of demographic changes on social conditions. One can indicate, for example, how changes in the age composition of the population have affected crime rates. Part of the explanation for our increasing crime rate is that most crimes are committed by young people and people in this age group are a larger proportion of our population than they used to be. By working with disaggregated statistics, one can quantify this fact. In similar fashion one can show how a change in the proportion of the population who are farmers has affected the incidence of poverty, how a change in the proportion of housing which is owner-occupied has affected the quality of housing and so forth.

So far, nothing has been said about the problem of nonexistent data. How can social problems be analyzed if the basic data which are needed have never been collected? In the past, most researchers, both in and out of government, have looked upon the existing supply of data as a fixed resource, something which could be twisted and shaped for their own scholarly purposes but not basically changed. In fact, many research projects have been planned around the availability of some interesting statistical series. Better social reporting could help to correct this situation by insuring that the statistics which are collected are those which are needed for the purpose of measuring broad social trends and learning their causes.

One type of data which would be particularly useful for studying a number of social problems is data on values or attitudes. Too often the hard data of economics and demography are used in preference to the softer data produced by public opinion polls or attitudinal surveys. For example, economists often study discrimination against Negroes by looking at differences in the income or earnings of similarly qualified blacks and whites. An alternative approach is to simply ask white people about how they feel about black people (and vice versa) as co-workers, as employees, as consumers, as next-door-neighbors, etc. This approach has its own problems but it would tell us more about the progress toward what is presumably our ultimate goal in this area, i.e., to reduce race prejudice. If there were no prejudice, income differentials, involuntary segregation, and the other undesirable results of discrimination would largely disappear, at least in the long run.

Social Indicators and Public Expenditure Decisions

Social Indicators as Benefit Measures—Social indicators should be designed in such a way that they provide reasonably good estimates of

the desired output or benefits of particular public programs.* As noted earlier, the actual form which an indicator takes depends on the level of decision-making involved or on how broadly programs are defined. In some cases, social indicators are general measures of health, education, and the like. In other cases, they are much more detailed measures of such things as the quality of primary education, or the health personnel available to a specific population. In other words, there is a hierarchy or system of social indicators with suboptimization at each level of decision-making below the most general one. In practice, only a small part of the complete system is analyzed at any one time, but this sort of piecemeal analysis should keep the larger picture in view.

Program analysts like to work, whenever possible, with output measures which are expressed in dollar terms. However, measuring benefits in this way often puts a serious constraint on the kinds of problems which can be analyzed. Social indicators will help to fill this gap. Most of the indicators in Table 1, for example, are expressed in nonmonetary units. It would even be possible to have social indicators with ordinal rather than cardinal properties. It is not always necessary to know *how* effective a program is; it would sometimes be sufficient to know whether or not it was effective and whether it was more or less effective than some other program.

If program analysts don't insist on using dollar values, they will have a more flexible yardstick with which to measure social conditions, but there will be disadvantages as well. In the first place, the benefits of programs in different areas of social concern will not be commensurable. There is no way to make direct comparisons between the benefits to be gained from an increase in life expectancy with those to be gained from an increase in social mobility because they cannot be expressed in the same units. What can be done is to present the alternatives to those responsible for making decisions and let them make the necessary translation of units in their own minds. Thus, information can be provided in the following form: x dollars spent on program A would improve life expectancy by 3 years while the same amount of money spent on program B would increase an index of social mobility by 2 percentage points. The decision between these two alternative uses of funds must then be made on a judgmental basis.

A second, and related, disadvantage of using nonmonetary indicators as benefit measures is that there is no way to aggregate benefits when more than one indicator is needed to judge the effectiveness of a particular public program. Many programs have multiple objectives and their success cannot be judged by a single indicator. The Model Cities program, for example, is designed to improve "the quality of life" in designated areas, and it would probably take a sizeable number of social indicators to measure its intended benefits. But unless these benefits are measured in dollar terms (or some other standard unit), there is no way to obtain an aggregate measure of output or to summarize the overall effectiveness of the program in each city.

Finally, when nonmonetary indicators are used, it is not possible to compute a benefit-cost ratio as it is when both costs and benefits are measured in the same units. Without such a ratio there is no way to

*Further discussion of this issue is found in the papers by Rivlin, Grosse, Brandl, and Levine in vol. 3 of this collection.

identify *quantitatively* those programs for which benefits exceed costs (i.e., those programs which show a benefit-cost ratio greater than one). As a result, the decision-maker will be forced to choose, on the basis of subjective valuation, whether the (nonmonetary) benefits exceed the (monetary) costs.

The above-listed disadvantages should not be overemphasized. In the first place, because of the uncertainties underlying most estimates of even those benefits which are readily expressed in dollar terms, some qualitative judgments are usually necessary anyway. Secondly, a single social indicator can often be used to measure the benefits of several programs designed to achieve a given objective, and in this case, a nonmonetary indicator will reveal the preferred alternative as unequivocally as monetary measures of benefit.

To summarize, whenever the economic returns associated with a public program seem to be the appropriate criteria for judging its success, they should be utilized since various measures of cost and benefit will then be commensurable. But, when economic values are not relevant, and when there is consequently no quantitative information about program benefits, an effort should be made to develop nonmonetary indicators. As we have seen, this will sometimes necessitate that decisions be based on a subjective evaluation of the pertinent facts concerning costs and benefits; but the availability of such information will sharpen judgment and thus lead to better decision-making and more rational policy choices.

Social Production Functions

In cost-benefit analysis perhaps the most difficult task is not measuring costs and benefits themselves but showing how they are related to each other—that is, estimating social production functions. Clearly, the causes of social change, or at least their correlates, must be identified before intelligent remedial action can be planned. It is particularly important to determine whether social conditions are improving in a given area because of the public expenditures made in that area or for other reasons. The learning of school children may increase at the same time that teachers' salaries are going up but the two are not necessarily related. Children may be learning more because they are better-fed or because their parents are better-educated than in the past. Presumably, changes in social conditions which are the result of Government activity can be controlled. Changes in social conditions which are the result of other factors cannot be controlled although they may be offset or counteracted by the appropriate public policies.

Thre are several methods of estimating social production functions, i.e., of relating social conditions (as measured by social indicators) to their major determinants. One method is to use a multivariate statistical model to estimate these relationships from survey data. An example, taken from *Toward A Social Report*, will help to illustrate this method. The difference in the average incomes of whites and nonwhites was used as an indicator of the relative economic position of Negroes. If one of our goals is to improve the relative condition of life for this group of Americans, we need to know why such an income gap exists. If it were the result of discrimination in employment, one remedial policy would be appropriate, but if it were the result of deficiencies in

education or other background factors which impair the qualifications of Negroes, another set of policies would be called for. Using special Census data and a somewhat sophisticated statistical model, it was possible to estimate that perhaps three-fifths of the total income gap is the result of an impairment of qualifications and the remainder the result of current discrimination in the labor market.[12] These estimates cannot be accepted uncritically, but with all their flaws they have important policy implications. They suggest, for example, that discrimination is still very important and that programs designed to assure equal opportunities in hiring, promotion, and pay deserve some emphasis.

Another way to measure the effects of government programs on social conditions is to use the experimental methods of the physical sciences. In essence, this approach consists of choosing human subjects, inducing them to participate in experimental government programs, and observing the results. Since there is always a control group involved, the effects of the government program can be separated from the effects of other influences on the relevant social indicator.[13] There are, of course, some problems associated with this approach; humans are not guinea pigs and they don't react in the same way to being used as experimental objects. However, given the gravity of some of our social problems and the gaps in our ability to solve them, experimentation in the social field must be thoughtfully pursued.

Social Indicators and the Decision-Making Process—Armed with social indicators and some knowledge of the determinants of social change, lawmakers and public administrators would have a greatly enhanced ability to make effective and timely decisions. First, they would be able to monitor social change and identify emerging problems more quickly. As the result of rapid advances in scientific knowledge and technology, social institutions are less stable and more complex than in the past, and there is a greater need to analyze the changes taking place in order to see whether or not they are contributing to our individual and collective well-being. Next, when undesirable trends or conditions were identified, a compensatory social policy could be adopted with government programs designed to allocate resources to the areas of special concern for the purpose of offsetting, retarding, or reversing these trends. As conditions changed, with some areas of our social environment improving and others getting worse, a reallocation of resources would occur to reflect shifting needs and priorities. Finally, social indicators would help to reveal the most effective programs for carrying out agreed-upon policies. Using social indicators as proxies for the benefits associated with various government programs, the most efficient means of achieving a given goal could be identified.

The division of responsibility, within the Executive Branch, required to carry out this decision-making process might be roughly as follows: The job of monitoring social trends would belong to a Council of Social Advisors who would produce an annual (or less frequent) Social

[12] *Toward a Social Report*, p. 25.

[13] This is the approach recommended by G. H. Orcutt and A. G. Orcutt, in "Experiments for Income Maintenance Policies," *American Economic Review*, vol. 58, No. 4 (September 1968). The Office of Economic Opportunity is now sponsoring a controlled experiment on the incentive effects of negative income taxation. (See the paper by Levine in vol. 3 of this collection.) Several similar projects to test the effects of other income maintenance schemes are currently being planned by the Department of Health, Education, and Welfare.

Report. The direction of social policy and the formulation of broad objectives would be the responsibility of the President but would be influenced by the findings and recommendations of the Council (as well as by the interests of Congress, the mood of the country, etc.). Finally, specific programs to implement these objectives would be proposed by the various agencies or departments concerned after a careful analysis of the alternatives available. This analysis would ensure that the Administration's social objectives were attained with the least possible expenditure of public funds—or, alternatively, that a given budget was being used in the best possible way.

Thus, social indicators would play a vital role. They would be used by the Council of Social Advisors to influence the direction of social policy. They would be used by program analysts to find the most efficient means for implementing that policy.

Conclusion

Much more discussion of the role of social indicators and social reporting is needed. Many issues remain unresolved. What social conditions should be measured? At what level of decision-making are social indicators most useful? How many of our existing statistics could be readily converted into good indicators? What new data need to be collected?

The recently-released HEW report represents a first step toward a systematic assessment of social conditions, but it is not a full-fledged Social Report. There is a need to collect more relevant data, to treat a wider range of questions, and to show how social conditions vary between different groups in our society. Former Secretary Cohen has recommended the continued allocation of staff resources in the Executive Branch for the purpose of producing a more comprehensive Social Report in the next two years. It is not yet clear whether this work will be continued, and if so whose responsibility it will be. My own view is that some kind of periodic assessment of social needs could make a significant contribution to social policy.

In the meantime, social indicators could be used much more extensively than they are at present as measures of the benefits associated with various public expenditures. This would help to improve the allocation of resources among programs and to ensure that public funds were being used as efficiently as possible.

Part III

Some Problems of Analysis in Evaluating Public Expenditure Alternatives

Section A

Risk, Uncertainty, and the Treatment of Time in Public Investment Analysis

ON THE DISCOUNT RATE FOR PUBLIC PROJECTS

BY WILLIAM J. BAUMOL

William J. Baumol is Professor of Economics at Princeton University.

"At stake in the choice of an acceptable discount rate is no less than the allocation of resources between the private and public sectors of the economy." On the basis of this observation, Professor Baumol discusses the principles which lead to the choice of an appropriate public discount rate concept. He asserts that "the correct discount rate for the evaluation of a government project is the percentage rate of return that the resources utilized would otherwise provide in the private sector." In discussing this basic proposition, Professor Baumol investigates the nature of opportunity costs and why they differ under certain circumstances. He focuses on the role of taxes in creating a disparate pattern of opportunity costs in the private sector. When a public activity draws resources from a number of sectors in the economy, "the correct social discount rate . . . will be a weighted average of the opportunity cost rate for the various sectors from which the project would draw its resources."

After discussing the difficulties in estimating an appropriate public discount rate caused by externalities and the problem of second best, Professor Baumol concludes "that with the help of the principles and data now available, one should be able to arrive at reasonable approximations for the social rate of discount. Certainly it should be possible to derive figures considerably more defensible than the conventional calculations that are all too often utilized to justify projects, not all of which are clear in their economic merit."

1. SIGNIFICANCE OF THE DISCOUNT RATE

It is generally recognized that the discount rate is a critical datum for the evaluation of any proposed Government project.* Even where there is little basic disagreement about the investment's prospective costs and benefits, the choice of discount rate figure may make the difference between acceptance and rejection. A project which seems to yield substantial net benefits when evaluated at a 3 percent rate may well appear extremely wasteful if the rate is 5 percent.

Yet, despite the critical nature of this parameter, in some calculations it is assigned a value almost cavalierly, with little attempt to show that the selected figure is not chosen arbitrarily and capriciously.** One sometimes encounters discount figures in cost-benefit calculations whose sole justification seems to be that similar figures were used in the past. Of course that can never be an acceptable argument, first because the earlier figure may have had as little justification as the one presently employed and, second, because changing circumstances change the appropriate value of the discount rate.

At stake in the choice of an acceptable discount rate is no less than the allocation of resources between the private and the public sectors

*Further discussion of this issue is found in the paper by Krutilla in this volume.

**Further discussion of this issue is found in the papers by Hirshleifer & Shapiro in this volume, and Searl in vol. 3 of this collection.

of the economy. The discount rate, by indicating what Government projects should be undertaken, can determine the proportion of the economy's activity that is operated by governmental agencies, and hence, the proportion that remains in the hands of private enterprise.

Moreover, even given the decision on resource allocation between the private and the public sectors, the choice of discount rate can affect profoundly the type of projects undertaken by Government agencies. When the discount rate is low this will encourage highly durable investments, the bulk of whose benefits will become available only in the distant future. Thus, when the economy is beset by urgent and immediate investment needs a relatively high discount rate will tend to be appropriate.

With so much at issue it is well worth exploring in some detail the principles that should be employed in arriving at a discount figure, and the rationale that underlies them. This paper undertakes to do so without heavy reliance on the jargon of technical economics.

2. The Basic Criterion: Opportunity Cost

The observation that the discount rate is the arbiter of the allocation of resources between private and public enterprise is the key to the principles which underlie the choice of an acceptable discount figure. The right discount rate becomes that number which indicates correctly when resources should be transferred from one sector to another.

More specifically, suppose one is considering the construction of a dam which will employ x man-hours of labor, y tons of cement, and z kilowatt hours of electricity. In an economy in which the level of employment is high, if those resources are made available to the Government they must be transferred out of the private sector. Just as in the guns versus butter case, each item added to the public sector involves some corresponding reduction in the volume of resources in private hands.[1]

We may now establish a rather obvious criterion to test the desirability of the proposed resource transfer. If the resources in question produce a rate of return in the private sector which society evaluates at r percent, then the resources should be transferred to the public project if that project yields a return greater than r percent. They should be left in private hands if their potential earnings in the proposed government investment is less than r percent. The logic of this criterion is self-evident. It states no more than the minimal dictate of efficiency: Never take resources out of a use where they bring in (say) 9 percent in order to utilize them in a manner which yields only 6 percent!

The standard which has just been described is the concept economists call *opportunity cost*. We have stated, in effect, that the proper

[1] On the other hand, if any resources used by the Government project would otherwise be totally unemployed, now and in the future, their use obviously incurs no opportunity cost in the private sector. In an economy such as that of the United States since World War II employment of resources has usually been so high that this consideration is often largely irrelevant to the facts of the matter, except when a project is designed specifically to make use of idle resources, e.g., to provide jobs to the unemployed. However, some types of project, or projects undertaken in particular sectors of the economy may well utilize a relatively large proportion of resources and particularly labor, that might otherwise go unusued. For a discussion of this phenomenon and its consequences see [5] and also the paper by Haveman in this volume.

criterion on which to judge the desirability of a governmental project, from the point of view of the general welfare, is the value of the opportunities which the private sector must pass by when the resources are withdrawn from that sector. A government project is desirable if and only if the value of the net benefits which it promises exceeds the cost of the lost productive opportunities which that investment causes.

It follows almost immediately that *the correct discount rate for the evaluation of a government project is the percentage rate of return that the resources utilized would otherwise provide in the private sector.* That is, the correct discount rate is the opportunity cost in terms of the potential rate of return in alternative uses on the resources that would be utilized by the project. An example will readily show why this must be so. Suppose these resources are capable of returning our hypothetical 9 percent in the private sector. Consider three proposed government projects: Project *A* which offers an average rate of return of 12 percent, project *B* whose expected return is 9 percent, and project *C* whose anticipated yield is 7 percent. It should be obvious that if we discount the returns of project *C* at the opportunity rate of 9 percent we will end up with a negative net present value figure (i.e., 7 percent discounted at 9 percent comes out to less than the principal invested). If we discount project *B*'s expected returns at 9 percent we will obtain a zero figure for the present value of net benefits (the returns will just cover the cost of the investment). Only project *A*, when discounted at 9 percent, will be assigned a positive net benefit figure. Thus the discount rate calculated at the opportunity rate works just as it should: it passes projects whose yield is greater than its resources could earn in the private sector and turns down projects whose benefits are not equal to the earnings the resources could provide in private hands.

The same illustration also shows immediately how an incorrect choice of discount figure—one not equal to the opportunity rate—can result in decisions harmful to the general welfare. For example, consider two extreme possibilities in terms of our hypothetical figures—a discount rate that is much too high (say, 15 percent) and one that is much too low (say, 5 percent). At the excessive 15 percent figure the usual cost-benefit criterion would reject all three projects, even project *A*. The Government would then fail to undertake an investment that clearly represents an efficient use of society's resources. On the other hand, a 5 percent discount rate would, on a cost-benefit criterion, lead to the construction of all three projects. Thus even investment *C* would be undertaken even though it uses resources that should better be left in private hands. For it takes resources from employments in which they return 9 percent and puts them into an occupation in which they bring in only 7 percent, a palpable 2 percent net loss to the community.

The upshot is perfectly clear. Any discount rate that is clearly above or clearly below the opportunity cost rate is indefensible because either of these will lead to decisions that reduce the general welfare. We must turn therefore to an investigation of the opportunity cost rate of resources, for once we have determined this we will have the requisite information for the choice of discount rate to be used in the evaluation of Government projects.

3. Reasons for Differences in Opportunity Costs*

Matters would be relatively simple if any batch of resources withdrawn from the private sector were to incur the same (percentage) opportunity cost as any other batch. If an opportunity cost of r percent were to apply throughout the economy, to determine the social disocunt rate one would simply proceed to estimate this number, r, and the resulting figure would then be the discount rate.

Unfortunately, for a variety of reasons, the magnitude of the opportunity cost varies with the source from which a project's resources are obtained. As we will see, resources which might otherwise have been used by a corporation will generally incur a higher opportunity cost than do resources drawn from direct consumer use.

In some sense, differences in opportunity costs of resources taken from different sectors of the economy are all a manifestation of imperfections in the market. In principle, if capital could flow without restriction to any sector of the economy where its earnings would be increased and all returns on capital were to accrue to its investors, resources would be forced by the market process to yield the same rate of return in every use. The standard argument is easily summarized. Suppose in such an economy that there were two industries, one of which returned k percent on capital while the other yielded some higher return, say $2k$. In that case investors would find it profitable to withdraw funds from the first of these industries and put them to work in the second. But the growing scarcity of capital in the first industry would tend to reduce its outputs, raising its prices and (therefore) its rate of return. Similarly, in the second industry the influx of capital would expand production, and force prices and returns downward. This transfer of capital would continue so long as any substantial difference in the two rates of return persisted. The flow of capital would cease only when the rate of return in the first industry had risen sufficiently and that in the second had fallen enough to make the two rates of return equal. At that point, clearly, capital withdrawn from the one sector would have exactly the same rate of return as capital withdrawn from the other—their opportunity cost rates would be identical.

Why does such equalization generally not occur in practice? There are two broad reasons: First, part of the return to capital may not accrue to the investor. A prime example of this is provided by taxation which siphons part of that return off to the Government, and since the burden of taxation varies from industry to industry, from the corporate to the noncorporate sector, and so forth, it means that opportunity costs will vary accordingly. A second broad reason for such variation lies in impediments, such as monopoly, to the influx of capital into some productive activities.[2] Specifically, we will consider three broad sources of such divergence because they must be taken into explicit account in the determination of the social rate of discount. These elements are the following: (a) taxes, (b) risk, (c) divergence between private and social benefits (externalities). Let us examine them in turn.

---

[2] For the relevance of monopolistic elements to the social discount rate calculation, see [2], p. 791, footnote 2.

*Further discussion of this issue is found in the paper by Hirshleifer & Shapiro in this volume.

4. The Role of Taxes: Analysis

As already indicated, the burden of taxes will vary from industry to industry, from one class of producer to another, depending on its mode of organization—that is, corporation, partnership, or do-it-yourself production by the consumer himself.

However, let us, to begin with, assume for simplicity that there are only two classes of producer: The corporation, which pays a 50-percent corporate income tax, and the private individual who produces certain items for his own use, and who pays no taxes on this production process. Assume also that there are available for sale riskless Government bonds offering a rate of return of, say, 5 percent.

In our example, it is easy to determine the opportunity cost of resources withdrawn from the corporate sector. Corporations must yield to their investors an after-tax return of 5 percent. For otherwise investors would be unwilling to provide any funds to the corporations and would instead put all their money into the safe Government bonds. But the required after-tax return of 5 percent on corporate capital means that these firms must earn 10 percent on capital *before* payment of the 50 percent corporate profit tax. In other words, the presence of special taxes on the output of this sector means that resources invested in it must produce goods and services valued at a level sufficiently high to yield a 10 percent return. The corporation can then engage only in the production of consumers' or producers' goods whose purchasers value them sufficiently to pay a price that yields a 10 percent return on corporate investment. A withdrawal of resources from the corporation, then, will cause a reduction in output whose opportunity cost in terms of consumer valuation is given by that figure: 10 percent.

Notice in evaluating resources from the corporate sector in this way, it does not matter whether the resources are used by the firm to produce consumers' goods or producers' goods (e.g., machinery). So long as the output in question is supplied by a corporation, the resources used in its manufacture will have a 10 percent opportunity cost. The relevance of this consideration will be clear presently.

Let us turn now to the more difficult problem, the opportunity cost of resources withdrawn from the other productive sector in our imaginary economy, the do-it-yourself consumers.

Since goods produced by consumers for their own use do not provide a rate of return that is measurable directly it is necessary to find some indirect means by which their opportunity cost can be inferred. Fortunately there is a straightforward way in which this can be done. When a consumer voluntarily purchases one of the bonds available in our imaginary economy he is indicating that their 5 percent rate of return compensates him for giving up the use of that money in his do-it-yourself operation. Put another way, if this investment in his own production were worth more than 5 percent to him he would not buy the bond, while if his investment in his own work were worth less than 5 percent to him he would purchase more bonds than he does. Thus without any conscious calculation on his part, a consumer's security purchases reveal something about what the rate of return of investment in the production of goods for his own consumption is worth to him—the opportunity cost incurred when a

dollar's worth of resources is prevented from going to him and is transferred to a Government investment project.

What about consumers who do not buy any bonds? It follows from the preceding discussion that their opportunity cost must be at least as high and probably higher than it is for bond purchasers.[3] To a man who chooses *not* to purchase bonds at 5 percent, the purpose for which he uses his money must by definition be worth more than 5 percent. This is no less true for a man who fails to purchase bonds because "he cannot afford them." That phrase merely implies that his consumption dollar means a great deal to him—perhaps even survival itself, and is therefore certainly more valuable than 5 percent.

We may summarize by saying that the opportunity cost of present consumption to nonbondholders must be at least as great as the figure for bondholders, and that to bondholders the opportunity cost of resources is indicated by the rate of return on their bondholdings.

In practice, of course, there is a broad spectrum of earnings tax conditions, among which are included the untaxed do-it-yourself productive process and corporate production with its (roughly) 50 percent profits tax. We have seen that the opportunity cost of resources from the untaxed sector (assuming the absence of risk and other such complications) is, roughly, r percent. In the corporation, half of whose earnings are taxed away, the opportunity cost rate is $2r$, where r is the bond rate of interest. Similarly, if the earnings of a partnership in a particular industry were reduced by, say, one-fourth through taxation, that firm would also have to earn enough before taxes to yield r percent after taxes to its investors. This company's pretax rate of return, call it s percent, must then satisfy the relationship

$$s-(1/4)s=r, \text{ that is, } (3/4)s=r \text{ or } s=(4/3)r.$$

This figure, which obviously lies between r and $2r$, will then be the opportunity cost rate for resources withdrawn from a firm paying 1/4 of its earnings in taxes.

5. Taxes and the Social Discount Rate: General Principles [4]

We may then generalize our results as follows:

(a) For production that is not subject to taxation the opportunity cost rate will generally be no lower than, and as an approximation, be equal to the r percent rate of return on riskless bonds.

(b) For production whose returns are taxed at a rate such that the fraction $1/k$ of the returns *remains* after taxes, then the opportunity cost rate for resources withdrawn from this sector will be kr percent. Thus, in the preceding example, $1/k$, the fraction of the partnership's pretax returns remaining after taxes, is ¾. Therefore, the opportunity cost rate for its resources is $kr=4/3r$ percent.

---

[3] In particular, the do-it-yourself producer may not only invest his own money in his productive activity, but may in fact borrow to help in its financing. In that case his productive activity must to him be worth at least as much as the rate of interest on the funds that he borrows, for otherwise he would not be willing to pay that amount for a loan. Since the rate of interest on loans to an individual can be expected to exceed the yield on a government bond of comparable life, the basic proposition enunciated in the text still holds for such a borrower—the opportunity cost of resources coming from him must be at least as high as the bond rate of interest (and it may well be significantly higher).

[4] This section draws heavily on an unpublished note [13] by Prof. David D. Ramsey of the University of Missouri. That note pointed out some shortcomings in an earlier version of my analysis [2].

What can be done with the diverse opportunity cost figures for resources drawn from different sectors in calculating a single discount rate for the evaluation of a Government project? The answer is simple. Suppose the Government is to withdraw some resources from the one type of productive activity and some from the other. Suppose, for example, 20 percent of these resources would have come from do-it-yourself production yielding a 5-percent return, and 80 percent from the corporate sector with its 10 percent opportunity cost yield. Then the Government project will offer a net benefit to society if on the whole it can produce a rate of return equal to their weighted average:

$$(20/100)\times 5\% + (80/100)\times 10\% = 9\%.$$

More generally, we have as a third rule:

(c) The correct social discount rate for a project will be a weighted average of the opportunity cost rate for the various sectors from which the project would draw its resources, and the weight for each such sector in this average is the proportion of the total resources that would come from that sector.[5]

The preceding rule, incdentally, will be seen to hold where there are differences in opportunity cost rates whatever their origin, whether produced by variation in the tax treatment of different economic sectors, or differences arising from any other source. The three rules (a), (b), and (c) also hold for an economy with many more than the two sectors in our simple illustration.

Before turning to the next general topic, it should be emphasized that the preceding analysis was couched entirely in terms of productive sectors and had nothing to do with the distinction between consumption and investment goods. A corporation can produce either or both of these two types of output and the same opportunity cost rate will apply to both. This means, if the analysis is valid, that the correct social discount rate will tend to be considerably higher than it would appear to be on the basis of some of the more standard modes of calculation.[6]

---

[5] The weighted average expression for the social discount rate can be derived by the following elementary argument. Let k dollars be invested in the Government project and let v be its value (including benefit flow) one year later. Suppose, if the project had not been built, that the proportion p of the investment would have been employed in an activity yielding r percent, and that the remainder, $(1-p)$, would have been used in an investment with an s percent yield. In total, at the end of the year the alternative use would then have yielded the total value

$$pk(1+r)+(1-p)k(1+s)=k[1+pr+(1-p)s]\ .$$

The Government's use of the k dollar investment will therefore be socially preferable to the private use if and only if

$$v>k[1+pr+(1-p)s],\ \text{i.e.,}\ k<\frac{v}{1+[pr+(1-p)s]}\ .$$

Thus, the discount rate is given by $pr+(1-p)s$, the weighted average of the returns in the two sectors described in rule (c) in the text. This argument obviously also applies directly to the case where resources are taken from more than two economic sectors.

[6] See, for example, Krutilla and Eckstein [8] and Eckstein [3]. On the other hand, I suspect that the radically different discount rate calculation recommended by Harberger [4] will yield results very similar to those that would be obtained by the methods proposed in the text. There also seems to be no conflict between these methods and the views expressed in the Joint Economic Committee's report ([7], pp. 13–14), which are deliberately (and, no doubt, wisely) vague on this issue, stating "while advocating the opportunity cost of displaced private spending as a correct conceptual basis for the Government discount rate, the subcommittee does not presume to advocate a precise method for calculating this rate. . . . We agree with the Bureau of the Budget's reluctance to 'adopt . . . the rate of return on private investment foregoing alone because government funds are withdrawn from both consumption and investment.'" The text of this paper obviously recognizes the relevance of the distinction between consumption and investment but maintains that the proper role of consumption in the discount rate of calculation is not quite what it is sometimes taken to be.

which, in terms of our simple example again, would employ the bond rate r for the discounting of all relevant consumers' goods production and the corporate rate only to (corporate) investment. An illustration will make the difference clear. Suppose the opportunity cost rates for our two sectors were the figures in our previous illustration (5 and 10 percent), and suppose the resources to be employed in the proposed project would otherwise have been divided into the following proportions:

	Percent
Consumers' goods production by consumer sector	20
Consumers' goods production by corporate sector	70
Producers' goods production by corporate sector	10

Since the conventional calculation applies the 5 percent bond discount rate to all consumers' goods, its weighted average calculation of the social discount rate would be

$$[(20+70)/100]\times 5\% + [10/100]\times 10\% = 5.5\%.$$

But, as has been shown, even consumers' goods produced by the corporate sector must yield a 10-percent return on investment, and hence the correct weighted opportunity cost rate must be (as before)

$$[20/100]\times 5\% + [(70+10)/100]\times 10\% = 9\%,$$

a very substantial difference.[7] It may be noted, incidentally, that the orders of magnitude of the figures are not totally unrealistic.

A final conclusion to be derived from the analysis is the implication that the appropriate discount rate may vary from one proposed Government project to another.* If they derive their resources from different sectors or vary in the proportions in which they draw upon the different sectors, then their opportunity costs may vary. Consequently, it may sometimes be possible to reduce the opportunity cost rate for the project by careful planning of the means by which its resources are obtained. In particular, as has already been noted, a project designed to draw heavily on resources which would otherwise be unemployed will for this reason incur an opportunity cost that is particularly low.

---

[7] The recommended procedure does of course conflict with the approach that has been taken in a number of highly reputable studies. It is therefore important to be very clear about the case for taking the opportunity cost of consumers' goods produced by a corporation to be dependent on the corporate income tax. The argument can be put very simply: assume that without a corporation tax all consumers' goods are associated with an r percent opportunity cost rate. Then the imposition of such a tax will lead the corporate sector to reduce its outputs and hence to raise the marginal utility of these goods to their consumers. Therefore a reduction of c units of corporate output of consumers' goods after the tax has gone into effect will incur a greater loss of consumer utility than the same c units of decrease in output would have caused before the tax. Hence if the opportunity cost was r when there was no tax it *must* exceed r percent after the tax has taken effect.

The argument clearly runs into all sorts of complications in practice in a world of risk and imperfect competition. But if the corporations were perfect competitors in a riskless economy it would apply precisely. Then the imposition of a tax would lead a sufficient number of firms to leave the production of any consumers' good to raise the pretax rate of return to $2r$ where r is the pure rate of interest. The marginal cost of a unit of output (excluding marginal return to capital) would, in equilibrium, equal price minus $2r\triangle k$ where $\triangle k$ is the incremental capital needed to produce the output, and hence the opportunity cost per unit of capital moved out of this productive activity must then be $2r$.

Note that this argument is independent of the incidence of the tax, i.e., it must hold whether or not price has gone up by the amount of the tax, or has even gone up at all.

*Further discussion of this issue is found in the paper by Hirshleifer & Shapiro in this volume.

6. The Role of Risk*

Risk will also typically vary from industry to industry and from firm to firm. This produces further differences in reported rates of return. A risky industry may be expected to offer comparatively high profit prospects in order to be attractive to investors. That is why returns on stocks are typically higher than bond yields and why the rate of interest on a corporate bond is normally higher than that on a comparable Government security.

The nature of the issue this poses for the calculation for a social discount rate as well as its resolution can be brought out clearly with another simple example. Suppose two firms which are in all other aspects alike produce respective rates of return before taxes of 10 and 12 percent. The 2-percent excess in the return of the riskier firm is, in effect, an insurance premium to compensate investors for the greater gamble involved in offering funds to the less safe of the two companies. The question then is whether resources withdrawn from the second firm incur an opportunity cost of 12 percent, or whether the true (riskless) opportunity cost is merely 10 percent as it is for the other firm.

One may argue that in fact the two companies yield the same real rate of return to their investors. Since the latter's 2 percent higher return merely covers investors' risk, as already noted, the real loss to investors of withdrawal of a unit of resources will therefore be equal.

However, a number of noted economists (see, e.g., Arrow [1], Hirshleifer [6] and Samuelson [14]) have argued with considerable cogency that the risk for the investment in an individual project taken by itself does not necessarily correspond to the risk it contributes to society's investment as a whole.[8] A bankrupt firm which continues to operate in the hands of its receivers is a loss to its investors but not to society. More important, because in the economy a very large number of projects are underway at any given time, the insurance principle, that is, the statistical "law" of large numbers, applies to them if their prospects for success are largely independent. It is true that because a recession can produce serious consequences for a large number of investment projects simultaneously, the prospects for their success are not completely unrelated. But if, as an approximation, we ignore this complication it can be shown that the larger the number of projects involved the smaller the risk contributed to total social return by any one of them. Taken as a whole, from the viewpoint of society, they become virtually riskless.

This is essentially the principle on which a life insurance company operates. It cannot tell when any one of its policyholders will expire and, taken by itself, the risk of insuring him is enormous. But when it insures a large body of policyholders the operation becomes *virtually riskless*. That is, barring calamities such as epidemics or wars,

[8] It should be noted, incidentally, that risk affects governmental as well as private investment projects. For example, many abandoned canals were no doubt built in the anticipation that their use would continue much longer than it did. And just as bankruptcy may prevent the completion of a private investment project, a change in administration after an election may result in the abandonment of some unfinished public undertaking.

*Further discussion of this issue is found in the paper by Hirshleifer & Shapiro in this volume.

the insurance company can forecast with virtual certainty how many of its policyholders will die in any given year, and the residual risk approaches zero as the number of policyholders increases.

Viewed in this way, we must conclude that while the risk involved in any individual investment project is apt to be substantial for the supplier of capital, it may be negligible from the point of view of society as a whole.[9]

Only one more step is needed to see what these considerations imply for the role of risk in the evaluation of the social discount rate. Going back to our illustration, consider once again the 12-percent rate of return of our riskier firm. The 2-percent differential between this return and that of the safer company represents, as we saw, no net gain to the investor who is merely compensated thereby for his risk relative to investment in the other company. But since to society either firm's investment program is virtually riskless, it follows that the social rates of return of the two are *not* equal. One produces a riskless 12 percent while the other yields a riskless 10 percent. This means that the corresponding social discount rate of resources drawn from these two firms would be a weighted average of these two rates, in accord with rule (c) of the preceding section.

Specifically, it means that in the social discount rate calculation there should be no deductions for a risk component in the rates of return observed in practice. When a corporation produces a before-tax yield of 16 percent, 6 percent of which might be judged as a risk premium, from the point of view of the social discount rate the corresponding opportunity cost rate is 16 and not 10 percent. For here risk plays the same role as taxation in its effects on the social rate of return. Private risk forces the firm to invest only in opportunities which offer a yield of 16 percent. The social opportunity cost is therefore correspondingly high.

We conclude that the correct discount rate is a weighted average, over all tax and risk circumstances, of the rates of return that would otherwise be earned by the resources to be used in a Government project. As already stated, rule (c) then holds, at least approximately, for variations in rates of return resulting from differences in risk as well as from disparities in tax rates.

7. Externalities *

There is one other element important for the social discount rate calculation, though we may treat it rather briefly. It is a matter that has been given considerable emphasis in the economic literature, and justly so. This is the issue of externalities—the differences between social and private costs and benefits of an investment program, of which risk compensation is a special case, as we have just seen.

The point is easily made with the aid of the most standard of illustrations of an externality. The firm whose output pollutes the atmosphere obviously provides a net social return significantly lower than

[9] I have made this point elsewhere in a slightly more formal manner (see [2], p. 794–5). Note that on this point my position differs somewhat from that of the Joint Economic Committee ([7], pp. 13–14) which calls for a special allowance in the discount rate applied to a particular public project for the risk to which that project is subject. On the argument in the text, the *incremental* risk contributed by any one such project may plausibly be considered negligible, and hence requires little or no discount rate adjustment.

*Further discussion of this issue is found in the papers by Davis & Kamien, and Kneese & d'arge in this volume.

the figure given by the yield on its capital. Similarly, the company that produces external benefits; perhaps training of the underprivileged, yields a social product greater than is indicated by its financial returns. In calculating the social discount rate, corresponding adjustments naturally have to be made in the discount figure. The polluting firm's rate of return figure has to be adjusted downward in order to base the calculation on its true social yield, and a corresponding upward adjustment is required to obtain the true social opportunity cost of the company whose operations provide training to the unskilled.

This principle is one that would no doubt be accepted by virtually all economists. Yet one must not overlook some reservations relating to its implementation. First, one must recognize difficulties that beset the evaluation of externalities. The cases where investigators have succeeded in providing some quantification of external costs and benefits are extremely rare. Many types of externality (that is, the impairment of health that is caused by pollutants) simply are not readily translated into financial terms; for example, into a numerical evaluation of their consequences for the social rate of return. Even where such a translation is possible, the effects are typically so widely diffused among the population that an estimate of their aggregate value becomes very difficult. These remarks are of course not meant to question the validity of the principle under discussion—the need to adjust the discount rate for the *social* costs and benefits of the resources concerned. It does however suggest that anything but the roughest sort of implementation of this principle will not be very easy.

A second comment on the externalities issue is the fact, not often alluded to in the literature, that governmental projects can also produce externalities. Army trucks pollute the air as effectively as private vehicles, and government work also can, as a byproduct, teach skills to ghetto residents. This means that in judging the desirability of a proposed Government project adjustments for externalities must be made on both sides of the ledger—the estimated yields of both the private and the governmental projects competing for resources must be revised in order to arrive at *net* social contribution.

8. Subsidy for the Future

There is one argument in the literature which has been used to maintain that the weighted average rate of return such as that proposed in rule (c) yields too high a social discount rate. This view maintains that private citizens tend systematically not to invest enough from the viewpoint of the Nation's future. It concludes that the discount rate should be kept very low in order to induce an increase in investment today, as a contribution to the Nation's welfare tomorrow.[10] A low discount rate will clearly lead to approval of a large number of governmental investment projects, and the greater the number of

[10] This view is clearly associated with the work of Pigou. See [12], part I, chapter II. Among others, Marglin also seemed to take such a stand in an earlier paper [10]. In a more recent discussion [11] his position is hedged considerably more. He argues "that individuals may not have unique preference maps, that instead preferences may depend on the role they play. . . . The preferences with respect to present and future which individuals hold in their political life as citizens may differ from the preferences they hold in their economic roles as consumers and producers." (Pp. 15–16). However, he does not indicate whether he believes the difference between the two follows any consistent pattern, one way or the other.

investment projects undertaken, the greater is the contribution to the Nation's prosperity.

Such a position is, however, questionable in several respects. First of all, there is no basis for the presumption that a transfer of resources from the private to the public sector will necessarily produce a net increase in the amount of investment undertaken by the economy. The increase in the amount of governmental capital construction is very likely to be offset, at least in part, by a decline in private investments in plant and equipment.

Surely, if society's investment for the future is considered to be inadequate the appropriate remedy is to institute simultaneous inducements to both private and public capital formation.[11] As we have seen, an arbitrarily low discount rate on public projects is certainly not the instrument for that purpose. Such artificially low discount rate on public projects introduce serious inefficiencies into the investment process by causing the withdrawal of resources from areas of use in which their yield is high and leading to their transfer to areas in which their return is low. This is a most peculiar way to encourage more effective investment in the future.

In any event, those who maintain that there is inadequate provision for the future draw incorrect general inferences from irrelevant particular cases. It is difficult to quarrel with the conservationist's view that the destruction of irreplaceable natural resources imposes a heavy cost on our posterity. The destruction of a portion of a canyon, the extinction of a species of wildlife, the erosion of soil all are matters of serious concern because once done they cannot be undone, and this is precisely the legitimate ground on which conservationists urge increased care in avoiding depletion of our resources.[12]

But it is not legitimate to jump from the valid point that one generation has no right to use up wastefully the resources that cannot be replaced by its successsors, to the questionable conclusion that each generation is constrained to engage in overall efforts to support its posterity beyond the level that is indicated by the free market. For that is precisely what is involved in a program of low discount rates or any other program of special inducements to investment.

The basic issue is whether it is desirable to subsidize in this way the commitment to the future of resources which could otherwise serve society today. Considerable real investment is already provided by the private sector and by the program of government projects which can pass the standard cost-benefit test utilizing the discount rate that

---

[11] This viewpoint is also taken by the Joint Economic Committee report ([7], p. 11). There are of course important cases where it is considered socially desirable to stimulate governmental output of goods and services which cannot be provided effectively by the private sector. Education, elimination of pollution and national defense are all services whose supply, it is generally felt, should not be left exclusively in the hands of the private sector. In such cases Government subsidy, and in some cases complete Governmental financing, is entirely appropriate because of the external benefits contributed by these activities. But this only requires us to take explicit account of such externalities in our cost-benefit calculations. It does not justify a particularly low *discount rate* on the Government project which would distort the allocation of resources between short- and long-term investments. If we need more expenditure on education now—better books and better teachers—a reduction in the discount rate would not provide them. It would only stimulate the construction of durable school buildings, the long-term investment portion of educational-expenditure.

[12] Paradoxically, in practice a low discount rate will probably increase the destruction of such resources. In the past low discount rates have been used to provide a cost-benefit justification for precisely those engineering projects which have constituted, at least according to conservationists, the most serious threats to national parklands, recreation areas and other such resources.

has already been described. Is there any justification for a program of subsidies designed to produce an even greater overall commitment of resources to the future?

A little thought shows that the grounds for such a program are tenuous at best. Our economy is characterized by a long run rate of growth of GNP of the order of 3 to 4 percent per year, compounded, and in recent years the growth rate has been even greater. Per capita income has risen persistently throughout our history and there is every reason to expect the rise to continue. We are therefore wealthier than our predecessors and it can quite safely be predicted that our successors will be richer than we. In effect, then, the subsidization of a program of added investment amounts to an inducement for the transfer of additional resources from the poor to the rich. It would take inputs whose product would be available for consumption today and make them available tomorrow when the supply of consumer goods is likely to be so much more abundant than it is at present.[13]

It should be made clear that the preceding discussion is not meant to imply that there is no need for increased activity by the public sector. The crisis of our cities, the problems of the impoverished and the underprivileged minorities and a variety of other critical issues may well require for their resolution governmental activity on an increased scale. But these all call for investments whose yield is obtainable quickly, not long-term investments, the bulk of whose benefits will become available in the more distant future. Advocacy of a very low discount rate in these circumstances is tantamount to the view that these immediate problems are not very pressing, and that society's resources are better transferred to the service of the future generations who in any case will surely be wealthier than we.

9. The Problem of the Second Best

One technical problem besets the analysis of the social discount rate presented in this paper. This is the well-known problem of the second best. We have seen how taxes, risk, and externalities produce discrepancies in the rates of return earned by various parts of the private sector. This means that capital simply is not allocated optimally within that sector. Society could obviously benefit by a transfer of capital from areas in which its social rate of return is relatively low to other parts of the economy offering a higher social return.

It is, then, not necessarily true that the rules of optimal investment procedure for the public sector are applicable, unless they are simultaneously put into operation in the private portion of the economy as well. For it has long been shown (see [9]) that if an optimality rule is enforced in one sector of the economy alone it can actually make matters worse.

In the present case, however, it does not seem plausible that this problem will be very serious. By the nature of the opportunity cost test, it calls for a transfer of resources only from a use in which its yield is relatively low to one where its return is comparatively high. This implies that the social welfare can be expected to increase in the

[13] This argument is based on Tullock [15]. For a similar position, see also the Joint Economic Committee report [7], p. 11.

process, assuming of course that there has been adequate adjustment for differences between private and social returns, so that the yield and opportunity cost figures take into account the more indirect contributions of the resources in question.

10. Concluding Comment

There always remains a formidable task of implementation even after the basic principles in an area have been enunciated. But a rational decision process does require examination of those principles. It is primarily to this task that this paper has addressed itself. It has proposed what is, in appearance, a relatively simple formula for the evaluation of the social discount rate, the weighted average of the rates of return in the various productive sectors from which resources would be withdrawn for the government project under consideration.

But the simplicity of this expression is deceptive. It is not easy to determine in practice from what productive sector a given project will be drawn, or even from where governmental investment as a whole obtains its resources. What one seeks in trying to obtain this information is the catalog of the decrements in the outputs of the various portions of the private sector which would result from a governmental investment program. When the matter is put in this way, its difficulties should be obvious.

Nor is it easy to judge the rate of return figures for the various sectors of the economy. A corporation offers different nominal rates of return on its various types of debt and equity and, as the recent literature on corporation finance has brought out so forcefully, it is not easy to arrive on the basis of these figures at a single number representing the rate of return on the company's capital. Perhaps a reasonable rule of thumb is the traditional weighted average of the rates of return on a firm's stocks, its bonds, and its other securities, where each of these sources of funds is weighted by the proportion of the company's total finances which it contributes.

In sum, the calculation of the social rate of discount is not something that can be left simply to traditional practices and conventions. The underlying principles should enable us, by a reasonable process of approximation to arrive at operational procedures that represent a significant improvement over current procedures. I have sought to emphasize the complications in order to avoid the impression that all the issues have been settled and can readily be reduced to routine. But I believe firmly that with the help of the principles and the data now available one should be able to arrive at reasonable approximations to the social rate of discount. Certainly it should be possible to derive figures considerably more defensible than the conventional calculations that are all too often utilized to justify projects not all of which are clear in their economic merit.

REFERENCES

[1] K. J. Arrow, "Discounting and Public Investment Criteria," in A. V. Kneese and S. C. Smith, eds., *Water Research*, Baltimore, 1966.

[2] W. J. Baumol, "On the Social Rate of Discount," *American Economic Review*, September 1968, pp. 788–802.

[3] Otto Eckstein, *Water Resources Development: The Economics of Project Evaluation*, Cambridge, Mass., 1961.

[4] Arnold Harberger, "Statement," in *Hearings Before the Subcommittee on Economy in Government of the Joint Economic Committee of the Congress of the United States*, July 30–August 1, 1968, Washington, 1968.

[5] R. Haveman and J. V. Krutilla, "Unemployment, Excess Capacity, and Benefit-Cost Investment Criteria," *Review of Economics and Statistics*, Vol. 49, No. 3, August 1967.

[6] Jack Hirshleifer, "Investment Decisions Under Uncertainty: Applications of the State-Preference Approach," *Quarterly Journal of Economics*, May 1966, *80*, 252–77.

[7] Joint Economic Committee, *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis*, Washington, 1968.

[8] J. V. Krutilla and O. Eckstein, *Multiple Purpose River Development, Studies in Applied Economic Analysis*, Baltimore, 1958.

[9] R. Lipsey and K. Lancaster, "The General Theory of the Second Best," *Review of Economic Studies*, 1956, *24*, 11–32.

[10] S. A. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quarterly Journal of Economics*, February 1963, 77, 95–112.

[11] ———, "The Discount Rate in Public Investment Evaluation" (unpublished) December 1968.

[12] A. C. Pigou, *The Economics of Welfare*, 4th ed., London, 1932.

[13] D. D. Ramsey, "On the Social Rate of Discount: Comment," 1968 (unpublished).

[14] P. A. Samuelson, "Principles of Efficiency: Discussion," *American Economic Review*, Proc., May 1964, *54*, 93–96.

[15] Gordon Tullock, "The Social Rate of Discount and the Optimal Rate of Investment: Comment," *Quarterly Journal of Economics*, May 1964–78, 331–36.

THE TREATMENT OF RISK AND UNCERTAINTY*

BY JACK HIRSHLEIFER and DAVID L. SHAPIRO*

Jack Hirshleifer is Professor of Economics at the University of California at Los Angeles. David L. Shapiro is Assistant Professor of Economics at the University of California at Los Angeles.

While the existence of uncertainties provide the rationale for the transfer of some private sector activity to the public sector, there is very little understanding of how the Government should adjust to the existence of risk and uncertainty when evaluating alternative public undertakings. Indeed, existing practice in the Federal Government has been extremely disparate, reflecting both the conceptual difficulties in dealing appropriately with uncertainty and the state of economic thought on this matter.

Professors Hirshleifer and Shapiro attempt to integrate the substantial progress made in economic theory in recent years "with traditional viewpoints, so as to provide a framework for the evaluation of ongoing practices with respect to risk and uncertainty." Following a review of the basic propositions on which the economic analysis of public investments rests, they discuss the different meanings of the terms "risk" and "uncertainty" and the implication of the different varieties of each for public investment analysis. They find the literature of economics which deals with the problem of risk and uncertainty to contain "a good deal of confusion." For example, while some observers see risk to be a social cost, others have argued that "risk premiums observed in the market are not socially relevant." In part, the controversy concerning the size of the appropriate public discount rate has revolved about these different views of the social relevance of risk and uncertainty. In the final section of their paper, Professors Hirshleifer and Shapiro discuss some recent theoretical developments and present a "model of time-state preference" as a framework for appropriate public sector response to the existence of risk and uncertainty. This model leads them to the conclusion that the application of a riskless rate of interest in discounting is inappropriate and would result in the adoption of inefficient projects. Rather, the observed "risky rate" implicit in the evaluation of comparable private projects provides an appropriate guide to governmental evaluation practice.

I. THE PROBLEM

We live in a world all too obviously dominated by risk and uncertainty.** Facing the issue of risk and uncertainty, in the analysis of public investment decisions, is therefore no mere theoretical flourish. Rather, it is very close to the heart of the matter. Unfortunately, the conceptual difficulty of the problem is very great, and the "state of the art" in economics does not yet provide for its full solution. In consequence, benefit-cost investigations typically employ the artificial model of a world of certainty—with perhaps some intuitive or rough-and-ready *ad hoc* adjustments. Thus, costs may contain an item for "contingencies," benefits may be said to have been evaluated "conserva-

* Research and clerical assistance furnished by the Water Resources Center, University of California, is hereby acknowledged.

** Further discussion of this issue is found in the paper by Zeckhauser in this volume.

tively," or a time-discount rate higher than the "pure" (riskless) rate of interest may be employed. These procedures are not necessarily incorrect, and may in fact be quite unavoidable, but the entire enterprise shows the need of a firmer theoretical foundation.

In the past decade, theoretical economists have made very substantial progress in dealing with risk and uncertainty. This progress has not yet been fully absorbed by the economists who specialize in the applied area of cost-benefit analysis, and of course even less so by practitioners and administrators in the field. The present paper represents an attempt to integrate recent theoretical development with traditional viewpoints, so as to provide a framework for the evaluation of ongoing practices with respect to risk and uncertainty.

Any analysis must be conducted within certain "ground rules." These are a set of propositions, which might themselves be debatable or even more or less clearly in violation of "the facts," but which are taken as useful and acceptable simplifications for the sake of the argument. Among such ground rules for this discussion are:

(1) *The principle of cost-benefit analysis.* This principle asserts that policy ought to be determined by a systematic comparison of costs and benefits—rather than, for example, by inspired intuition or by the interplay of political pressure groups.

(2) *The efficiency criterion.* This criterion asserts that costs and benefits are to be calculated in terms of the overall achievement of individual desires, as expressed in the valuation of goods and services (commodities) produced or foregone. Maximizing the difference between the values of benefits and costs represents a goal akin to maximizing real national income, except that the value placed by individuals upon voluntarily chosen leisure is counted along with the values of material consumption and investment. The *distribution* of the net benefits is not taken into account, presumably because a larger total income could in principle be redistributed so that everyone benefits in comparison with whatever could be attained with a smaller aggregate income. The serious questions and objections that can be raised concerning the efficiency criterion have been discussed to some extent in other papers and will not be considered here.

(3) *Reduction to dollar values.* Actual projects may involve the production or sacrifice of a great variety of different commodities: familiar goods like wheat and electric power, as well as less obvious ones such as leisure and collective goods (e.g., national defense), and "intangibles" like better neighborhoods. For this paper, the problems involved in the treatment of multiple commodities—as well as the difficult questions connected with the evaluation of "externalities" and the consequences of market imperfections—are presumed to have been solved. (Again, these topics have been very fully discussed in a number of the other papers.) In consequence, we will need only to deal with risk and uncertainty as related to the dollar totals of costs and benefits associated with investment alternatives.

We can now consider the meaning of the terms "risk" and "uncertainty." While the words have rather different connotations ("risk" suggests the potential variability of the objective configuration of events, while "uncertainty" underlines our subjective lack of knowledge as to which configuration will obtain), no formal distinction between them will be made here. Both terms will be used to express a

situation in which, whether for objective or subjective reasons, analysis requires us to take into account the possibility of a number of alternative outcomes or consequences of actions. There is one tradition in the literature that attempts to formulate a distinction between risk and uncertainty on the basis of ability to express the possible variability of outcomes in terms of a probability distribution (Knight [18: 225–231]). According to this tradition, when we do not know the specific outcome but do know the probability distribution, we have "risk"; when we do not even know the probability distribution, we have "uncertainty." This distinction has proved to be sterile. And indeed, we cannot in practice act rationally without summarizing our information (or its converse, our uncertainty) in the form of a probability distribution (Savage [28]).

There are two different senses of the word "risk" that are often confused. In the first sense, risk is the danger that reality might somehow fall short of overoptimistic beliefs or promises. Thus, an investment adviser might say: "There is a high risk that this railroad bond, though promising a 10-percent coupon, may be defaulted." On the other hand, there is a more neutral use of the word in the technical literature. In this latter sense risk refers to the fact of variability of outcome, whether favorable or unfavorable. Thus, the same adviser might say: "Common stocks are riskier than Government bonds," by which he means that the investor holding the former might either do much better or much worse than an investor holding the latter.

The distinction may be clarified by introducing the concept of *expected value* (or *mathematical expectation*). Given a probability distribution of numerical (dollar) outcomes, the expected value is the probability-weighted average. Thus, the expected value in the toss of a single die is 3.5—the average of the equally probable outcomes 1, 2, 3, 4, 5, 6. The 10-percent coupon yield was an optimistic if-all-goes-well estimate of the actual yield from holding the railroad bond mentioned above; allowing for the probability of default, the mathematical expectation of yield might be only 5 percent. Use of the mathematical expectation is an adjustment allowing for the risk of unfavorable outcomes. It can be regarded as a *correction for optimistic bias*, and must not be confused with adjustments that might be made to allow for attitudes toward risk in the sense of variability.

In dealing with variability risk, we are no longer in a position to make allowances or corrections merely in terms of mathematics. For, an element of taste enters in: the investor's degree of risk-aversion (if he dislikes risk) or risk-preference (if he likes risk). In the case of the railroad bond with 10 percent nominal but 5 percent expectation of yield, consider the alternative of a Treasury bond yielding 5 percent with certainty. An individual characterized by risk-aversion would prefer the latter, an individual with risk-preference would prefer the former. Whereas taking the *expected value* represents a correction or allowance for bias, allowance for risk-aversion (or risk-preference) leads to the concept of the *certainty-equivalent value* of an uncertain outcome. Thus, for an individual characterized by risk-aversion, the railroad bond above might have a certainty-equivalent yield of only 4 percent—that is, be indifferent to a Treasury bond yielding 4 percent, even though the railroad bond has a mathematical expectation of yield equal to 5 percent.

27–877—69—vol. 1——34

One important issue, where the two concepts of risk have caused some confusion, concerns the ability of Government to "pool" a large number of independent investments and therefore (it has been alleged) to "ignore risk." The underlying idea here rests upon the statistical law of large numbers. This law states that, if there are a large number of independent experiments, the average outcome obtained will be very close to the mathematical expectation of outcome. It will be immediately clear that the thrust of the argument is that Government may sometimes be in a position to "ignore risk" in the sense of *variability*—the mathematical expectation of return on Government investments may become almost a certainty overall. But if the risk in question is due to the fact that returns from Government projects are typically stated in an overoptimistic way, the fact that Government engages in many such projects will in no way tend to counteract or eliminate this bias.

II. The Riskless Case*

Consideration of the principles of investment decision in the world of risk and uncertainty is best preceded by a review of the principles that would be applicable in an artificial but simpler world of certainty. For, some of the errors and confusions that bedevil analysis of the more complex case can be dispelled by a thorough understanding of the simpler certainty model.

In this certainty model we of course need not take any account of risk-preferences. But as we are dealing with intertemporal choices (investment decisions), the *time*-preferences of the members of the community are relevant. Indeed, given the assumed reduction to dollar values, we are in effect dealing with a one-commodity model in which the only distinguishable objects of choice are consumption claims now and in the future. Coordinate with the taste information summarized under the heading of time-preference is the technological information that can be summarized under the heading of time-productivity. Time-preference refers to the *willingness* of consumptive decisionmakers (individuals and households) to exchange units of current consumption for units of future consumption; time-productivity refers to the *ability* of productive decisionmakers (firms and governments) to physically convert potential consumption of the present date into potential future consumption. It is conventional to measure both time-preference and time-productivity in terms of the percentage premium in the exchange of future claims for current ones. Thus, if an individual is just willing to give up $1 of current consumption for $1.10 of consumption 1 year from now, his marginal 1-year rate of time-preference is said to be 10 percent. And if a producer can convert (invest) $1 of current consumption so as to return $1.20 1 year from now, his marginal rate of time-productivity is said to be 20 percent. Finally, if in the market $1 of titles to current consumption exchanges against $1.05 of titles to consumption 1 year off, we say that the market *rate of interest* is 5 percent.

Standard economic theory leads to a number of conclusions that will be relevant to our present discussion. These are summarized briefly below. The wording runs in terms of comparisons between present

*Further discussion of this issue is found in the paper by Baumol in this volume.

consumption claims and future claims dated 1 year from now. The generalization to a multiplicity of future dates, while not without its complications, does not introduce any really novel problems.

1. *The principle of rational intertemporal consumptive decision.* Each consumptive decisionmaker will distribute his wealth between current and future consumption in such a way that his marginal rate of time-preference is just equal to the market rate of interest. For if there were any disparity between the consumptive premium and the market premium, it would pay him to modify his consumption plans by lending or borrowing on the market.

2. *The principle of rational intertemporal productive decision.* Each productive decisionmaker will continue to invest until his marginal rate of time-productivity is just equal to the market rate of interest. For, if there were any disparity it would pay him to adjust his production plans by increasing or decreasing the scale of investment.

3. *The principle of market equilibrium.* The interaction of rational consumptive and productive decisionmakers leads to a market equilibrium characterized by an interest rate that simultaneously measures (is equal to) the marginal rates of time-preference and the marginal rates of time-productivity everywhere in the economy. For, if the latter conditions failed for even one decisionmaker, he would be trying to engage in additional transactions—the economy would not be at equilibrium.

The principles above represent "positive" conclusions from the standard behavioristic postulates of economic theory. That is, given such assumptions as rational individuals, perfect markets, etc., the statements above can be inferred as scientific predictions to be tested against observation. The fourth principle below, however, is a "normative" conclusion or policy recommendation—based on value judgments in addition to purely scientific elements.

4. *The present-value rule for government investment decision.* The government ought to adopt that set of investment projects which maximizes the net balance of the values of aggregate present sacrifice and aggregate future return, where the future return is "discounted" at the market rate of interest.

To explain this principle, it will be convenient to introduce some formal symbolism. We will denote current investment of an economic unit by i_0, and the return on investment 1 year from now by q_1 (thus, the subscript represents time from the present). The present value of the entire set of productive investments undertaken by the unit in question is defined as V_0 in:

$$V_0 = -i_0 + \frac{q_1}{1+r_1} \tag{1}$$

Here r_1 is the market rate of interest effective for exchanges between present and future funds (borrowing and lending).[a] A particular project

[a] The generalization to multiple future time-periods leads to an equation of the form:

$$V_0 = -i_0 + \frac{q_1}{1+r_1} + \frac{q_2}{(1+r_2)(1+r_1)} + \cdots + \frac{q_T}{(1+r_T)\cdots(1+r_2)(1+r_1)}$$

Here $r_1, r_2 \ldots, r_T$ are successive 1-year interest rates up to the economic horizon T.

will involve some increment Δi_0 of investment and Δq_1 of future return, so the present value *of the project* is defined as:

$$\Delta V_0 = -\Delta i_0 + \frac{\Delta q_1}{1+r_1} \tag{2}$$

Consequently, the present-value (PV) rule can be expressed in two essentially equivalent forms: (a) Maximize overall present value V_0, or (b) Adopt any and all projects whose present values ΔV_0 are positive.[1]

The justification for the PV rule runs in terms of the positive conclusions listed above, together with the value-judgment as to the efficiency criterion listed among the "ground rules" in section I. For, if the goal is to be maximization of the net balance of costs (investment) and benefits (returns from investment), and if the relative values placed by consumers upon the commodities current and future consumption are to be accepted (in the form of the ruling market rate of interest used as a discounting factor), then the government should also in its sphere attempt to maximize the value of its net output in the interests of its ultimate clients—the consumers.[2]

The present-value rule could be challenged either by the rejection of the efficiency criterion as a value judgment, or by the rejection of one or more of the positive principles listed above. Thus, consumer irrationality would cast great doubt upon the significance of the market rate of interest as representative of marginal time-preferences, while producer irrationality would have the same effect with regard to marginal time-productivities. And market disequilibrium would mean, in effect, that there is no single rate of interest serving the function of representing these magnitudes. But we will be holding to the efficiency criterion as a ground rule; [3] as for the positive principles of individual behavior, we need not discuss their "realism" in the artificial certainty model since the issue will arise even more forcefully in the context of risk and uncertainty.

One noteworthy aspect of the present-value principle is its *objective* nature. No subjective considerations of the productive decisionmakers' *personal* time-preferences enter into their decisions—but only the integrated weight of all individuals' preferences as they interact in determining the market rate of interest. It is this feature that facilitates

[1] The two formulations are strictly equivalent only when projects are independent, in the sense that adoption of any one does not affect the investment required or the returns yielded by any others. There are many possible patterns of interdependence: projects may be mutually exclusive, or one may be prerequisite to another, etc. For interdependent projects, the version (a) of the present-value principle remains valid and thus is the more fundamental conception.

[2] Here again the statement is true in simplest form only if an independence condition holds, in this case independence between the productive costs and returns on government and private projects. Should this condition fail, it might pay the government to adopt projects of negative present value in themselves, but having favorable "spillover effects" on projects of private firms (or, conceivably, of other levels of government). In principle, such spillover effects are already allowed for in the established procedures of government cost-benefit calculations.

[3] Among economists, the most prominent deviating school of thought accepts efficiency as a goal but rejects the valuation of temporal consumption that runs in terms of the supposedly biased preferences of the *present* generation—the only one in a position to cast "dollar votes." For varying points of view on this issue of "social time-preference," see: Pigou [25], Marglin [20], Feldstein [10], Hirshleifer [15]. For the purposes of this paper, we need only assume that government takes appropriate measures (by a system of taxes on present consumption and subsidies on investment for future consumption) to correct the market expression of these biased preferences. If this were done, the remainder of the analysis now running in terms of the "corrected" time-preferences would remain valid.

the voluntary organization of many individuals for combined production through instrumentalities like the corporation.

Since the present-value principle would also be naturally used by private firms in maximizing wealth of their owners, a question arises that may be put in the following form: If private firms and government agencies are both to be using exactly the same decision rule, does this model leave any real role for government as a distinct type of social agency? The point of the question is that, since there surely is a real role for government, there must be something wrong with the present-value principle. This inference would be incorrect. Private firms and government agencies can be both using the PV rule, but may be valuing benefits and costs differently. In particular, private firms would be expected to take into account only benefits and costs expressing themselves in market valuations. But government is in a position to separate the provision of the commodity or service from the attendant financial receipts and outlays. Consequently, it is able to take into account, on either the cost or benefit side, the values placed by consumers upon nonappropriable goods (e.g., national defense), positive or negative spillovers (as in water-resource management), etc. Of course, our postulated reduction to dollar values combines goods of all kinds in one analytical category, but this was solely for expository convenience. The real differences in appropriability or marketability of goods suffice to provide a role for government as a distinct productive agency.[4]

III. Review of the Literature

In the enormous literature on cost-benefit analysis for public investment decision, the problem of risk and uncertainty has received only sparse coverage.[5] Some authors raise the topic only to deny its relevance; others propose more or less specific "adjustments," but without substantial attempts at theoretical justification. (The recent development of the concept of "State-preference" does, however, provide an analytical basis for the treatment of risk—this topic will be reserved for sec. V below.)

The problem of risk obtrudes itself upon the analyst or decision-maker when he tries to determine "the" interest rate r_1—or more generally, the temporal sequence of interest rates $r_1, r_2, \ldots r_T$—for use in present-value formulas like those of section II above. For, in the world of reality we observe not a single rate but a bewildering variety of interest rates, even between a given pair of dates. Two main hypotheses come to mind in connection with this diversity. The first calls on *market imperfections* to explain why, for example, consumers must pay more for personal loans than savers receive on time deposits (e.g., Eckstein [7: 503]). The other school of thought maintains that the divergences of observed yields conceal an underlying *harmony* represented by the systematic and in principle predictable influence of risk (see, for example, Sharpe [29]). In the nature of the case, this issue can probably never be fully resolved; the hypotheses are programmatic statements rather than empirical generalizations. We shall

[4] In some cases the goods produced may be perfectly marketable, but the underlying desire is to achieve a wealth redistribution by not maximizing the attainable commercial returns. Here again, the key feature is the separation of financial compensation from provision of the service.

[5] See, for example, the recent review of cost-benefit analysis by Prest and Turvey [26].

be concerned here mainly with those analysts who do regard the divergences of realized interest rates as having at least something to do with differential risk and uncertainty.

The major analytical controversy that has occupied the literature is the following: Does the risk compensation sought by individuals—in the form of a "risk-premium" over and above the riskless rate of interest—represent a social as well as a private cost? There are also subcontroversies concerning the policy implications of each of the basic conclusions on the main issue. All these topics will be reviewed here. It should be kept in mind, however, that the risk-premium representing compensation for risk-aversion (reluctance to bear variability of outcome of investments) must always be distinguished from compensation for risk in the sense of optimistic bias in the statement of costs and benefits. Unfortunately, failure to make this distinction has been responsible for a good deal of confusion in the literature.

DOES RISK-AVERSION EXIST?

Before turning to the main issue, there is a prior question that ought to be settled. If individuals were on balance "risk-lovers," they would be willing to accept a discount to bear variability risk—rather than insisting on a market premium. And in fact, there is a tradition in economic thought which so maintains. Thus, Adam Smith declared:

> The chance of gain is by every man more or less overvalued, and the chances of loss is by most men undervalued, and by scarce any man who is in tolerable health and spirits, valued more than it is worth. ([30], book I, ch. 10.)

And of course, the phenomenon of gambling is further evidence along this line. On the other hand, Alfred Marshall came to the opposite conclusion ([22], book V, ch. 7 and book VI, ch. 8). And the institution of insurance is evidence in favor of his viewpoint.

There have been a number of attempts to explain the seeming conflict here (see Friedman and Savage [11], Markowitz [21], Hirshleifer [14]. All of these tend to support the conclusion that gambling is likely to be a relatively minor perturbation, with the bulk of substantial investment behavior subject to the influence of risk-aversion. And indeed, that risk-aversion predominates does not appear to be a matter of controversy in practice. Risk-aversion is almost always taken for granted not only in the cost-benefit literature for government investment decision, but in theoretical and applied works in the area of corporate finance and security markets (see, for example, Modigliani and Miller [24]).

RISK-AVERSION NOT A SOCIAL COST—THE POOLING ARGUMENT

The major support for the viewpoint that private risk aversion, accepted as a fact, is or ought to be socially irrevelant stems from the pooling argument. As indicated above, this argument is based upon the operation of the law of large numbers in making the mathematical expectation of outcome almost a certainty where there are many independent experiments. The position is well expressed by Vickrey:

> But insofar as riskiness is concerned, the reason risky investments carry an expected return greater than that of secure investments is that in the market, facilities for pooling of risks are

> imperfect, so that investment by private investors in risky investments given a limited portfolio and a certain indivisibility in the market, is unavoidably associated with a considerable dispersion of individual incomes. Given a certain risk aversion on the part of individuals * * * maximization of expected utility leads naturally to a bias against risky investments * * *. On the social level, however, the risk associated with a given public venture is inevitably pooled and averaged over the entire population of the country in some fashion, along with the risks of other projects, and this pooling or averaging of risks for public projects is accomplished without any cost of extra financial transactions. [32:89]

And at the same meeting, Samuelson supported Vickrey's view:

> One can look at much of government as primarily a device for mutual reinsurance. General Motors can borrow at a lower rate than American Motors because it is a pooler of more independent risks. It would be absurd for G.M. to apply the same high risk-interest discount factor to a particular venture that A.M. must apply. The same holds for We, Inc., which is a better pooler of risks than even G.M. [27:96].

The pooling argument will be vigorously evaluated in terms of the model of State-preference in section V. A few remarks may be in order here, however. Vickrey's contention shows awareness of the fact that the argument does depend in some unspecified sense upon market imperfections. However, these authors do not seem to fully appreciate that there is at least one other condition necessary for the validity of the argument. The second condition is that the returns on the private and Government investments being compared are not correlated with social variability of outcome. If the higher risk premium on American Motors is due largely to, for example, a higher correlation of outcome with the general business cycle—one type of "social risk"—it would be incorrect to say that private A.M. risks do not reflect social costs. And similarly, if the returns on Government investments vary systematically in procyclical fashion, the existence of many Government projects will not eliminate this source of variability. Kenneth Arrow [1] is aware of both of these necessary conditions, but nevertheless defends the pooling argument. He maintains that markets for reinsurance of private risks do not exist and that the returns on new investments tend to be uncorrelated with social risks. It would seem that both of these contentions are rather extreme. Diamond has indicated recently that Arrow underestimates the efficiency of security markets in effectively pooling the private sector's risks [5]. And it would seem preferable to explicitly account even if only approximately for the procyclical or anticyclical impact of particular Government investments rather than making an unwarranted assertion of neutrality on this score.

POLICY IMPLICATIONS OF SOCIALLY IRRELEVANT RISK

For the proponents of the argument that private risk aversion is a socially irrelevant cost of investment, the policy implications are treated as straightforward. Arrow's statement is typical:

> * * * the proper procedure is to compute the expected values of benefits and costs, and discount them at the riskless rate * * * [1: 28].

As will be seen below, this also would seem to provide the rationale for the existing procedures of many Government agencies.

The effect of this procedure is to encourage the expansion of public investment in the place of the private investment deterred by socially irrelevant risk. That this is the best method of coping with the problem has been challenged. For, granted the premises, it would clearly be more efficient for the Government to induce, by subsidy if necessary, adoption of the higher-expected-yield private investments not being undertaken [14: 270]. Indeed, as will be shown below, it is hardly a question of subsidy but rather a correction of the effect of a tax that places an enormous burden upon private risky investment—the corporate income tax. Opposition to a policy of divergence between public and private discount rates has also been expressed by Baumol*:

> Thus, nothing said so far argues for or against a low rate of discount. It states merely that society will not benefit if it increases long term investment in a wasteful and inefficient manner, by forcing the transfer of resources from employments with a high marginal yield to uses with a low marginal yield. For that is exactly what can be expected to result from the usual sort of figure of say 5 percent for discount rates on public projects when the corporate rate of return is perhaps three times that high. [3: 797].

Students of economics will recognize herein the familiar "second best" problem of welfare theory. In this context the "first-best" solution is the expansion of the deficient risky *private* investment, by correction of the influence of a discriminatory tax—or by subsidy, if necessary. If for some reason the "first-best" solution is unfeasible, we may want to consider "second-best" policies. But the expansion of risky *government* investments is a very inefficient "second-best," since some of the funds so used are preempting more productive investments in the private sector. Even if it were impossible to do anything to favor risky private investment, which is evidently not the case, government discounting at higher than the riskless rate is indicated—to strike a balance between the desired expansion of risky investment in general but the undesired preempting of more productive private investment.[6]

VIEWS THAT RISK IS A SOCIAL COST

Those who argue that the risk premiums observed in the market are not socially relevant have reference, of course, to premiums for bearing variability risk. We have noted earlier, however, that part of what is regarded as the market risk premium may be a correction for the optimistic bias of the borrower. We refer here to the "premium" of higher nominal yield that must be offered on securities considered risky by investors. Those who advocate the riskless rate for government investment are generally concerned, of course, that optimistic bias be

---

[6] Essentially the same point arises in the "social time-preference" argument, which maintains that there is inadequate private provision for the future since only the current generation casts "dollar votes" on the disposition of current resources. Attempts to overcome this supposed distortion by expansion of government investment alone lead once more to the same sort of difficulty (Hirshleifer [15]). In general, it is necessary to compromise between the objectives sought.

*Further discussion of this issue is found in the paper by Baumol in this volume.

eliminated from the cost-benefit stream (see Samuelson [27:96]. Where the issue is made explicit, economists seem unanimously to agree that risk in the sense of optimistic bias of proposed government investments is a social cost and must be allowed for. Some authors would prefer, however, to make the adjustment through the discount rate. Bain, Caves, and Margolis declare:

> The only general justification for introducing a "risk allowance" of one sort or another into investment calculations would be that some or all water agencies seem to have shown a propensity to make unjustifiably optimistic estimates of future benefits of projects; thus, reducing their estimates by such a means as increasing the rate of discount by two or three percentage points would compensate for their optimistic bias in estimating. [2:272]

These authors argue against tampering with the cost and benefit streams directly.

Turning to authors who seem to accept implicitly or explicitly that aversion to variability risks is a social cost to be considered in Government investment decision, the vast majority agree that adjustment has to be via the discount rate. One important exception, however, is Fred Hoffman, who as Assistant Director of the Bureau of the Budget testified:

> While I certainly do not wish to argue that Government programs are riskless—on the contrary they are often subject to considerable risk—I believe that better decisions are likely to result from considering the risks explicitly by adjusting the expected costs and benefits than by attempting to relate the average risk of peculiarly public programs to "similarly risky" investments in the private sector [17:27].

Most other authors, while possibly agreeing that such a procedure might be desirable in principle, rule it out as impractical (Hirshleifer, DeHaven, and Milliman [16: 144] or as lacking any foundation in traditional approaches to uncertainty (Dorfman [6: 149]).

The mainstream of debate on the evaluation of risky Government investment projects has turned upon the selection of the appropriate rate of discount to allow for "optimism bias" and/or "variability risk." Here the different views may be classified according to whether they presume that divergences in observed interest rates fundamentally represent (a) the influence of market segmentation or other imperfections *or* (b) the systematic and predictable influence of differing riskiness. Any such classification cannot be entirely hard-and-fast. Some authors may maintain an intermediate position, and others may not pose the issue clearly one way or the other. Nevertheless, it is possible at least to a first approximation to classify those whose views are primarily based upon the "market imperfections" hypothesis on the one hand, or the "harmony" hypothesis on the other.

For those thinking in the former terms, the inclination is to apply some across-the-board discount rate for all Government projects. Thus, Eckstein proposes a general risk premium of from one-half to 1 percent [9: 86]. In another work, Krutilla and Eckstein [19] conduct an elaborate analysis to determine the social cost of Federal financing. Clearly presuming market segmentation, they examine the differential impact upon the various investing and consuming sectors of ways of securing

Federal funds. This is done in order to provide weights for the averaging of typical interest rates or yields in these various sectors. The Government discount rate derived by this procedure incorporates some unknown average degree of risk-premium—in both the optimistic-bias and variability senses of the term. A somewhat analogous estimation procedure has been adopted by Harberger, who however recognizes the need for an explicit risk adjustment [12].

Those authors thinking primarily in terms of the harmonistic hypothesis are led to seek discount rates in the market that are somehow related to or reflective of the same sort of risks as the Government projects considered. Thus Hirshleifer, DeHaven and Milliman:

> * * * attempt to determine the *real marginal opportunity rate* which the market insists upon in providing capital to *private* companies whose investment decisions are most comparable to those of public agencies in water supply, namely, corporations in the public utility fields [16: 146].

And Professor Harberger testified:

> A third and still better approach would be to try to identify especially risky Government investments, investments of medium risk, and investments of demonstrably low risk and to make separate risk adjustments for each of these three categories. If we could simply identify investments in those three classes and have a higher than average discount rate for those that bear the earmarks of being highly speculative and a lower than average discount rate for those of types with assured histories of proven payoffs, we would be doing a decent job, though obviously not the best conceivable job [12: 73].

Bains, Caves, and Margolis seem to espouse a similar harmonistic viewpoint:

> Briefly, however, it would appear that, in order to secure the optimal or best attainable suboptimal allocation of resources to water development (given existing organization and performance in the private sector, including the organization of markets for funds), the appropriate rate of discount should be roughly equal to the *marginal* rate of return in marginal long-term investment in the private sector, and also equal to the marginal rate of time preference of the taxpayers of agency constituents who ultimately finance the bulk of investment in water projects. These two rates tend generally to coincide and to be approximated by the going net rate of interest on private savings invested in real estate [2: 268].

Note that the first and third citations here select rather different private investment classes as relevant and comparable in riskiness to public investment in the water-supply field. Still another related viewpoint is that of Stockfisch, who takes the estimated overall rate of return in the private sector as the comparable rate for military investment decisions [34]. A very important practical consideration here is whether rates are to be computed gross or net of corporate income tax—a point which will be taken up in section VI.

There are a number of other viewpoints worthy of notice, though not conveniently classifiable in terms of the categories above. McKean regards risk as an "intangible," and leaves its resolution to the sphere

of judgment [23: 64]. Dorfman attempted to apply statistical decision theory to the water-resource field, but came up with pessimistic conclusions as to the present applicability of these ideas [6: 144]. He did suggest one modification of traditional risk-adjustment approaches: to choose projects on the basis of minimizing probability of experiencing a specified disaster outcome. Finally, Haveman has advocated an interesting variant of discounting procedures ([13], appendix B). He would add a premium to the riskless rate when discounting benefits but would use a rate lower than the pure rate in discounting costs. The basic idea is that risk-aversion dictates writing *down* the present-value estimate of future benefits but writing *up* the present-value estimate of future costs. Moreover, in view of the greater uncertainty of benefits, the former adjustment should be larger than the latter. However, this discussion does not lead to a specification of how the discount rates are to be determined. Doing so would require a theoretical model of risk. The state-preference analysis of section V, which provides such a model, indicates that it is not actually necessary to separate the two rates—as will be seen below, it is possible in principle to determine a single risky discount rate that will generate the correct present certainty-equivalent value of an investment project. However, as a practical matter Haveman's two-rate proposal would probably be a defensible and workable improvement, in comparison with using a single discount rate that purports to be related to the average riskiness of the separate cost and benefit streams.

SUMMARY OF QUANTITATIVE RECOMMENDATIONS ON DISCOUNT RATES

The discount rates shown in table I represent estimates and proposals by different authors for Government investment decisions. The figures are not fully comparable, since they were made at varying dates in a period of changing conditions in the financial markets. Also, in some cases differing types of Government decisions were under consideration, so that the "comparable" private rates would not be expected to be the same.

Against all these should be kept in mind the recommendation of some authors that the *riskless* rate be used. This is usually interpreted as the government borrowing rate for the appropriate (usually long term) period of investment. This rate has varied in recent years between 5 and 5¾ percent. However, it is clear that the figures here include an adjustment for inflationary expectations; the anticipated *real* riskless rate has probably been rather steady in the neighborhood of 4 percent.[7]

TABLE 1.—DISCOUNT RATE RECOMMENDATIONS—RISKY RATES

Author	Year	Recommendation
Krutilla and Eckstein [19]	1958	5 to 6 percent.
Eckstein [9]	1958	½- to 1-percent risk premium.
Eckstein [8]	1968	8 percent.
Hirshleifer, DeHaven, and Milliman [16]	1960	10 percent.
J. A. Stockfisch [34]	1967	13.5 percent.
Bain, Caves, and Margolis [2]	1966	5 to 6 percent.
Harberger [12]	1968	10.68 percent.

[7] Although the nominal rate on long-term Government bonds has ranged from 5 to 5¾ percent in recent years, the rate of inflation in the economy has grown from 1 to about 3½ percent. This implies that the rate of return on these bonds has fallen from 4 percent to as low as 2 percent at the present time. However, the erosion of *realized* real returns by inflation appears to be systematically underestimated by investors. It seems reasonable to infer, therefore, that the *anticipated* real rate of return that enters into investors' calculations has remained in the neighborhood of 4 percent.

IV. Past and Present Government Practice

The analytical controversy, over whether premiums paid in the market to overcome individuals' risk-aversion represent a social as well as a private cost of investment, finds its counterpart in the practices of Government agencies.[8] The issue was epitomized in the testimony of the Comptroller General:

> One school of thought holds that the rate should be determined by and be equal to the rate paid by the Treasury in borrowing money. A second school of thought holds that the rate should be determined by what is foregone; namely, the return that could have been earned in the private sector of the economy when the decision is made to commit resources to the public sector [31 : 6].

The first position cited by the Comptroller General clearly corresponds to the prescription of discounting by the riskless rate; the second to discounting by a risky one.

It is very worthy of note that some Federal agencies do not discount future benefits (and costs) at all, while others employ only a haphazard method for doing so. A recent study by the Comptroller General (see table 2 below) showed that only 10 of 23 agencies surveyed employed the discounting technique in evaluating agency investment. Of the remaining 13, eight do have plans to employ discounting in the future. Some agencies engage in a form of implicit discounting by imposing an arbitrary cutoff date on future benefits. The Department of Labor, for example, is said to use a 1-year time horizon even though its program benefits are thought to continue for periods of 5 to 25 years. This method is equivalent to a zero discount rate up to the cutoff date, and an infinite discount rate thereafter. Other agencies that simply cumulate costs and benefits regardless of date are in effect employing a zero discount rate throughout.

In an effort to standardize analytical procedure, particularly among water resource agencies, the "Green Book" [35] was produced in 1950 by a committee of several agencies. With regard to the question of risk, the Green Book followed Frank Knight in attempting to distinguish between "predictable risk" and "nonpredictable risk."[9] Examples cited of predictable risk were fires, storms, pests, and diseases. Examples of nonpredictable risk were fluctuations in the level of economic activity, and innovations and technological change. For predictable risk the Green Book recommended what were in effect expected-value adjustments of the cost-benefit stream. For nonpredictable risk, a variety of adjustments were recommended: conservative estimates of net benefits, safety margins in planning, or the inclusion of a risk component in the discount rate. The rather surprising last possibility was advocated because of the impracticability in some cases of arriving at risk-free estimates of benefits.

The quantitative recommendation of the Green Book included the use of 2½-percent discount rate on government costs. (It must be remembered that, at the date of the Green Book, interest rates were far lower than those we have since become accustomed to. And, furthermore, they were then regarded as abnormally high and due for a fall

[8] Attention here will be restricted to agencies of the Federal Government.
[9] In sec. I above, we discarded this distinction as sterile.

toward the more "normal" ranges experienced in the 1930's.) A rate of not less than 4 percent was recommended for private costs and all benefits. This higher rate was justified on the ground that it would:

> . . . approach the rate of return needed to induce private investment and participation. This rate corresponds to the minimum current costs to private borrowers for obtaining funds through mortgage loans secured by real property or other substantial assets. [35 : 23].

The report even contemplated that still higher rates would be used in certain circumstances. Thus, the Green Book actually went a considerable distance toward the risky-rate prescription for discounting—the premise being that the projects under consideration were at least comparable in risk to "mortgage loans secured by real property or other substantial assets."

This advice went unheeded, however. Later revisions were made in the discount rate but the quest was merely for a representative riskless rate. In 1962 an *ad hoc* council appointed by President Kennedy recommended a formula for calculating such a rate. This provided:

> The interest rate to be used in plan formulation and evaluation for discounting future benefits and computing costs * * * shall be based upon the average rate of interest payable by the Treasury (i.e., the *coupon rate*) on interest-bearing marketable securities of the United States outstanding at the end of the fiscal year preceding such computation which, upon original issue, had terms to maturity of 15 years or more. [36 : 12]

In 1968, another revision was sought and the most current proposal is:

> The interest rate to be used in plan formulation and evaluation for discounting future benefits and computing costs, or otherwise converting benefits and costs to a common time basis, shall be based upon the average yield during the preceding fiscal year on interest-bearing marketable securities of the United States which at the time the computation is made, have terms of 15 years or more remaining to maturity: Provided, however, that in no event shall the rate be raised or lowered more than one-quarter percent for any years. [4 : 13]

The intent of the proposed reform is to eliminate the influence of historical coupon rates in favor of the yield data that reflect the current state of the funds markets. The net effect of the proposal was to raise the rate from 3¼ to 4⅝ percent. While this has the salutary effect of bringing the rate more up-to-date, it is clear that the intention remains to discount at a riskless rate.[10]

Many agencies follow this lead and employ the Treasury's borrowing cost (see [33], app. 1). Other Government agencies, however, have used rates that at least implicitly include a risk premium. The leader and prime mover in this area is the Department of Defense, which employs a discount rate of 10 percent. This rate was selected:

> To reflect the amount of time preference for current versus future money sacrifices that the public exhibits in nongovern-

[10] The current rates on money loans do reflect, however, the market's state of inflationary expectation. Since price-level increases of around 1 to 3 percent per year have been experienced in recent years it is reasonable to infer that the market's estimate of the *real* riskless rate of interest must be somewhat lower than the Treasury's borrowing rate. It follows that if future benefits are calculated at today's prices without allowance for inflation, the current Treasury borrowing rate really is too high to be a riskless rate. However, this result seems to be inadvertent, and not a concession in the direction of use of a risky discount rate.

mental transactions. The 10-percent rate is considered to be the most representative point within a range of plausible rates obtained from and considering this public time preference. [33 : 57]

We see that Government practice is highly varied. This is not surprising when it is realized that the establishment of the discount rate is left to administrative decision rather than being the subject of legislative direction. Not only are Government planners given considerable discretion and latitude in the estimation of benefits and costs, they are also left to exercise discretion regarding the applicable rate of discount. The potentialities for optimistic distortion are evident, if only because the decisionmakers within any given agency tend to be enthusiastic about and committed to its own particular goals. This review would seem to indicate the desirability of constraining the now-excessive degrees of freedom of the agencies, in the interests of a more efficient balance of investment within the Federal Government.

TABLE 2.—*Discounting Practices of Federal Agencies*

I. Federal agencies not using discounting in the analysis of individual programs in fiscal year 1969

A. *Agencies that plan to use discounting in future—*

1. Department of Housing and Urban Development
2. Federal Power Commission
3. Peace Corps
4. National Science Foundation
5. National Aeronautics and Space Administration
6. Department of Labor
7. Post Office Department
8. Federal Communications Commission

B. *Agencies that have no stated plans to use discounting or had no comments on their plans—*

1. Interstate Commerce Commission
2. Veterans Administration
3. Department of the Treasury
4. Export-Import Bank of Washington
5. Department of Commerce

II. Agencies Employing Discounting—

1. Tennessee Valley Authority
2. General Services Administration
3. Department of Agriculture
4. Office of Economic Opportunity
5. Department of Transportation (Federal Aviation Administration)
6. Atomic Energy Commission
7. Department of Defense
8. Agency for International Development
9. Department of the Interior
10. Department of Health, Education, and Welfare

NOTE: Survey did not cover agencies involved in water and related land resources programs that come within the purview of the Budget Circular A–76 entitled "Policies for Acquiring Commercial or Industrial Products and Services for Government Use." Source: [33 : 47]

V. Recent Theoretical Developments: The Model of Time-State Preference

In the past decade, theoretical economists have made considerable advances in the direction of an exact theory of risk. While many problems remain, this theoretical development now makes possible a better understanding and pinpointing of the unresolved issues in the traditional literature reviewed in section III above.

The key idea of the recent theoretical development is the picturing of an uncertain future as a set of hypothetical alternative "states of the world" at each date. Although one and only one of these states will in fact occur at each date, decisionmakers now, in the present, must contemplate and allow for all the possibilities if they are to act rationally. In so acting, they will naturally take into account, where relevant, their beliefs as to the probabilities of the several states, the values they place upon larger and smaller income claims, and their time-preferences. The interaction in the market of individual decisions will in effect determine a set of market prices for contingent claims to income at each future time-state. It is as if these contingent claims themselves become the elementary commodities that are traded in markets, although in reality trading occurs in sets of claims packaged into the forms we call *securities*.

These concepts can be elucidated with a simple model. In expounding the nature of the riskless solution, we dealt with the elementary case of choice between consumption now and consumption 1 year from now. The minimal extension of this model to allow for risk is the case in which there is no uncertainty about the present, but at a single future date (1 year from now) just one of two alternative states of the world "a" or "b" (e.g., war or peace, depression or prosperity) must come about. Given the reduction to dollar values, there are in this situation three objects of choice: claims to present consumption, claims to future consumption valid if and only if the state "a" obtains, and claims to future consumption valid if and only if state "b" obtains.

In the riskless case, the market was said to determine the interest rate r_1 as a discounting factor for future claims. But it is possible to express this in another way: we can say that the market determines the relative price ratio between future and present claims. Thus, taking the price of current claims P_0 as unity, the price of future claims P_1 is given by the discount factor: $P_1=1/(1+r_1)$. This permits us to rewrite the present value equation (1) for the certainty case in the instructive form:

$$V_0=-P_0i_0+P_1q_1 \qquad (1')$$

In the simple model of risk dealt with here, the market will determine the price ratios among the three objects of choice. Again taking the price of current claims P_0 as unity, the prices of the future contingent claims P_{1a} and P_{1b} will be determined. As before, a number of conclusions will follow from standard economic theory: there will be a principle of rational consumer behavior, of rational producer behavior, of market equilibrium, and a normative rule for Government investment decision. The extension of the "positive" principles to this risky model is reasonably clear, and they need not be explicitly re-

stated here. But it will be important to specify the *present certainty-equivalent value rule* for Government investment decision as the generalization to risky situations of the present-value rule of the riskless model. This rule is that the Government ought to adopt that set of projects maximizing the net balance V_0 of the market values of present sacrifice and future returns in the several states of the world contemplated:

$$V_0 = -P_0 i_0 + P_{1a} q_{1a} + P_{1b} q_{1b} \tag{3}$$

Or,[11] that each and every project should be adopted for which the incremental present certainty-equivalent value ΔV_0 is positive, where ΔV_0 is given by:

$$\Delta V_0 = -P_0 \Delta i_0 + P_{1a} \Delta q_{1a} + P_{1b} \Delta q_{1b} \tag{4}$$

As in the case of the riskless rule, the present certainty-equivalent value (PCEV) is strictly *objective*. It does not depend at all upon the decisionmaker's personal time-preference (or even probability beliefs), but only upon the net integration of all individuals' preferences and beliefs as expressed in the market in the form of prices for present and future claims.

A numerical illustration may be helpful at this point. Suppose that, with P_0 set at unity, the market-equilibrium prices for contingent claims are $P_{1a}=0.3$ and $P_{1b}=0.5$. Now imagine a government decision-maker considering an investment project involving a current outlay of \$1 and the future receipt of \$3 conditional upon state *a* coming about (and returning zero otherwise). This opportunity may be suggestively symbolized in the form $\{-1,{}^{3}_{0}\}$. Its present certainty-equivalent value ΔV_0 is $-1+3(0.3)+0(0.5)=-0.1$; this project, if adopted, would decrease overall PCEV and so should be rejected. Alternatively, consider the opportunity symbolized by $\{-1,{}^{2}_{2}\}$. Here a dollar outlay in the present leads to \$2 of return regardless of which state obtains 1 year from now (this is a riskless investment). Since $-1+2(0.3)+2(0.5)=+0.6$, the project has positive PCEV and so should be adopted. Note the strictly objective nature of the determination of acceptability of projects.

While productive decisionmakers need not consider subjective preferences or beliefs, as in the riskless case the *consumptive* decisionmakers will take these into account in distributing consumption (within their wealth constraints) over the available objects of choice. Thus, an individual might desire to purchase and hold more claims to future consumption in state *a* because he thinks state *a* is very likely, or because he is already well covered if state *b* comes about, or because he is generally satiated with consumption in the present, and so forth. All these factors will enter into the supplies and demands for the several objects of choice, and thus into the determination of the market prices.

In the certainty model, we saw that the simple relationship $P_1=1/(1+r_1)$ holds between the rate of interest r_1 and the price of future claims P_1. It is possible to generalize this relationship to the risky case by expressing the value of a future *certain* claim as P_1 in:

$$P_1 = P_{1a} + P_{1b} \tag{5}$$

[11] As in the simple PV rule, the two forms here are equivalent given project independence.

That is, a unit of future consumption can be purchased with certainty by combining unit future claims for each state of the world. This leads directly to the definition of a *riskless rate of interest* r_1 *in an uncertain world*, in the form:

$$P_{1a}+P_{1b}=\frac{1}{1+r_1} \tag{6}$$

In terms of the numerical illustration above, with $P_{1a}=0.3$ and $P_{1b}=0.5$, it is evident that the price of a future certain claim $P_1=0.8$, so that the riskless rate of interest r_1 is 25 percent.

We can now employ this theoretical development to reconsider the key issues debated in the literature review in section III. The first and most central issue is: Is risk (in the sense of variability) a social cost, or a merely private cost that government ought to ignore? The latter view rests, of course, upon the "pooling" argument. Since government is in a position to undertake many independent projects, it can treat the mathematical expectation of yield as in effect certain. Operationally, proponents of this position would take the mathematical expectations of the benefits and costs at each date, and then discount with the riskless rate of interest r_1. In contrast, the view that private risk aversion is reflective of a real social cost leads operationally to discounting the mathematical expectations of net benefits by a "risky interest rate" r_1^* suitable for the characteristic pattern of risk of the investment in question. Since we know that the present certainty-equivalent value rule gives correct results, the question may be posed in terms of the consistency of each of these operational procedures with that rule.

The difference between the recommended procedures can be illustrated by the respective evaluations of the project symbolized above by $\{-1,{}^{3}_{0}\}$. In this numerical example, as before, we let $P_{1a}=0.3$ and $P_{1b}=0.5$. But let us now assume in addition that the two states can be considered equiprobable; thus, $\pi_{1a}=\pi_{1b}=\frac{1}{2}$. Then, the mathematical expectation of the future return is 1.5. With a *riskless* interest rate of 25 percent, the discounting equation takes the form $-1+\frac{1.5}{1.25}=+0.2$. But we know that this is a project whose PCEV is negative ($=-0.1$) and which should *not* be adopted. Thus, the example shows that the recommendation to use the riskles rate in discounting the mathematical expectation of future returns cannot be correct in general. For the contrasting recommendation, the key problem is the question of how to determine the "suitable" risky rate r_1^*. Waiving for the moment the source of this figure, it is possible to verify numerically that 66⅔ percent is the correct discount rate to employ. That is, $-1+\frac{1.5}{1.667}=-0.1$, which equals the PCEV.

More formally, the procedure recommended by those supporting the pooling argument—discounting the mathematical expectation of future return by the riskless rate—would indicate adoption of an incremental project when the $\Delta V_0'$ is positive, $\Delta V_0'$ being defined in terms of the state probabilities π_{1a} and π_{1b} as follows:

$$\Delta V_0'=-\Delta i_0+\frac{\pi_{1a}\Delta q_{1a}+\pi_{1b}\Delta q_{1b}}{1+r_1} \tag{7}$$

It is intuitively evident that this procedure is correct—i.e., is logically equivalent to the PCEV rule—*if and only if the probabilities assigned to states are proportional to the prices of the state claims.*[12] In terms of the numerical example, if the state probabilities were 3/8 and 5/8, respectively, the mathematical expectation of future return would be $3(3/8)+0(5/8)=9/8$, so that $\Delta V_0'=-1+\frac{9/8}{1.25}=-0.1$, the correct result.

Can we assume that the prices of state-claims must be proportionate to the state-probabilities? No; because for one thing the consumers might simply have a disproportionately intense preference for consumption in one state rather than in another. A still more important, because systematically predictable, consideration is the following: Prices of future state-claims will depend not only upon relative probabilities but on the relative *scarcities* of income in the several states. If individuals are predominantly risk avoiders, claims to incremental consumption in a given state will be more highly valued, the probability being the same, if that state is associated with low overall income.[13] And indeed this prediction is borne out empirically by the relatively high realized percentage yield (implying relative low prior valuation of income) earned by procyclical securities like industrial equities in comparison with stable or anticyclical securities like government bonds.

Returning to the procedure of discounting mathematical expectation of future return by a *risky* interest rate, the correct r_1^* is given by the condition:[14]

$$1+r_1^*=\frac{\pi_{1a}\Delta q_{1a}+\pi_{1b}\Delta q_{1b}}{P_{1a}\Delta q_{1a}+P_{1b}\Delta q_{1b}} \tag{8}$$

It is important to note that the suitable risky rate is dependent only upon the *proportional* state-pattern of future returns. Thus, the projects symbolized by $\{-1,^{3}_{0}\}$, $\{-1,^{2}_{0}\}$, $\{-3,^{5}_{0}\}$ would all be discounted at the same rate; i.e., they constitute a "risk class" in the terminology of Modigliani and Miller [24: 268]. Note also that the riskless interest rate can be regarded as a special case of the more general logical category of risky interest rates; specifically, as the rate applicable to the certainty "risk class" represented by projects like $\{-1,^{2}_{2}\}$ or $\{-5,^{6}_{6}\}$.

---

[12] If $\frac{\pi_{1a}}{P_{1a}}=\frac{\pi_{1b}}{P_{1b}}$, then their common ratio equals $\frac{\pi_{1a}+\pi_{1b}}{P_{1a}+P_{1b}}=\frac{1}{P_1}$. Then $V_0'=-\Delta i_0+\frac{\Delta q_{1a}(P_{1a}/P_1)+\Delta q_{1b}(P_{1b}/P_1)}{1+r_1}$. This reduces immediately to the PCEV, ΔV_0, upon making the substitution $P_1=1/(1+r_1)$.

[13] This is sometimes described as the condition of "diminishing marginal utility of income."

[14] This condition leads to the r_1^* yielding the correct PCEV in

$$\Delta V_0=\Delta i_0+\frac{\pi_{1a}\Delta q_{1a}+\pi_{1b}\Delta q_{1b}}{1+r_1^*}.$$

This equation is the discounting version of equation (4).

We may now begin to consider some of the problems of practical application by asking the question: If the suitable risky discount rate is determined by consistency of results with those yielded by the PCEV rule, why go through this roundabout procedure? Why not simply evaluate all projects on the basis of incremental ΔV_0, calculated directly as in equation (4) from the outlay Δi_0, receipts Δq_{1a} and Δq_{1b}, and objective market prices P_{1a} and P_{1b}? Here it is necessary to reveal the skeleton in the closet! In actuality, the number of imaginable states even at a single date is enormously great. There do not exist in fact a complete set of securities in the form of elementary state-claims whose prices would directly provide the needed data for the use of the PCEV rule directly. Actual securities—e.g., stocks and bonds—represent more or less complex packages of claims to income distributed over many different states and dates. It may sometimes be possible to draw inferences from the pattern of prices for securities about the implicit prices for time-state claims, but it seems unlikely that all the necessary prices could be so inferred.

This does not mean, however, that we are completely helpless in practice. Here is where the concept of a "risk class" comes to our aid. For, it may well be that a proposed government project can reasonably be placed in the same risk class as a private project. It is then only necessary to estimate what risky discount rate r_1^* the market imposes on expected returns from such private projects. Concretely, the procedure would work as follows: Suppose we are evaluating a public power project, and there seems to be no reason to assume that the project is not in the same risk class as private power projects in general. Let us suppose that the latter are typically financed half and half by stocks and bonds, and that market prices for securities are such that the mathematical expectation of yield on the stocks is 8 percent and the bonds 5 percent. Then the suitable r_1^* for use in discounting expected returns of the comparable government project would be 6.5 percent.[15] (Actually, it is not strictly necessary to know mathematical expectations of returns; provided that the degree of *bias* in the private and public estimates of returns were the same, the implied private discount rate could still be employed.)

Low discount rates tend to make investment projects look good; high interest rates tend to make them look bad. In this connection, it is of some importance to appreciate the fact that the "risky" interest rate, while usually higher, could conceivably turn out to be lower than the "riskless" interest rate. Consider our earlier numerical example, with state probabilities still $\pi_{1a}=\pi_{1b}=\frac{1}{2}$, prices $P_{1a}=0.3$ and $P_{1b}=0.5$, but where we now have a project whose state pattern of returns is the reverse of that considered earlier; to wit $\{-1,{}^{0}_{3}\}$. Direct calculation from equation (4) shows ΔV_0 to be $+0.5$, while equation (8) shows the suitable "risky" rate r_1^* to be zero percent! This figure is *lower* than the riskless rate r_1 of 25 percent. The explanation for this possibly surprising result is that the riskless rate is calculated in terms of a state-pattern of returns yielding *equally in the two states*, while in the present example we have a situation in which the returns fall disproportionately (entirely, in fact) in the relatively more highly

[15] This discussion omits the problem of dealing with the corporate income tax, a complication that may be of great practical importance. Its role will be reconsidered in sec. VI below.

valued state. Proponents of government investment programs often argue that government should ordinarily employ a lower discount rate than does private industry in evaluating investment projects. Such a recommendation would be valid in this model if it were the case that returns from government projects tend to fall disproportionately in such highly valued states, in comparison with private industry. Thus, the "typical" private investment would, on this argument, not be the "comparable" one (in terms of risk class) from which the correct risky rate r_1^* is to be calculated.

Even among those who support the idea that the market's reflection of private risk aversion represents a real social cost, there is a school of thought opposed to *adjustment* for risk through the discount rate. One argument in this connection says that risk need not be a simple compounding function of time, and so no single overall interest rate adjustment would be suitable. The argument is correct, in that in principle r_2^*—the discount rate appropriate for moving claims dated 2 years from now 1 year closer in time—need not be the same as r . But, of course, there is nothing in the theory developed here that precludes the use of a series of different rates for discounting returns through time; we dealt only with r_1^* above merely because of the simplicity of the expository model used here. Furthermore, it will always be possible to find a single time-averaged discount rate R_T^* that will give the same results as calculating with a series of 1-year discount rates $r_1^*, r_2^*, \ldots, r_T^*$.

The alternative to use of the risky interest rate for discounting future uncertain returns would be to attempt to determine, at each date, the *certainty equivalent* of the net return at that date. If this could be done, the riskless interest rate could then be employed to discount these future certainty equivalents into a present certainty-equivalent value. However, the calculation of these future certainty-equivalent values seems to require direct knowledge of consumers' utility functions, and so is not a feasible procedure.

VI. Unresolved Issues and Directions of Future Research

It may seem surprising that the "pooling" argument fails so completely, in view of the plausible arguments that can be made in its favor. One can, in fact, construct a model in which the pooling argument does make more sense. The key is to distinguish between private and social states. To take the most extreme case, suppose that on the social level the amount of income at the single future date is constant; the only thing that may vary is its distribution over individuals. For concreteness, suppose that the individuals can be grouped into equally numerous "even" and "odd" categories. Imagine that each individual has an investment opportunity of the form $\{-1, {}^{3}_{0}\}$, and define state a as the situation where the evens obtain 3 and the odds obtain 0, while state b is the reverse situation. Then, we can see, private risk aversion may make every individual disinclined to undertake his personal investment project—whereas, on the social level, there is no risk at all.

The crucial point of this example, and the hidden assumption in the usual pooling argument, is that *the two classes of individuals cannot trade* state-a for state-b claims. For, if they traded at par, say, they could all convert their respective investment opportunities into the form $\{-1, {}^{1}_{1}; {}^{5}_{5}\}$ which would be *privately as well as socially* riskless.

Given risk aversion, everyone would benefit from such an exchange. Why then do they not make the exchange? The same argument can be generalized to a situation where there are multiple social states (some degree of social risk exists), but where the investment opportunities available to individuals have private state-patterns of returns that are all uncorrelated with these social states. Again trading among individuals should eliminate all private risk that is not reflective of social risk.

The pooling argument rests ultimately for support upon market imperfections—which hinder the trading that would otherwise tend to eliminate private risks (unbalanced state-patterns of returns) not reflective of social risks. Now the number of conceivable time-state contingencies is enormous, far greater than the number of distinct securities traded. This disparity is not surprising, given that there are real costs of providing markets. Thus, imperfect marketability of time-state claims is a fact, and so there is a degree of validity in the pooling argument. On the other hand, imperfect ability of Government to distribute time-state claims in accordance with consumers' marginal preferences is surely an equally signficant fact.[16] As a practical matter, the present authors believe that the use of the "comparable" risky rate of discount from the private sector is the best general guide—though the justification falls short of being airtight.

As a final note, let us consider the role of the corporate income tax.* The existence of this tax makes an enormous difference in the computation of the net returns from investment, and therefore in the calculation of risky interest rates for use in evaluation of Government projects comparable to those in the private sector. In the example above of a project to be equally financed with bonds of an expected yield of 5 percent and stocks of an expected yield of 8 percent, in the absence of tax the required real yield for the project to be profitable would be 6.5 percent. But with corporate income tax at a 50 percent rate, the real yield would have to be 10.5 percent.[17] Thus, the tax places at an enormous disadvantage the risky projects that have to be financed in whole or in part from equity investment.[18] It is therefore, just the reverse of the policy that might be suggested by the degree of validity in the pooling argument: namely, that risky investment ought to benefit from a subsidy. *The effect of the corporate income tax is very likely greater than imperfection of capital markets in deterring private risky undertakings.*

Now it may be that the corporate income tax in its present form reflects some kind of "social judgment" that private individuals are *excessively* inclined to undertake risky investment. If that is the case, then the effective required market yield (10.5 percent in the example) is the one that represents the suitable risky interest rate for use in evaluating a comparable government project. That is, the "social judgment" presumably says that the risk-class of projects that individuals would be inclined to undertake at a 6.5 percent expected yield ought

[16] This point was made in private correspondence by Peter A. Diamond.

[17] The real yield would be the average of 16 percent gross of tax (8 percent net) yield to pay on equity, and 5 percent yield to pay on bonds.

[18] On the other hand, it should be mentioned that the *personal* income tax has a partially counterbalancing opposite effect. The returns on risky securities can often be taken in the form of capital gains that benefit from a reduced tax rate.

*Further discussion of this issue is found in the paper by Baumol in this volume.

not to be undertaken unless they yield 10.5 percent—and this would hold true for Government projects as well. On the other hand, if the bias against risky private investments is to be taken as a mistaken or merely accidental result of policies adopted for other ends, presumably some figure between the 6.5 percent and the 10.5 percent is to be used in discounting Government investments (as a "second best" to the more desirable policy of modifying the tax law). The precise balance would have to depend upon the degree to which the public project under consideration absorbed funds that would be invested otherwise in higher-yielding private projects. For, here the "market segmentation" hypothesis is indeed validated. The Government tax places a wedge between the real productive rate of return on investment and the after-tax rate in terms of which investors make their time-preference decisions.

The main contentions of the last two sections can be briefly summarized. The model of time-state preference leads to the present certainty-equivalent value (PCEV) criterion as the generalization of the present-value (PV) criterion applicable in a world of certainty. Within the framework of this model, the "pooling argument" is incorrect. That is, discounting the mathematical expectation of future return by the riskless rate of interest r_1 may result in the adoption of projects of negative PCEV (or failure to adopt projects of positive PCEV). For discounting mathematical expectation of return, the model shows how a "risky rate" r_1^* appropriated for the risk class of the proposed investment should be calculated. In practice we do not actually have enough data to *calculate* this r_1^*. However, it can usefully be approximated by *observing* the r_1^* implicit in the evaluation of private projects in the same or closely comparable risk class. Any validity that the pooling argument has must rest upon market segmentation that prevents trading of time-state claims. Such market imperfections do indeed exist. But it seems unlikely that government has the ability to distribute time-state claims more in accordance with consumer desires than even these admittedly imperfect capital markets. Indeed, one form of government intervention—the corporate income tax—is an enormous deterrent to the risky private investment whose alleged insufficiency is what the "pooling argument" seeks to correct.

BIBLIOGRAPHY

1. K. J. Arrow, "Discounting and Public Investment Criteria," in A. V. Kneese & S. C. Smith, eds., *Water Resources Research* (Johns Hopkins Press, 1966).
2. J. S. Bain, R. E. Caves, & J. Margolis, *Northern California's Water Industry* (Johns Hopkins Press, 1966).
3. W. J. Baumol, "On the Social Rate of Discount," *American Economic Review*, v. 58 (Sept., 1968).
4. H. P. Caulfield, Jr., Statement of Henry P. Caulifield, Jr, Executive Director, Water Resources Council, before Subcommittee on Economy in Government, Joint Economic Committee, *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis*, 90th Congress, 2nd Session.
5. P. A. Diamond, "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," *American Economic Review*, v. 57 (Sept., 1967).
6. R. Dorfman, "Basic Economic and Technologic Concepts: A General Statement," in A. Maass, M. Hufschmidt, R. Dorfman, H. A. Thomas, Jr., S. A. Marglin & G. M. Fair, *Design of Water Resource Systems* (Harvard U. Press, 1962).
7. O. Eckstein, "Reply by Mr. Eckstein" (to Comments of J. Hirshleifer), in National Bureau of Economic Research, *Public Finances: Needs, Sources, and Utilization* (Princeton U. Press, 1961).
8. O. Eckstein, Statement of Otto Eckstein before Subcommittee on Economy in Government, Joint Economic Committee, *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis*, 90th Congress, 2nd Session.
9. O. Eckstein, *Water Resource Development* (Harvard U. Press, 1958).
10. M. S. Feldstein, "The Social Time Preference Discount Rate in Cost Benefit Analysis," *Economic Journal*, v. 74 (June, 1964).
11. M. Friedman & L. J. Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, v. 56 (Aug., 1948).
12. A. Harberger, Remarks before Subcommittee on Economy in Government, Joint Economic Committee, *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis*, 90th Congress, 2nd Session.
13. R. H. Haveman, *Water Resource Investment and the Public Interest* (Vanderbilt U. Press, 1965).
14. J. Hirshleifer, "Investment Decision Under Uncertainty: Applications of the State-Preference Approach," *Quarterly Journal of Economics*, v. 80 (May, 1966).
15. J. Hirshleifer, "Preference Sociale a l'Egard du Temps," *Recherches Economiques De Louvain*, v. 34 (1968).
16. J. Hirshleifer, J. C. DeHaven, & J. W. Milliman *Water Supply: Economics, Technology, and Policy* (U. of Chicago Press, 1960).
17. F. S. Hoffman, Statement before Subcommittee on Economy in Government, Joint Economic Committee, *Economic Analysis of Public Investment Decisions: Interest Rate Policy and Discounting Analysis*, 90th Congress, 2nd Session.
18. F. H. Knight, *Risk, Uncertainty, and Profit* (Harper & Row, 1965).
19. J. V. Krutilla & O. Eckstein, *Multipurpose River Development* (Johns Hopkins Press, 1958).
20. S. A. Marglin, "The Social Rate of Discount and the Optimal Rate of Investment," *Quarterly Journal of Economics*, v. 78 (May, 1964).
21. H. Markowitz, "The Utility of Wealth," *Journal of Political Economy*, v. 60 (April, 1952).
22. A. Marshall, *Principles of Economics*, 8th ed. (London: Macmillan, 1964).
23. R. N. McKean, *Efficiency in Government Through Systems Analysis* (Wiley, 1958).

24. F. Modigliani & M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review*, v. 48 (June 1958).

25. A. C. Pigou, *The Economics of Welfare*, 4th ed. (New York: Macmillan, 1960).

26. A. R. Prest & R. Turvey, "Cost-Benefit Analysis: A Survey," *Economic Journal*, v. 75 (Dec., 1965).

27. P. A. Samuelson, "Principles of Efficiency—Discussion," *American Economic Review*, v. 54 (Papers & Proceedings, May 1964).

28. L. J. Savage, *The Foundations of Statistics* (Wiley, 1965).

29. W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, v. 19 (Sept., 1964).

30. A. Smith, *The Wealth of Nations* (Modern Library reprint of Cannan edition).

31. E. B. Staats, Statement of Hon. Elmer B. Staats, Comptroller General of the United States, Joint Economic Committee, Subcommittee on Economy in Government, *Interest Rate Guidelines for Federal Decisionmaking*, 90th Congress, 2nd Session (1968).

32. W. Vickrey, "Principles of Efficiency—Discussion," *American Economic Review*, v. 54 (Papers & Proceedings, May 1964).

33. U.S. Congress, Joint Economic Committee, Subcommittee on Economy in Government, *Interest Rate Guidelines for Federal Decisionmaking*, 90th Congress, 2nd Session (1968).

34. J. A. Stockfisch. Prepared Statement of Jacob A. Stockfisch before Subcommittee on Economy in Government, Joint Economic Committee, *The Planning-Programing-Budgeting System: Progress and Potentials*, 90th Congress, 1st Session.

35. U.S. Congress, Subcommittee on Benefits and Costs of the Federal Inter-Agency River Basin Committee, *Proposed Practices for Economic Analysis of River Basin Projects*: Report to the Federal Inter-Agency River Basin Committee (Washington, D.C., Government Printing Office, May 1950).

36. U.S. Senate Document No. 97, *Policies, Standards Evaluation and Review of Plans for Use and Development of Water and Related Land Resources*, 87th Congress, 2nd Session.

Section B

Shadow Prices for Unpriced or Imperfectly Priced Inputs and Outputs

SHADOW PRICES FOR INCORRECT OR NONEXISTENT MARKET VALUES

BY JULIUS MARGOLIS

Julius Margolis is Professor of Economics and Director of the institute of Public Policy Analysis at Stanford University.

A major difficulty in evaluating the benefits and costs of a public undertaking is the absence of market values for most outputs and many inputs. Indeed, one of the reasons why Government undertakes certain activities is because there is no market to enable private sector production and distribution to take place. As Professor Margolis points out in this paper, it is the job of the government analyst responsible for evaluating the economic worth of an activity to develop values for outputs and inputs which coincide with those which would have been generated by a market had one existed.

The process of forming these shadow values is made difficult not only because of market limitations, but also because there is no widespread agreement on the basis for generating social values. Should one try to aggregate the preferences of people in the society or should one attempt to rely on the preferences of program planners or administrators? Professor Margolis argues that it is the willingness of people to pay for government outputs which signifies their social value. He discusses the techniques for estimating this willingness to pay and shows how the existence of legal and institutional constraints can lead to errors in the estimation of shadow prices. Professor Margolis concludes with the reminder that "shadow prices are useful pieces of information, but unless the decisionmakers benefit by acting on the basis of those prices, little is gained."

Introduction

The State of Israel retaliated for the machine gunning of one plane and the death of a passenger by sending troops across a border and destroying 13 planes. The United Nations judged that the exchange was not appropriate and condemned Israel. Other retaliatory raids have not been condemned—was there an implied judgment that the "price was fair?" Many persons would object in principle to retaliation, a forced exchange of losses, and though most of us would be prepared to state whether a specific retaliatory exchange was reasonable or not we would be very hard-pressed to offer a "scientific" defense of our opinion. The oratory at the Security Council may seem far removed from the more prosaic questions of public expenditures analysis, but agreement about shadow prices is the essence of both problems, that is, the evaluation of exchanges which are not carried out under optimal market conditions. The problem of estimation of shadow prices exists wherever the market is imperfect or nonexistent and of course government services and regulation are instances *par excellence.*

In economic analysis the problem of evaluation of goods or activities is solved for the bulk of cases by market exchanges, where money enters on one side of the exchange. Whenever money is used, we can form a ratio of the amount of the good to the amount of money; the ratio is the price and this can be quoted for all goods and activities.

However there are many cases where exchanges occur without money passing hands; where exchanges occur but where they are not freely entered into; where exchanges are so constrained by institutional rules that it would be dubious to infer that the terms were satisfactory; and where imperfections in the conditions of the exchange would lead us to conclude that the price ratios do not reflect appropriate social judgments about values. Each of these cases gives rise to deficiencies in the use of existing price data as the basis of evaluation of inputs or outputs. The enumeration and refined analysis of market imperfections has reached a high level; unfortunately the analysis of how to replace market numbers is still primitive.

Previous papers have gone into detail concerning the imperfections of the competitive market which both justify public intervention, and at the same time, make prices inappropriate as measures of social value.* I shall focus on the analysis of how governments grope towards assigning price-like values, or shadow prices to their inputs and outputs. But before dealing directly with this issue there are several general background points about the market which should be discussed. The basic question asked by the analyst when he searches for shadow price is: what would the users of the public output be willing to pay. The analyst tries to simulate a perfect and competitive market for the public output, estimate the price which would have resulted, and accept this as the shadow price. Unfortunately, even if the analyst succeeds in finding this price it is not clear that the Government would accept it as sufficient to establish a value. Markets and quasi-markets are one form of social interaction; it is not obvious that other forms of social interaction should be interpreted so as to be consistent with the logic of market exchanges.

I. Market Limitations

A simulated market is a useful approach so long as the market is the process by which all, or at least the overwhelming number of goods are exchanged. Economics has been subject to attack for its attempts to subject all social activity to the measuring rod of money; though there is much merit to this criticism, fortunately, the economist has persisted in extending his calculus. An aggregate index of the state of social welfare is still beyond our ingenuity but the national income figures which were created to estimate the level of welfare have proven invaluable for a wide variety of purposes and efforts to extend the concepts to social accounts are likely to bear fruit. But more important for our analysis is the huge area of valued social activity contained in the market sector of the economy and which is *not* directly priced. The costs of goods produced do not necessarily measure the value of resources which could have produced other benefits and the prices paid for goods need not be an appropriate measure of final products received.

Consider the case of production of a marketable commodity. Productivity indices are highly variable among nations; within nations, the range of variation among industries or plants is very great. Certainly, some of this variation can be attributed to variable amounts of consumption which occur in the sphere of production. In some cases,

* See the papers in Part I of this volume.

Government controls are exercised to encourage this consumption, but in other instances more mysterious processes are at work. It would not be surprising if the Government rejected the market price as not properly reflecting these other consumption aspects.

One of the most dramatic expansions of consumption in our society has been the reduction in the workday, workyear, and worklife. If leisure could be considered a commodity, it is likely that its volume has increased more than any other commodity. "As a rough measure of past growth in free time, the employed worker has about 1,200 hours per year more nonworking time than his 1890 counterpart * * *. In addition to a shortened workyear, nonworking years have grown by about nine for a male at birth, with present life and worklife expectations." [1] Over the last half-century, about one-third of the gains from productivity have been taken by increased leisure—a nonmarketed consumption good. The point of interest to us is that our society has chosen to take a very large percentage of its potential income in the form of reduced work and thereby reduced money income but increased leisure. It should be pointed out that this reduction has taken place by market choices and through Government actions. Numerous legislative actions have limited work at every stage in the production process. It would not be surprising to find comparable losses in productivity elsewhere, which might also be judged as consumption.

I would hazard that a very large part of "inefficiency" in production is attributable to consumption. We know too little about inefficiency but certainly the range of difference in productivity among firms, industries and nations is huge, even in identical facilities. Productivity gains in applying well-known techniques are often in the range of 50 percent and great efforts are often necessary to initiate them.[2]

Why is there a willingness to accept lower productivity? Is it too unreasonable to say that a large part of inefficiency are potential losses which are really acts of consumption, and the resistance to reduce inefficiency is the unwillingness to surrender consumption.

A willingness to accept lower productivity can be viewed in another perspective—a reluctance to change. It should not be surprising that men resist change in work patterns, especially as they grow older. New methods involve learning, an investment. What had been previously the consumption of potential income now becomes an actual investment and with advanced age, the expected value of increased income falls, so that investment becomes less desirable. The adaptation to a new process involves not only setup costs in new equipment and losses of production but also sacrifices in search for information and learning new habits at the expense of old comfortable routines. The organizational changes are often the most difficult to make.

I do not want to belabor the social role of "inefficient" production processes. Not only could it be interpreted as extra-market consumption, but I believe it casts some understanding about the production

[1] Juanita M. Kreps, *Lifetime Allocation of Work and Leisure*, Research Report No. 22, Office of Research and Statistics, Social Security Administration, U.S. Department of HEW, 1968, p. 36.

[2] For a survey and interpretation of the many studies, see Harvey Leibenstein, "Allocative Efficiency vs. X-Efficiency," *American Economic Review*, June 1966.

of public goods. For instance risk aversion is often stated as a common characteristic of individual decisionmakers and sometimes it is argued that it is sufficient to explain why governments must act in some circumstance.* The adoption of this position may mean that the government does not accept the individual's evaluations as expressed in market prices as binding; the rejection of this price has led to a lengthy and inconclusive debate. The sorting out of utility payoffs in production and investment has proven to be extremely difficult.

A second source of "inefficiency" which has plagued the analyst of the public sector is the immobility of resources. Inputs do move in response to income differentials, but it is clear that there are strong resistances to move, and these resistances grow with age of residents. Individuals have poor information about alternatives and the full range of social activities in the site of a new residence. A move means a necessity of building up their private social capital in a new place instead of consuming the private social capital already accumulated for the old place. Under these conditions, individual resistance to movement is not surprising.

It is equally not surprising that governments in their provision of public services accept this immobility. Many governments assume regional objectives as significant in shaping the public services.** Governments are often not persuaded to abandon a depressed district by demonstrations that the costs of moving a population from a declining region may be less than the extra costs of providing services for depressed and less dense regions. On the contrary, the government may supplement the public services with efforts to encourage the expansion of economic activity to support the inefficiently located public services. The government may be responding to extra-market income and consumption.

There is nothing novel about the preceding remarks; the existence of nonpecuniary advantages and disadvantages is well-known. Unfortunately knowledge of their existence has not led to an appraisal of its importance. I hazard that it may be very important especially in the sectors where the government operates.

In the preceding illustrations we have stressed private evaluations of market activity which are not readily captured in the market. For these reasons, governments may not respect the market process and values. It is also true that there may be social valuations of the market process which vary drastically from the economists interpretation of prices. These social evaluations may lead to a rejection of market guides. For instance the economist's interpretation of market activity leads him to distinguish carefully between efficiency and distribution effects.*** If an entrepreneur discovers a better way to produce a commodity, it may lead to a price reduction. His profits may increase; the profits of his rivals may fall; the consumers will gain. Where we evaluate the desirability of the innovation we just look at the gain of the innovator; the losses of other firms and the

*Further discussion of this issue is found in the paper by Zeckhauser in this volume.

**Further discussion of this issue is found in the paper by McGuire in this volume.

***Further discussion of this issue is found in the papers by Weisbrod, Bonnen, and Freeman in this volume.

gains to the consumer are considered to be offsetting and merely transfers. A well-organized competitive economy is constantly generating these transfers. Governments do not share the indifference of the economist toward these distributional effects.

Compensation is granted in many cases where the losses or damages would be considered transfers. If these were paid by the government as part of a project the benefit-cost calculus would include them in the costs. And yet it may be true that the project might never be authorized unless compensation were paid. It could be said that these payments are necessary to have the project be adopted though they do not reflect the value of any resources used for the project. Those who receive losses may have more political influence than the gainers and thereby be able to affect the political decision, or there may be a social judgment that the losses were incurred because of the government's decision, and it would be improper to allow any sector to have undue losses.*

Government policies or programs are often massive in nature. Their effects are dispersed throughout the economy but they have major impacts in very restricted areas. A new bridge may create extreme losses along old traffic arteries; a zoning shift may create millionaires and shatter dreams. "Inequities," are charged; lengthy delays result; designs are altered; the distributional effects are relevant for policy but they will also affect the resource cost of getting the public service supplied. Further the certainty or uncertainty of compensation as a consequence of a public act will affect how the private sector behaves.

The thrust of these remarks is to cast some doubt on the standard interpretation of the price data available in the market. There are many situations, often characterized as transfers, where government has accepted transfers as costs and thereby has been criticized by economists, but where it may be that the government has been responding to distributional effects. Possibly, a more explicit accounting of these distributional effects might lead to a more effectively designed public program. Shifts in prices always have distributional consequences and we may err in too readily refusing to consider these "nonefficiency" aspects.**

Finally, the market and the data it generates is limited in that it relies on a model of individuals as consumers, whereas in reality the purchasing decisionmakers or the job choosing decisionmakers are one or both parents. Children ride in the car bought by the parent. If the parent chooses to balance his journey to work against garden space, the child does not face the same tradeoffs, he can only enjoy or suffer the consequence of the parental decision. In principle the parent considers the family welfare in making his decision. However, we need only look at the divorce rates to make it apparent that harmony in the family is far from universal. Many public services are directed toward the special problems of those who are disenfranchised in the private market. However, market prices reflect the values of the private decisionmakers in the families, not the social units for whom they act. Would mass transportation or neighborhood play grounds be more numerous if children could spend their "equitable" share?

*Further discussion of this issue is found in the paper by Dorfman & Jacoby in this volume.

**Further discussion of this issue is found in the paper by Feldman in vol. 3 of this collection.

My assignment is not the limitations of the market and therefore I will not extend these remarks. I believe it necessary to introduce the points about these limitations because my subsequent comments about shadow prices accept much of the economists views about the value content of market processes and prices. From here on, I will in large measure presume that if the market were not beset by technical imperfections as externalities, indivisibilities and information costs, relative prices would equal relative costs and therefore relative prices would reasonably reflect relative social values.

II. Shadow Prices: Whose Shadow?

Shadow prices are computed to reflect social values; the estimation procedures assume that social welfare is derived by aggregating individual valuations. In practise, economists have accepted the task of generating support for this view of social welfare. Consistent with this position has been the rejection of the legitimacy of an active role of the political process or administrative structure as formulators of the public interest. The result which could be anticipated, is a tension between the economic analyst, with his view of the public interest, and the political and administrative decisionmakers who do not share the economists' view of the public interest.* Those who accept the authority of administrative officials or political leaders have quite different views of appropriate objective functions and what numbers should be used to evaluate outputs and inputs.

The conflict between the two approaches can be crudely phrased by asking: Who is the client, who is the employer of the economist? The typical answer of the economist is that he is true to the principles of serving the public interest as defined by the profession in their scientific journals i.e. the aggregation of individual preferences. He selects his models and criteria so as to maximize the professional view of the public interest; the economist has selected a client who is neither an employer or decisionmaker. This is a noble perspective, unfortunately the purchasers of the economist's information and advice) the administrative officials or political leaders) are neither persuaded by the economist's insight nor do their incentives impel them to accept the perspective of the national interest as formulated by the economist. For instance, a municipal official, concerned with urban renewal, may assign a benefit to a project if there are net gains to his city while offsetting losses in adjoining cities are of "purely academic" interest. Or, a national transportation agency may assign benefits based upon improved traffic flows, but the agency is not expected to consider the losses elsewhere, for instance, the increase in noise or air pollution. Payoffs to decisionmakers, administrative or political, are not based upon what happens to an index of national welfare, even if it could be constructed, and therefore it is not surprising that evaluatory measures developed in response to an agency's needs are partial and sometimes inconsistent with social welfare.

It may be presumptuous for the economist to insist on the primacy of his imputed prices against the views of the legitimately constituted

*Further discussion of this issue is found in the paper by Steiner in this volume.

authorities who employ him. However disrespect for authority was not introduced by our rebellious students; professional bodies have long lobbied to influence policy on the basis of their special competency. In the next section we shall assume the economist's vision of the public interest, despite our reservation that there is an alternative scheme by which to evaluate public outputs and inputs. There are social processes by which public decisions are guided. We have alluded to some, e.g. political bargaining, bureaucratic myopia, professional self-interest; it is likely that outcomes of each process considered independently would be far from socially optimal, but this does not mean that the total set of processes are not optimal. For instance, when we analyze the market system we recognize that a unit may not be at its best position though the economy may be operating optimally. We realize that a firm tries to maximize its profits and that its behavior is socially optimal if the economic system is so organized that it behaves competitively. Is it possible that a similar structure exists among different decisionmaking systems? Are there conditions of social equilibrium which are related to economic or political equilibrium just as general economic equilibrium is related to the equilibrium of the firm? I am far from convinced that a concept of social equilibrium analogous to economic equilibrium is useful but certainly economic criteria are partial and it would be wise to keep an open mind about more general formulations. Meanwhile we see efforts to extend economic equilibrium concepts to political behavior, but as I said, hereafter I will accept the economist's limited view.

III. Shadow Prices and Input Costs

The correction of market prices, or the imputation of prices, is done for both inputs and outputs. Most of the controversies deal with outputs though the same conceptual difficulties apply to both cases. In principle inputs are valued in terms of what they could have produced elsewhere and, therefore, we are immediately driven to the valuation of outputs. In practice, it is usually assumed that the market price of the inputs reflects their alternative values and therefore they could be used to measure the value of the inputs. Unfortunately there are a few dramatic cases where this is incorrect and others where the correctness is disputed.

The most commonly advocated adjustment of an input price is for unemployed labor.* Clearly, if labor had no alternative use then it would be incorrect to assign a price to it. This adjustment would make a project less costly and it would affect the design of the project. Of course, it would not be appropriate to assume the labor flow of a project over time would have been unemployed, but the problem of timing often holds for even the fairly shortrun. The time lag between analysis and the initiation of a program may extend for years and therefore current unemployment may not be an appropriate assumption. If one could not assume unemployment for the period of actual use of the resource then the full thrust of this adjustment, the substitution of unemployed labor for other inputs, would be lost.

*Further discussion of this issue is found in the paper by Haveman in this volume.

A second input which may often be incorrectly priced, and too often does not bear an adjustment, is land.* A program on public land is usually assessed the cost of the land to the agency rather than the opportunity costs of the land. Therefore a project on public land is "less costly" and would receive preference. The value of public land is often misstated because of the nature of the constraints imposed on government, that is, private alterative uses are rarely considered. Private land is also incorrectly valued, though for a different reason—a consequence of using a shadow price adjustment to value benefits overtime for the benefits from the publicly held land but not applying a similar adjustment to the benefits derivable from the same land, if privately held.

If the Government does not accept the market interest rate as a basis of valuing its benefits and costs over time, it adopts a "social rate of discount," typically well below the market rate.** In applying this rate the Government is in all likelihood grossly underestimating its land costs. For instance, the Government might purchase an agricultural area and convert it to recreational use if the recreational benefits exceeded the costs of the land which presumably should equal the agricultural benefits. However, the value of the land would be the sum of the returns to agricultural use capitalized at the market rate of interest, possibly twice the assumed social rate. Therefore it is possible that a project may show a favorable benefit-cost ratio solely due to applying a shadow price—the social rate of discount—in the public sector and not extending the adjustment to the private sector.

The preceding error: the setting of shadow prices for a public discount rate but no adjustment for the private market rate is a general problem. This criticism can be extended to all of the capital inputs used by the public sector. Since their foregone use is market values lost, valued at the market interest rate, they will be underestimated relative to public benefits solely because the use of too low an interest rate.

IV. Shadow Prices and Constraints

We mentioned in the preceding section that public land is often incorrectly priced since private alternative use is not considered. This difficulty can be generalized to many parts of the Government. Institutional rules are established and then prices are estimated within the confines of the rules. For instance, the value of a water supply to an area may be based upon the cost of supply to that area, but there may be a very low cost supply which is ruled out because of a Government policy of dedicating the water to another use. Therefore the assigned value of the water may be grossly overstated from the perspective of the Nation. Or, an agency may value its inputs in program A in terms of loss of benefits in program B for which the inputs are substitutes. However there may be a much more productive use in another agency, but the possibilities of transfer of funds may be very meager. For this reason, the evaluation of the worth of inputs may be understated.

---

*Further discussion of this issue is found in the paper by Knetsch in vol. 3 of this collection.

**Further discussion of this issue is found in the paper by Baumol in this volume.

A classic example of a set of constraints which may run counter to estimated shadow prices arises from budget allocations in conjunction with some discretionary authority residing with agencies. Funds for an agency are raised through taxes or bonds; they are not a concern of the decisionmakers of an agency. The agency has no knowledge of the social costs of these funds, but they are very sensitive to the fact that their actions are very limited by the funds assigned to them. They are told that if there exists unemployment or arbitrarily fixed exchange rates, payments made for labor or imported goods inputs will not reflect their approximate social costs. The central authority might calculate a shadow price and instruct the agencies to use, say, a very low price reflecting that it is in surplus. If the agency were to use this price in its calculations it would use relatively more labor than otherwise, in sensible agreement with the social evaluation of its labor inputs. But if the agency is subject to a budget limitation, which is usually the case, it would find that it would have spent relatively more on the process using labor, that is, from its perspective a dollar spent on labor is worth a dollar spent on machines and though it may cost society less for them to use labor they make the same budgetary sacrifice whether labor had been over or underpriced. They will be highly resistant to adopting these shadow prices as operational tools in their choices of design of programs.

V. Estimation Procedures

The estimation procedures used to assign values to public outputs can be described as efforts to simulate market outcomes. It is assumed that for a variety of technical reasons—externalities, public goods, and so on—the private market must be supplemented by public production and distribution. The essential character of the objective function according to this view requires the assignment of prices which would have resulted from market behavior, if there had been some way to overcome the technical limitations which gave rise to public supply and a perfectly competitive market were able to operate. This individualistic view is usually supplemented by a judgment in regard to the distribution of income, a far from trivial departure from consumers sovereignty. The measurement rule used to determine the value of a government output is: estimate what the users of the public product would be willing to pay. Since the products are distributed at zero or low, conventional prices there are no direct measures by which the price can be discovered and therefore several indirect procedures are used to reveal the price.

1. The most common technique used to evaluate public output is to consider the product as an intermediate good and then to estimate the value of the marginal product of the good in further production, that is, assume the user is a producer and then ask: by how much does the public output increase his income? Illustrations are found in natural and human resources development. Some goods are easily and naturally treated in this fashion. For example, water supply is used for home consumption, but the great bulk of it is consumed in agriculture, power generation and industrial processing. Let us consider agricultural water. Productivity studies of irrigated farms are used to estimate the value of the product of an incremental acre-foot of water;

this is the agricultural value of the water. It is assumed that the farmer would pay this amount as a price for the water and therefore the marginal product is identified as the "imputed market price," or "shadow price" of water. The computation of this figure is not a simple task, but beyond these problems there are frequent errors in the application of this approach, even when the product is easily treated as an intermediate good.

In practice we find that the farmer is rarely asked to pay as a price to the Government the increment in income attributable to the water—an amount which is interpreted as the price he would be willing to pay. It is usually asserted that the unwillingness to charge this indicates a subsidy to agriculture. While this may be true, it also may be true that the computed shadow price may be far greater than a simulated market price for several reasons.

a. There may be alternative sources of supply and the cost of this supply would put a ceiling on the price that the farmer would pay for the water, for example, ground water pumping costs may be greater than the charged price but below the imputed value. The failure to consider this alternative is a shortcoming of this analysis.

b. The budget studies usually computed an expected value over several years. If the farmer is a risk averter he will assign a lower value to this anticipated income. It might be correct for the government to plan its operations on the basis of average expected values, but then it would have to absorb the risk and sell the water below the price. This problem of individual risk and government preference for expected values has not been studied in reference to the design and management of programs.

c. The computation of incremental productivity usually refers to averages rather than marginals. If the water is supplied at a low or zero price then we know that the user will purchase it to the point where the value of the marginal product would equal price. Since the average value will be greater than the marginal, multiplying the total amount which will be purchased by the average value will greatly overstate the total revenue from selling the product at a market price.

The divergency between average and marginal suggests another ambiguity in the concept of the price the users would be willing to pay. A simulated market price equates demand and supply, it balances off marginal gains and costs. But the price users may be willing to pay for the quantity rather than do without it may be much greater. In principle, the users would be willing to pay a sum equal to the area under the demand curve in order to receive the quantity rather than do without the project. If there are reasonably good alternatives to the public output the demand curve will be highly elastic and the difference between the area under the demand curve and the product of the market-clearing price and quantity would be small. As we shall see the limiting factor of alternative supply is considered in the estimation of benefits in some cases, but not others.

The above difficulties of estimating the productivity values of public services are small compared to those which arise when we consider services like education and health which have been valued as intermediate goods. For educational output, statistical studies have found a correlation between years of schooling and income. Therefore, it is argued that the earning capacity of an individual has been

increased by the additional years of schooling.* Further, it is said that the individual should be willing to pay that increment of expected income, less his foregone earnings while he is in school, for the educational services provided him. In practice, of course, an individual would not be expected to make such a payment, but this is attributed to the immature judgment of a student and his lack of capital. But does this relationship between education and income, even if it could be convincingly established, exhaust the reasons for public education or is it even the dominant factor? Clearly, the government is interested in many consequences beyond the income of the person. For instance, will he be a better and more responsible participant in the political process? Will he be a better neighbor? Will he have developed values and insights which will make him a more effective parent? Many more goals for education have been asserted, but all we want to establish here is that governments will find unsatisfactory a rule for the design of educational programs, or for the determination of its scale, which is based only upon the enhancement of an individual's expected income.

In the case of health services, a different set of problems develop.** Health is much more of a consumer good than education. Health, as an intermediate product, is valued by the additional working time and increased productivity associated with a reduction in disability. A saved life is valued at the present value of the expected income stream it would have earned. (Some would say that it should be net of the consumption of the saved life and therefore it should be the value of his savings.) Certainly individuals are concerned about the loss of working time but pain, discomfort and the fear of incapacity and death may be even more of a basis of willingness to pay to avoid illness. If the health program were designed to maximize benefits measured by income growth, then the diseases of the aged would be ignored and diseases of women would receive relatively little support. Neither individuals nor governments are prepared to accept the enhanced income as the sole basis for determining the private or social benefits of health.

2. A second indirect technique commonly used to estimate what individuals are willing to pay is based upon the cost savings of the public service, that is, the reduction in the costs that individuals would have incurred if the public service were not supplied. This approach is most commonly adopted in the fields of transportation and power. Generally, it is assumed that there is an inelastic demand for the output and, therefore, the public somehow would have managed to transport the goods or to develop energy though the costs would have been much higher. The major sources of savings are private carrier costs in the case of transportation and private generation of energy in the case of power.

The cost-savings approach to benefit measurement faces two problems: the identification of the real alternatives which were saved and the constraints imposed upon policy if the user savings are the

*Further discussion of this issue is found in the paper by Brandl in vol. 3 of this collection.

**Further discussion of this issue is found in the papers by Wholey in this volume, and Grosse in vol. 3 of this collection.

basis of evaluation. In the case of transportation, the costs savings are realized by shippers and travelers. An improved highway would reduce the travel time and operating costs of traveling between any two points. Presumably the users of the highways would be willing to pay that difference.

The savings for many public services should be the alternative facilities or programs which would have been provided by the government. Unfortunately, public agencies are notoriously poor in considering alternatives. For instance, an apprenticeship training program may be an alternative to vocational training in a school, but the educational agency is unlikely to consider an inservice program as an interesting or feasible alternative. Importing goods may be an alternative to agricultural expansion, but this is not likely to be considered. Organizations are not active searchers for information about alternatives they are not likely to pursue it. It is also true that many of the most feasible alternatives may be those ruled out by legal or administrative constraints and never even considered.* For instance, a change in the tax structure may provide a very different set of incentives for private consumption but it might never be considered by an operating agency. As a consequence, estimates of the real costs of alternatives which are saved are notoriously bad, except for the calculation of user savings in some cases.

The assumption of inelastic demand leads to greatly inflated multiplier of overstated benefits. The increased traffic generated by a public project would not have existed at the old cost structure and therefore it would be an error to assume a benefit equal to the unit cost savings of the old traffic multiplied by the augmented new traffic, including the increased flow. This problem becomes very acute when activity shifts among cities or ports in response to slight improvement of facilities. There may be large movements responding to very slight gains in private transportation costs, possibly far below the amount necessary to pay for the capital costs of the facilities.

The limitations of the use of costs of alternative programs as a measure of benefits is reflected in the analytical studies of the PPBS groups. These studies are appropriately labeled as cost-effective rather than cost-benefit analyses. The objectives of the program are usually specified in physical terms and the alternatives where trade-offs are considered as restricted to those under the control of the agency. The narrow view is typified by the label given the function of the military system analysts: the most bang for the money rather than the most national security or social welfare.

The evaluation of public output in terms of costs or savings by users is based upon an assumption that we are to be guided by the efficiency calculation of the individual beneficiaries of the project. Political and administrative leaders resist the policy conclusions drawn from these studies since they see them as restrictive of their freedom to plan the development of the nation. For instance, the benefit-cost calculations might indicate the most efficient transportation network but efficiency has never been a sufficient criterion for governments in their locational policies. It is possible that governments have erred in trying to sup-

*Further discussion of this issue is found in the paper Achinstein in this volume.

port their declining regions, to populate their empty places, or to stem the flow of population to their capitals, but the public support of these programs has been great. Though economic arguments of external economies and diseconomies have been used to defend these programs, it is clear that regional objectives have been pursued for their own sakes, with a willingness on the part of the government to sacrifice national income for these benefits.

Cost savings on the part of individuals or agencies need not be the only frame of reference in trying to discover shadow prices. There is an alternative formulation which views the legislature or administration as representatives of the aggregate of citizens. It is argued that unrestrained political bargaining is not optimal but that the addition of appropriate information to the decisionmaking process would lead to optimal outcomes. The first step of the analyst is to design a quantitative measure of the product of the public output. This is not difficult in the case of most commercial, marketable commodities like food, clothing, or machines, but for services like recreation, education, or national defense, an appropriate unit of output is not obvious. The most common measures refer to use of the service rather than their desired qualities; e.g., years of schooling rather than increased productivity, socialization, and so on. The second step is to estimate the real costs of resources necessary to produce the outputs. The above information is equivalent to the marginal rate of transformation between two public services. The decisionmaker is then asked to revise the expenditure levels among the public services—the revision, if the authorities are responsive to the public, would be in the direction of equating the marginal rate of substitution in utilities to the marginal rate of transformation in production. A decisionmaker who seeks to be most effective will welcome this information-structure, but there is a great gap between the decisionmaker's objective function and the social welfare, defined by aggregating individual preferences. In practice, this pattern of information of real costs and benefits for decisionmaking is appropriate to the view of a public interest, defined as a social ordering expressed by the government.* The net result of applying this criterion, besides optimal allocation, is the derivation of an appropriate set of shadow prices. These are the relative costs of the different public services when the government has decided that the budget has been optimally allocated. The ratio of marginal costs of two public outputs will be the same as the ratio of their social values and therefore the marginal costs figures will have the same interpretation we give to competitive market prices.

3. The third major technique of shadow price estimation is to estimate directly the users prices by appeal to market information. This is the most difficult task, but it may prove to be the most fruitful. In many cases there are near substitutes for collective consumption. There is usually a private educational, health, or recreational market; the extensive study of this market may provide the needed price information. The difficulty facing the analyst is that the comparable private commodities are sometimes very different. The characteristics of service of a private medical clinic may be sufficiently similar to a public

* Further discussion of this issue is found in the paper by Steiner in this volume.

clinic that the private data may be usable, but the differences between a public park and a private camp are huge and difficult to compare.

Another form of use of market data is more indirect, it relies on the responses of the private sector in gaining access to the free public services. Public services are free, but access to them may be costly. Parks are free, and since they are desirable men will pay higher rents for sites located close to them. There would be similar shifts in the demand for land because of differential quality of schools, medical facilities, highway systems, and so forth. Households will reveal their preferences by their locational decisions, and further, the revelation will be quantitative. An analysis of the household's costs may provide information about the value they assign to these public services. This form of analysis will require complicated econometric studies, since changes in behavior will be due to many factors and some shifts will be due to the initial changes of the users of the public services rather than to the public services themselves.

VI. Shadow Prices and Incentives*

We have alluded to the problem of incentives to public officials, but that brief allusion greatly understates their importance. Shadow prices are useful pieces of information, but unless the decisionmakers benefit by acting on the basis of those prices little is gained. It is utopian to assume that disinterested scientists will compute the shadow prices and devoted public servants will accept them as binding. When I discussed the use of a low shadow price for unemployed labor I pointed out that since the agency is subject to a budget constraint they are not likely to be receptive to a rule which tells them to treat labor as though it were a free good when it is obvious to them labor inputs place a drain on their limited budget. A similar set of improper incentive is reflected in the continuous battling between highway authorities and conservationists. Public parks provide "free or cheap land" from the perspective of the construction agency and since their budget is limited—the use of public land maximizes the amount of traffic served per dollar of their budget.

The problems of incentives and shadow prices may be of even greater magnitude when we deal with the many decentralized programs initiated by the Federal government in the urban areas. The Federal government may identify targets and even assign values to outputs but if incentives to local governments are not consistent with these values the federal objectives are not likely to be achieved. For instance, fiscal profitability is of minor concern to the national interest but the payoffs of a project along this dimension often dominate the local design and execution of a program.

The thrust of the above remarks is that the problem of shadow price determination is not simply one of calculation. If the shadow prices are to guide behavior then those who must make and implement the decisions require incentives to provide the "correct" information needed for calculation and to use the prices. Therefore the study of shadow pricing rules opens up the even more difficult study of the optimal structure of government.

*Further discussion of this issue is found in the paper by Schultze in this volume.

EVALUATING PUBLIC EXPENDITURES UNDER CONDITIONS OF UNEMPLOYMENT

BY ROBERT H. HAVEMAN*

Robert H. Haveman is Associate Professor of Economics at Grinnell College, currently on leave as Senior Economist of the Subcommittee on Economy in Government of the Joint Economic Committee.

Under conditions of full employment, the monetary costs of public expenditures closely approximate the social costs. This is so because the prices paid for the labor and capital inputs to public investments accurately reflect the value of the outputs which are being foregone when resources are diverted to government projects. However, public investments are often undertaken under conditions of less than full employment. "To the extent that otherwise unemployed resources are drawn into use by the public expenditure, the social cost of the expenditure . . . is less than the market or monetary cost . . . by definition, these unemployed resources are not producing other things." This divergence between monetary and social costs can be observed even under conditions of relatively low overall unemployment, because of the "substantial variation of unemployment rates by occupation, by industry, and by region, around the national average unemployment rate."

Using the results of a recently developed computational model, Professor Haveman explains how the true social cost of a public expenditures can be estimated. This framework proceeds by estimating the occupational, industrial, and regional pattern of labor and capital demands generated by the project; relating this pattern of demands to unemployment data for comparable categories: and then adjusting the monetary costs for the idle resource use which is shown to be incorporated in them.

He then describes a method of adjusting for the divergence between "nominal" (monetary) benefit-cost ratios and the more relevant "opportunity cost" (social) benefit-cost ratios. As a result of this adjustment, projects whose benefit-cost ratios might have been unacceptable under full employment assumptions may become acceptable under conditions of less than full employment. He analyzes the effect of using social cost estimates on the design, location, and priorities of public investments and concludes that "only through substituting social opportunity costs for nominal monetary costs in the expenditure criterion can public decision-makers isolate expenditures which are both intrinsically economic and substantial employment generators."

Introduction

In the post-war period, the expenditure side of the public sector has been the subject of a substantial amount of analysis. While some of this work developed and refined the theory of public expenditure analysis, other work was aimed at improving the methods for evaluating the economic gains and the costs of alternative expenditure projects. This paper reflects the latter of these two emphases. It presents the reasons why monetary costs fail to reflect real social costs when the economy is experiencing unemployment and excess capacity, and elaborates a method for adjusting monetary costs when such conditions prevail. The implementation of this method eliminates the overstatement of real costs by monetary costs when some of the resources drawn into use by a public expenditure have idleness as their alternative.

*The views expressed in this paper are the author's and not necessarily those of the Joint Economic Committee.

I

An essential proposition in public expenditure economics is that, in evaluating the economic worth of a public expenditure, it is the *social* costs and *social* benefits which must be evaluated and not the private (or monetary) costs and private (or monetary) benefits. A second proposition is that, while social and private values may well be identical, they are not likely to be when there is some imperfection in the operation of the market system.* Clearly, the existence of unemployed resources represents market system imperfection which would cause these two values to diverge.[1]

If resources in the economy are fully employed, the monetary costs of the labor and capital purchased by a public expenditure are likely to be a good approximation of the value of the things which society would be able to enjoy if the expenditure were not undertaken.[2] This is so because the price paid to resources employed in any enterprise tends to equal the value of what these resources are producing in that use. When they are hired away from that activity, society is forced to forgo the output which they would have produced.

Consider, for example, a $100 public expenditure which is used to purchase $100 worth of labor and capital. If there were full employment and if the economy were functioning ideally in other respects, these resources would have been used to produce $100 worth of goods and services of some unknown composition which would have been purchased, used, and enjoyed by members of the society. This $100 worth of goods and services, then, is the social cost necessitated by the public expenditure. Stated alternatively, because of the $100 public expenditure, resources of that value are diverted from producing $100 worth of other things and society is forced to forgo the opportunity of using and enjoying these particular "other things."

However, when there is not full employment of labor, or when plant capacity is not fully used, some of the resources hired by the public expenditure may not have to be diverted from alternative uses. Some of them may be drawn from the pool of unused or idle labor or capital resources. In the case in which the public expenditure employs otherwise idle resources, society does not have to forego the opportunity to use and enjoy other things. By definition, these unemployed resources are not producing other things. Consequently, to the extent that otherwise unemployed resources are drawn into use by the public expenditure, the social cost of the expenditure—the value of the alternative uses that would have been made of the required manpower and capital—is less than the market or monetary cost.

II

It is clear that public expenditures made during a period of substantial unemployment would call into use some resources which would otherwise have been unemployed. Not so obvious, however, is the fact

[1] William J. Baumol, *Welfare Economics and the Theory of the State* (second edition, Cambridge: Harvard University Press, 1967), pp. 135–142. In that volume, Baumol treats unemployment as an external diseconomy requiring a collective remedy outside of the market system.

[2] This, of course, assumes that the market system is operating at its "efficient best" in all other respects.

*Further discussion of this issue is found in the papers by Davis & Kamien, Kneese & d'Arge, Zeckhauser, and Demsetz, in this volume.

that a similar result is likely when the national unemployment rate is quite low. However, this latter proposition—that some resources called into use by the marginal public expenditure would otherwise be idle, even under conditions of relatively full employment—is a point which must not be neglected.

The reason for this effect is the substantial variation of unemployment rates by occupation, by industry, and by region around the national average unemployment rate. If the Nation showed a full employment rate of 4 percent and if every occupation in every region showed that same rate of unemployment, a public expenditure could in all likelihood cause 100 percent of the resources which it required to be diverted from other uses. However, if the 4-percent national rate is composed of a 2-percent unemployment rate in some occupations (regions) offset by a 6- to 7-percent unemployment rate in other occupations (regions) it is not likely that all of the labor which is hired by the public expenditure would be diverted from other uses. A part of these resources would be drawn from the high unemployment occupations (regions) in which case the public expenditure would call into use some resources which would otherwise have lain idle. Indeed, in the case in which there is substantial variation of occupational, industrial, and regional unemployment around the national unemployment rate, it is conceivable that the full set of resource demands imposed by the public expenditure might be met by units of labor and capital drawn from the idle pool.

Table 1 shows that, in fact, there is a substantial amount of variation of occupational, industrial, and regional unemployment rates around the national average. The distribution of unemployment rates around the national average is presented for 1960. In that year the national unemployment rate was 5.6 percent. From the data in these distributions, it is clear that it is necessary to know the structure of demands which a public expenditure imposes on the economy in order to determine the extent to which the expenditure does or does not use resources which would otherwise be unemployed.

TABLE 1.—VARIATION OF OCCUPATIONAL, INDUSTRIAL, AND REGIONAL UNEMPLOYMENT RATES AROUND NATIONAL AVERAGE, 1960

Unemployment rates	Occupational [1]	Industrial [2]	Regional [3]
Below 2.6	4	1	0
2.6 to 3.6	3	2	6
3.6 to 4.6	0	0	8
4.6 to 5.6	2	4	15
5.6 to 6.6	3	4	11
6.6 to 7.6	0	1	7
7.6 to 8.6	1	2	2
8.6 or more	5	2	1

[1] Data for 18 major occupational categories.
[2] Data for 16 major industry categories.
Data for 50 States.

When this result is related to the opportunity cost logic presented in section I, it becomes clear that much public spending in the postwar period imposed social—or opportunity or real—costs on the society which were less than the monetary costs.[3] Moreover, and more impor-

[3] If an unemployment rate of 5 percent is defined as full employment, 10 of the 21 years since World War II were years with idle productive capacity in excess of this minimum; 15 of the 21 years saw unemployment in excess of the frictional minimum if 4 percent is the full employment rate.

tantly given today's unemployment situation, some spending which occurs during periods of rather full employment may entail the use of resources which would otherwise have been unutilized or underutilized. This would be especially true if the pattern of resource demands imposed by a particular public expenditure emphasized the occupations, industries, and regions which had substantial unemployment even though, overall, the economy was rather fully employed. Again, the use of these resources entails zero opportunity costs.[4] Clearly, the accurate economic evaluation of the social costs of a public expenditure requires a detailed estimate of the pattern of the occupational, industrial, and regional demands imposed by the expenditure and a comparison of these demands with the existing pattern of occupation, industrial, and regional unemployment, *both* when the economy is not fully employed *and* when it is.

III

The first step in evaluating the opportunity costs of a public expenditure is to estimate the pattern of the demands generated by the expenditure for labor, by occupation, for capital, by industry, and for both labor and capital, by region. While the pattern of labor and materials employed directly by the expenditure is not difficult to ascertain, the set of final labor and capital demands imposed after the material inputs are traced through the several rounds of the production process is far more difficult to estimate. Recently, this estimation task has become possible because of the national input-output matrix assembled and published by the Office of Business Economics of the U.S. Department of Commerce [5] and the industry-occupations matrix completed by the Bureau of Labor Statistics of the U.S. Department of Labor.[6] When these empirical matrices are incorporated along with basic estimates of the direct resource demands of a public expenditure into an appropriate computational model, the full catalog of direct and indirect demands placed on factor sources—by occupation, industry, and region—can be estimated.

In one particular model developed for the purpose of estimating the complete pattern of labor and capital demands imposed by a public expenditure, the sequence of computations proceeds as follows:[7] Given the basic data on the direct material, equipment and supply inputs required by an expenditure, the complete pattern of industrial demands can be calculated through use of the input-output matrix. Then, on the basis of a set of relationships which grant a preferred status to the region in which the expenditure is undertaken and the

---

[4] Implicit in this position is the proposition that involuntary leisure has zero benefit to either the unemployed worker or the society.

[5] Morris R. Goldman, Martin L. Marimont, and Beatrice N. Vaccara, "The Inter-Industry Structure of the United States," A Report on the 1958 Input-Output Study, *Survey of Current Business*, 44 (Nov. 1964), 10–29; Norman Frumkin, "Construction Activity in the 1958 Input-Output Study," *Survey of Current Business*, 45 (May 1965), 13–23; National Economics Division Staff, "The Transactions Table of the 1958 Input-Output Study and Revised Direct and Total Requirements Data," *Survey of Current Business*, 45 (Sept. 1965), 33–49.

[6] U.S. Department of Labor, Bureau of Labor Statistics, *Handbook of Methods for Surveys and Studies*, Bulletin No. 1458 (Washington, 1966), Chap. 7, and U.S. Department of Labor, Bureau of Labor Statistics, *Occupational Employment Statistics, Sources and Data*, Report No. 305 (Washington, June 1966).

[7] An elaboration of the details of this model and its use can be found in Robert Haveman and John V. Krutilla, *Unemployment, Idle Capacity and the Evaluation of Public Expenditures* (Baltimore; Johns Hopkins Press, 1968).

geographic location of each industry's capacity, the geographic distribution of these total industrial demands is estimated. Third, by using the industry-occupation matrix, the labor demands imposed on each region because of the industry output demands are estimated, in occupational detail. Fourth, these occupational labor demands generated by purchases of materials, equipment, and supplies—and distributed among the regions—are added to the onsite occupational pattern of labor demands by region. Finally, the pattern of demands imposed on capital are determined by industry and by region by applying appropriate capital-output ratios to the total output demands, by industry, which were estimated in the second step.

By employing this model, the pattern of resource demands can be computed for any public expenditure in 156 occupation, 80 industry, and 10 region details. Table 2 shows, in substantially consolidated form, the kind of detailed estimate furnished by this model. In that table the pattern of occupational, industrial, and regional demands is shown when a multiple-purpose—including hydroelectric power generation—water development project is constructed in the Lower Atlantic States.

In the final column of this table, it is shown that a total gross output of $1,032 per $1,000 of total project cost is generated by the direct purchases of materials, equipment, and supplies required for the project.[8] Of this total gross output demand, 29 percent of it, or $300, is imposed on the Lower Atlantic region—the region where the project is assumed to be constructed. In addition, because of the heavy demands which this kind of installation places on durable equipment manufacturing, a substantial set of demands are imposed by the project on the Mid-Atlantic and East North Central regions where these industries are concentrated. Together, these three regions account for over 70 percent of the total gross output stimulated by the expenditure. That it is the durable goods industries which account for this regionally concentrated result is also seen in the table. Of the total gross output of $1,032, durable goods production accounted for $539, or over 50 percent. Of the $539 of durable goods output, over $320 or 60 percent is produced by the Mid-Atlantic and East North Central regions.

In the lower portion of table 2 the labor demands required by the project are shown in occupational detail and by region. Because of the nature of this kind of construction installation, many of the labor demands are required on the construction site. This accounts for the heavy concentration of labor demands in the region in which the project is assumed to be constructed. The Lower Atlantic region supplies nearly three-fourths of the total labor demand generated by the project. Consistent with the gross output estimates which demonstrated the concentration of durable goods in the highly industrialized regions, it is seen that those labor demands which the project imposes on other regions are concentrated in the Mid-Atlantic and East North Central regions and among the craftsman and operatives occupations.

[8] This $1,032 represents the total gross output, both direct and indirect, generated by the final expenditure. Its size and industrial breakdown was estimated by an input-output calculation. The value of final demand, representing direct purchases of materials, equipment, and supplies, was $514 (out of $1,000 of total project cost). It is this final demand which, through the input-output calculation, generated the $1,032 of total gross output. The portion of the $1,000 of total project cost not represented by direct purchases of materials, equipment, and supplies is largely accounted for by the direct demand for on-site labor.

TABLE 2.—GROSS OUTPUT BY INDUSTRY AND TOTAL LABOR COST BY OCCUPATION IN EACH OF 10 REGIONS FOR A $1,000 PUBLIC EXPENDITURE ON A MULTIPURPOSE WATER RESOURCE PROJECT ASSUMED TO BE CONSTRUCTED IN THE LOWER ATLANTIC REGION, IN DOLLARS PER $1,000 OF EXPENDITURE[1]

INDUSTRY	New England	Mid-Atlantic	East north-Central	West north-Central	Southeast	Lower Atlantic	West south-Central	Mountain and west Coast	Total
Agriculture, forestry, and fisheries	[2] (5)	[2] (6)	1(15)	2(21)	1(16)	[2] (4)	1(13)	2(21)	8
Mining (including crude petroleum)	[2] ([3])	4 (5)	4 (5)	4 (6)	2 (4)	35(58)	10(14)	7 (9)	77
Construction	[2] (2)	1(11)	1(16)	[2] (3)	[2] (3)	5(54)	[2] (4)	[2] (6)	9
Nondurable goods manufacturing	7 (7)	20(20)	24(23)	5 (5)	10(10)	11(11)	14(14)	11(10)	103
Durable goods manufacture	31 (6)	126(23)	197(37)	27 (5)	29 (6)	49 (9)	25 (5)	56(10)	539
Lumber and wood products	1 (5)	1 (7)	2(12)	[2] (4)	2(16)	1 (8)	1 (7)	6(40)	13
Stone, clay and glass products	4 (5)	15(19)	19(24)	6 (8)	7(10)	13(17)	5 (6)	9(12)	78
Primary metals industries	6 (4)	37(28)	47(36)	4 (3)	6 (5)	14(11)	3 (3)	14(10)	129
Fabricated metal industries	6 (6)	23(23)	31(31)	6 (6)	8 (8)	12(12)	5 (5)	11(11)	100
Machinery (excluding construction and electrical)	8(10)	22(29)	34(45)	3 (4)	2 (2)	2 (1)	1 (1)	4 (5)	75
Construction machinery	1 (1)	4 (8)	27(52)	4 (7)	1 (3)	2 (3)	9(17)	4 (9)	52
Electrical machinery	5 (8)	19(29)	26(40)	3 (5)	2 (3)	3 (5)	1 (1)	5 (8)	64
Transportation equipment	1 (4)	3(14)	11(52)	1 (4)	1 (5)	1 (7)	1 (3)	2(10)	20
Miscellaneous manufacturing	1(14)	3(45)	1(21)	[2] (5)	[2] (3)	[2] (2)	[2] (2)	[2] (7)	7
Transportation and warehousing	1 (2)	6 (8)	8(11)	2 (2)	2 (3)	49(58)	2 (3)	3 (4)	72
Wholesale and retail trade	1 (2)	6 (6)	9 (9)	1 (2)	2 (2)	70(75)	2 (2)	3 (3)	93
Services	4 (3)	14(11)	21(16)	4 (3)	4 (3)	71(54)	6 (5)	7 (6)	131
Total gross output[4 5]	45 (4)	177(17)	264(26)	45 (4)	52 (5)	300(29)	61 (6)	89 (8)	1,032

OCCUPATION									
Professional, technical, and kindred	2	6	9	1	1	49	2	3	72
Managers, officials, and proprietors	1	5	8	1	2	43	2	3	66
Clerical, and kindred workers	2	6	9	1	1	22	2	3	46
Salesworkers	(2)	1	1	(2)	1	6	(2)	(2)	10
Craftsmen, foremen, and kindred	3	11	18	3	3	274	3	5	319
Carpenters	(2)	(2)	(2)	(2)	(2)	48	(2)	(2)	49
Cement finishers	(2)	(2)	(2)	(2)	(2)	4	(2)	(2)	4
Iron and metal workers	(2)	(2)	(2)	(2)	(2)	22	(2)	(2)	22
Construction equipment operators	(2)	1	1	(2)	(2)	45	(2)	(2)	48
Other building trades	(2)	1	1	(2)	(2)	37	(2)	(2)	41
Mechanics	(2)	2	3	(2)	(2)	12	(2)	1	18
Labor foremen	1	2	4	1	(2)	48	1	1	58
Other craftsmen, foremen, and kindred	1	5	9	1	1	58	1	2	88
Operatives and kindred workers	4	13	20	3	4	54	3	7	108
Truck and tractor drivers	(2)	2	2	1	(2)	15	1	1	22
Sailors and deckhands	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)
Other operatives and kindred	3	12	17	2	3	39	3	5	86
Service workers	(2)	1	2	(2)	(2)	3	(2)	(2)	7
Laborers	1	3	4	1	1	77	1	1	89
Farmers and farmworkers	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	2
Total labor cost [4] [6]	12 (2)	47 (7)	70(10)	12 (2)	14 (2)	527(73)	13 (2)	24 (4)	719

[1] Percent of row totals stated in parentheses behind the dollar values.
[2] Less than $0.1 but greater than 0.
[3] Less than $0.5 but greater than 0.
[4] Rows may not add because of rounding.
[5] Percent of total gross output stated in parentheses behind regional total gross output figures.
[6] Percent of total labor cost stated in parentheses behind regional total labor cost figures.

Finally, the substantial disparity in the pattern of industrial and occupational demands generated by various project types should be noted. While all of the project types analyzed were water resource investments, the anatomy of their industrial, occupational, and regional impacts is far more diverse than is generally recognized. While some project types require very little on-site construction (dredging), others require the installation of huge capital facilities (multipurpose projects). For 12 project types analyzed, the ratio of labor compensation (direct and indirect) to total project cost ranges from .52 to .72. The range in the ratio of on-site labor cost to total labor cost extended from .25 to .58. The ratio of durable goods demanded to gross material demands extends from .2 to .65. When all of the project types are assumed to be constructed in the Lower Atlantic region, the percent of national gross output retained in that region ranges from 24 to 32 percent; the percent of national labor cost retained extends from 69 to 75 percent. Even more radical disparities among project types are noted as detailed industrial or occupational sectors are studied.

IV

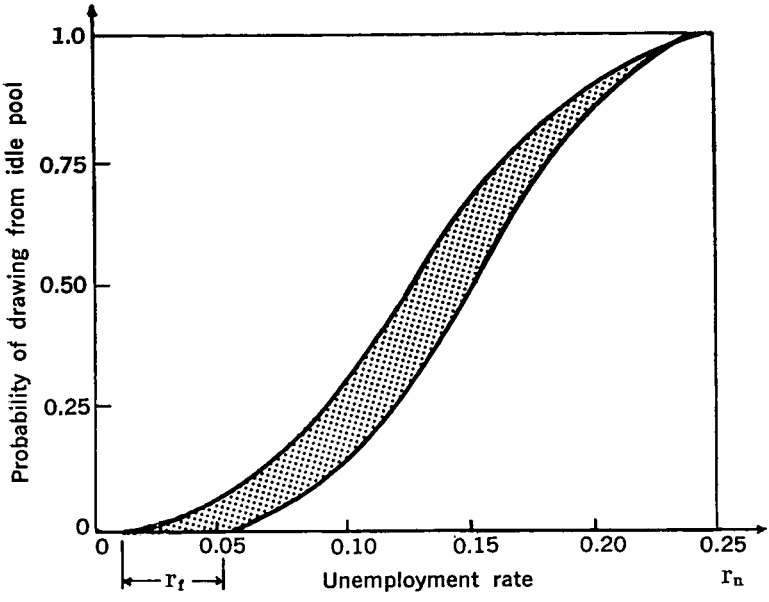
Having ascertained the pattern of resource demands imposed by a public expenditure, the next step in evaluating the social costs of these demands is to compare them with the occupational and regional pattern of labor unemployment and the industrial pattern of excess plant capacity. As described above, the monetary costs of a public expenditure represent real opportunity returns forgone at the margin only if all of the resources used had alternative employs. If there is unemployment, however, some labor used is likely to be drawn from the idle pool. This labor has no comparable opportunity cost. Similarly, the opportunity rate of return on otherwise idle capital drawn into use by the expenditure is zero. However, because capital services are largely storeable, depreciation charges are a real cost properly imputed to the expenditure even when otherwise idle industrial capacity is drawn into use.

On this basis, the occupational, industrial, and regional breakdowns of monetary costs can be modified to the extent that the units of labor and capital represented would have been otherwise idle. To estimate the extent to which any labor and capital demand employs otherwise unused resources, it is necessary to trace each unit of labor and capital employed to its source and to inquire concerning its alternative use. In the absence of data necessary to implement this counsel of perfection, the model discussed here assumes that the levels of occupational unemployment (or, in the case of capital, industrial excess capacity) are significant determinants of the proportion of labor drawn from any occupation and region (capital drawn from any industry) which would have, in the absence of the expenditure, been idle. For example, this approach treats an increase in the demand for labor at low levels of unemployment as simply shifting workers among jobs without reducing unemployment below the frictional minimum. However, as the rate of unemployment (excess capacity) rises, so too does the probability that the incremental demand will draw otherwise unemployed labor (idle capital) into use. Because accurate knowledge on the pattern of labor and capital market response does not exist, a set

of synthetic response functions is employed. These functions relate the probability that a given increment in the demand for labor and capital will be drawn from otherwise unemployed resources to the level of occupational unemployment and industrial excess capacity on the basis of reasonable assumptions concerning market operation.

In figure 1, the kind of relationships used to estimate the extent to which labor demands are supplied from otherwise unutilized resources

Figure 1



in this model are shown.[9] The set of synthetic functions inscribed within the area between the two curves states that the higher the unemployment rate, the greater the proportion of labor from any given occupation which is hired from the idle pool. The region labeled r_f describes the range of unemployment rates at which each of the major occupational categories is said to be fully employed.[10] Full employment for each occupation is defined by the national unemployment rate experienced by that occupation in 1953—a year with minimum unemployment without undue inflationary stress. The point labeled r_n signifies the rate of unemployment at which an increment of demand would be entirely supplied from otherwise unutilized resources. For the set of relationships included in the shaded area of figure 1, this unemployment rate is .25, which is the estimated rate of unemployment at the height of the depression of the thirties. It is assumed that under such conditions, increments to the demand for labor and capital are satisfied with no displacement of alternative outputs. The relationships incor-

[9] The area between the two curves is the region within which the response functions, one for every major occupational category, fall. A separate set of functions, not shown here, was used to estimate the proportion of capital demands, by industry, which were satisfied by otherwise idle capacity.

[10] There is a single rate for each occupational category within r_f.

porated into the curves which lie in the shaded area are offered as an accurate portrayal of actual labor market behavior.

These relationships (and similar ones for estimating the withdrawal of capital from the idle pool) are used with detailed data on the level of occupational unemployment by region and industrial excess capacity to provide the basis for estimating the real costs of public expenditure. By combining the labor and capital response relationships with detailed evidence on the occupational, industrial, and regional patterns of unemployment, an estimate of the proportion of the labor and capital withdrawn from the idle pool in each pertinent occupation, industry, and region is obtained. By multiplying these percentages by the dollars of monetary cost in each category, the amount of monetary cost which, because of the use of otherwise unemployed resources, is not matched by social cost is estimated. When the monetary costs are adjusted for the idle resource use which is incorporated in them, the remainder represents the true social cost of the public expenditure.[11]

V

In tables 3 and 4, some estimates of social cost and its relationship to market cost are shown for a sample of public expenditures. These estimates are based on the unemployment conditions experienced in 1960. As noted, the unemployment rate was 5.6 percent in that year. Consequently, the adjustment to market cost required for investments undertaken in that year was substantially greater than for investments undertaken in more recent periods with unemployment levels below 4 percent. In table 3, the variation in the percentage which social costs form of total *labor* costs is shown for five public expenditure categories in the water resources area. While this data shows the influence of regional unemployment differentials on the degree to which *total* monetary labor cost requires adjustment, the model described above also generates tables which highlight the variation in social cost as a percentage of monetary cost for numerous detailed occupational and industrial categories.

In table 4, estimates of social cost as a percent of the total *expenditure* are shown for the same public projects constructed in each of the 10 regions, again with unemployment conditions prevailing in 1960. Table 4 demonstrates the substantial variation in the required cost adjustment which exists among project types. It also shows that the variation in adjustment for any single project type as its geographic location changes is even more significant than the variation among project types. In no case does the range for the former variation fail to exceed 15 percentage points. The influence of geographic unemployment on required cost adjustment is clearly seen by comparing the cost adjustments for projects constructed in the high unemployment lower Atlantic region with similar data for project construction in other regions. For every project type, the cost adjustment required for construction in this region is at least 10 percentage points below the median adjustment for all regions.

[11] On the basis of this formulation, it is seen that the exercise of adjusting nominal costs for the unemployed labor and idle capital which is used is an example of "shadow pricing." See the paper by Margolis, in this volume.

TABLE 3.—ESTIMATE OF SOCIAL LABOR COST AS A PERCENT OF MARKET LABOR COST FOR 5 REPRESENTATIVE PUBLIC EXPENDITURES IN 10 REGIONS OF PROJECT LOCATION, 1960

Region	Large earthfill dams	Local flood protection	Medium concrete dams	Large multiple-purpose projects	Dredging
New England	88	82	88	87	83
Mid-Atlantic	80	74	82	81	76
East North Central	90	89	92	92	89
West North Central	87	84	89	89	88
Southeast	93	92	94	94	88
Lower Atlantic	75	65	74	73	73
Kentucky-Tennessee	81	75	81	80	77
West South Central	92	91	93	93	86
Mountain	91	92	93	93	94
West Coast	85	82	86	86	74
Range of percentages	75–93	65–92	74–94	73–94	73–94.0
Median percentage	87.5	83	88	88	84.5

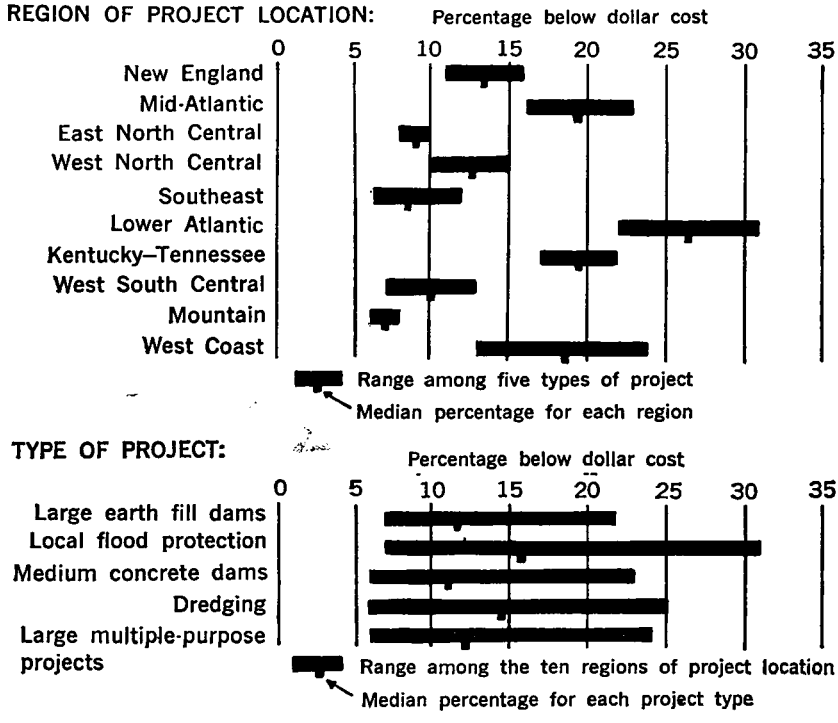
TABLE 4.—ESTIMATE OF TOTAL SOCIAL COST AS A PERCENTAGE OF TOTAL EXPENDITURE FOR 5 REPRESENTATIVE PROJECT TYPES IN 10 REGIONS OF PROJECT LOCATION, 1960

Region	Large earthfill dams	Local flood protection	Medium concrete dams	Large multiple-purpose projects	Dredging	Range of percentages	Median percentage
New England	89	84	88	87	84	84–89	86.5
Mid-Atlantic	82	77	84	82	78	77–84	80.5
East North Central	91	90	92	92	90	90–92	91.0
West North Central	88	85	90	89	89	85–90	87.5
Southeast	93	93	94	94	88	88–94	91.5
Lower Atlantic	78	69	77	76	75	69–78	73.5
Kentucky-Tennessee	83	78	83	81	79	78–83	80.5
West South Central	92	92	93	93	87	87–93	90.0
Mountain	92	92	93	93	94	92–94	93.0
West Coast	86	84	87	87	76	76–87	81.5
Range of percentage	78–93	69–93	77–94	76–94	75–94		
Median percentage	88.5	84.5	89	88	85.5		

The results of both tables 3 and 4 are summarized in figure 2. The charts shown there display the percentages by which the dollar costs of the selected public expenditures undertaken under economic conditions similar to those of the 1957–64 period—of which 1960 is taken to be typical—would overstate the social costs. The differences vary with the unemployment levels and other economic conditions in the region where a project is located, and also with the amounts, kinds, and origins of labor and materials required for each type of project.

Examination of this data suggests that the social cost of public expenditures for investment projects undertaken in 1960—and by inference from 1957–64—is between 70 and 90 percent of nominal monetary expenditures. The precise percentage depends on the category of expenditure, the region in which it is undertaken, and the nature of the relationship used to relate the rate of idle resources to the proportion of resources demanded by the public expenditure which will be withdrawn from the idle pool.

Figure 2



VI

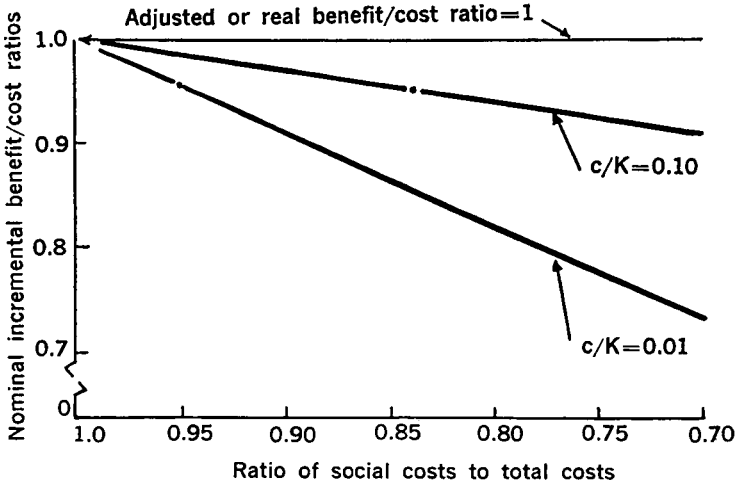
Given the estimated social costs, the question remains as to the extent to which the "nominal" benefit-cost ratios computed for projects constructed during the slack conditions between 1957–64, for example, or for chronically depressed areas, diverge from the more appropriate "opportunity cost" benefit-cost ratios.

While nominal capital (construction) costs overstate opportunity costs by the magnitude suggested, the effect on the benefit-cost investment criterion will be dampened to the extent that future operating, maintenance, and interim replacement costs occur in a fully employed economy. It is convenient, although not essential, to assume that investments made under conditions of unemployment do not operate in a less than fully employed economy. With this assumption, the component of annual project costs requiring adjustment will be only the capital charges—that is, the interest and amortization component of annual costs. The required adjustment in annual costs, therefore, will be a function of the ratio of annual operating, maintenance, and interim replacement costs (c) and the capital (or construction) costs (K).

The impact of construction cost adjustment on benefit-cost ratio calculations is shown graphically in figure 3 on the assumption that only construction costs require adjustment. Here the divergence in the nominal benefit-cost ratio from the true social benefit-cost ratio is analyzed for capital intensities (c/K) of .01 and .10. The upper curve shows the relationship of the social to the nominal benefit-cost ratio at different ratios of social capital costs to total capital costs for a public expenditure in which the annual operating costs (c) are 10 percent of initial construction costs (K). The lower curve shows the same relationship for a more capital-intensive expenditure—one in which annual operating costs are only 1 percent of initial capital costs. From these relationships, it is seen that when social costs for initial project construction are, say, 75 percent of total monetary costs, the nominal benefit-cost ratio for an economically efficient project can be as low as .78:1 if the undertaking is very capital-intensive (say, c/K=.01). On the other hand, for a less capital-intensive expenditure (say c/K=.1), a ratio of social capital costs to total costs of 75 percent would require a nominal benefit-cost ratio of at least .9:1 for the investment to be efficient. On the basis of this evidence, it can be asserted that most proposals for heavy construction projects bearing an unacceptable benefit-cost ratio from .85–.99 when evaluated under full employment assumptions would be deemed efficient under the conditions of unemployment and excess capacity of the sort prevailing in the 1957–64 period.

Figure 3

Ratio of Social Capital Costs to Total Market Costs



VII

From the analysis of this paper, it is clear that the level and distribution of unemployed resources in the economy does affect the evaluation of the social cost imposed on the economy by public expenditures. Moreover, the pattern of unemployment and excess capacity should influence the *design*, *location*, and *priorities* of public investments to be constructed during any time period. While the general proposition which follows from the study is that the monetary cost of public expenditures overstates the true social cost when otherwise unemployed resources are drawn into use by the expenditure, there are a number of more specific conclusions which are corollaries to this general proposition:

- If the national unemployment rate exceeds the frictional minimum or if there is variation of occupational or regional unemployment around a national full employment rate, it is likely that more of all expenditures, public and private, can be justified than would be implied by the efficiency criterion using monetary benefit and cost estimates.
- If either of the idle resource conditions described above exist, the ranking of projects by the standard benefit-cost ratio using social value estimates would differ from the ranking which would occur if monetary estimates were used. Those expenditures, either public or private, which place heavy demands on occupational, industrial, and regional sectors showing idle resource rates above, the frictional minimum would rise in the ranking relative to those which place predominant demands on other sectors.
- If either of the idle resource conditions exist, the design of projects relying on social benefit-cost computations will differ from the design resulting from adoption of the full employment assumption. Those projects placing relatively heavy demands on occupational, industrial, and regional sectors showing high idle resources rates will be oversized relative both to their full employment design and to the scale of projects which place demands on other sectors. Moreover, all expenditures, public and private, which rely on social benefit-cost calculations for design, will make relatively heavier demands on occupations, industries, and regions showing relatively high idle resource rates than if the design criterion were based on the full employment assumption.
- Because of the implications of the above propositions, the problem of unemployment, regional stagnation, and high unemployment occupations and industries would tend to be eased by use of the social benefit-cost design criterion rather than the market cost design criterion based on the full employment assumption. This is so because unused resources are evaluated at a very low cost in the social benefit-cost criterion. Use of these resources is, consequently, encouraged.

As a word of caution, it should be emphasized that the results of this study should not be taken to imply that every public expenditure project which has been rejected because of an inadequate benefit-cost ratio should be undertaken when the unemployment rate rises above 4 percent. The conclusion to be drawn is that there is an operational framework by which to re-evaluate projects in terms of their opportunity costs when regional or national unemployment rates depart

from frictional minima.[12] Moreover, to avoid biasing public expenditures in the direction of a single program, all public investments (including tax cuts) should be similarly analyzed to determine what, if any, differences exist among them. A second warning concerns the extent of adjustment required in the benefit-cost ratio when otherwise unutilized resources are a part of monetary costs. As seen in the calculations of tables 3 and 4, the level of social costs typically falls only about 10 to 15 percent—at most 25 percent—even when the rate of unemployment is 8 to 9 percent, as it was in a number of regions in 1960. At a time when there is much expectation that the incorporation of "redevelopment benefits" or "secondary benefits" into benefit-cost analysis will lead to the justification of many projects not otherwise meeting the efficiency criterion, this conclusion should be sobering.

Only through substituting social opportunity costs for nominal monetary costs in the expenditure criterion can public decisionmakers isolate expenditures which are both intrinsically economic and substantial employment generators. Through such shadow pricing efforts, more discriminating judgment can be applied to public expenditure policy in general and especially to public expenditure policy in the chronically depressed, high unemployment, and declining areas of the Nation.*

[12] See Haveman and Krutilla, *op. cit.*

* Further discussion of this issue is found in the paper by McGuire in this volume.

Section C

Multiple Objectives and Regional Impacts in a Consistent Analytical Framework

PROJECT DESIGN AND EVALUATION WITH MULTIPLE OBJECTIVES

BY A. MYRICK FREEMAN III*

A. Myrick Freeman III is Assistant Professor of Economics at Bowdoin College.

Because the resources available to the Government are not unlimited, choices must be made among the many legitimate public sector objectives. If the benefits of all programs were commensurate and if they could all be valued in terms of some common unit of account, "the efficiency criterion [would be] adequate as a public expenditure criterion. It calls for the selection of those projects with positive net benefits (money valued benefits less costs)." Unfortunately, it is difficult to value some benefits and costs. The outputs of many government programs are not marketed and have no price; others manifest externalities or are not divisible into discrete units; still other outputs are not yet measurable even in physical terms. "One of the most pervasive of these project effects which is difficult to measure and evaluate concerns the redistributive or equity impacts of the project."

Professor Freeman outlines three approaches which have been suggested for choosing among alternative programs when the distribution of benefits and costs is a factor in determining project worth. Of the approaches, Professor Freeman favors the construction of a system of explicit weighting of benefits and costs. He argues that this procedure is likely to be more open and consistent than the others proposed. He discusses the difficulties of determining a meaningful set of weights. Attempts to extrapolate them from previous legislative expenditure decisions, for example, have encountered problems of inconsistencies and insufficient information.

Professor Freeman concludes that if weighting functions are to be used to assist government decisionmakers in choosing appropriately among projects with multiple objectives, "they must be based on consideration of legislative intent, guidelines established in law and precedent, and ultimately subjective evaluations by politically responsive decisionmakers."

I. *Introduction*

In an open, nondictatorial, and rationally governed society all government expenditures are made ultimately in pursuit of a single objective, the improvement of human welfare or well-being in that society. This lofty goal is above dispute. But since we have no direct measure of welfare it is also incapable of giving effective guidance to decisionmakers who are trying to make the most of a limited budget by choosing among alternative expenditure plans.

Decisionmakers must view their expenditures as producing not simply increments to welfare, but many separate kinds of benefits: flood control, education, transportation, cleaner air, more livable cities, et cetera. In these terms there is not one but a multiplicity of objectives

*I wish to express my appreciation to Professors Robert Haveman, Otto Eckstein, Martin McGuire, Thomas Hopkins, and William Shipman for helpful comments and criticisms of an earlier draft of this paper. Of course they bear no responsibility for remaining errors, nor do they necessarily endorse the views in this paper.

which are to be sought with government expenditures. Yet it is inappropriate for decisionmakers to try to maximize the amount of flood control plus the number of students educated, et cetera. We cannot have more flood control *and* more education and more of everything else. It is "either/or." Choices must be made. The identification of multiple objectives in itself does not provide adequate guidance to decisionmakers who must make these choices.

Neither of these views of objectives, the single, all-encompassing welfare goal or the multiple, noncommensurate objectives, can provide any practical basis for allocating limited government budgets among competing programs. The task of this paper is to explore the shape of this dilemma and to examine our progress in finding possible solutions at the practical level.[1]

The progress so far has been substantial, but this is because the easy steps have already been taken. We have developed fairly acceptable techniques for attaching money prices or market values to some kinds of benefits from government expenditures. Where this can be done for the principal benefits of a project, the efficiency criterion can be employed. By this criterion projects are evaluated on the basis of the difference between the money values of their benefits and costs. However for some kinds of benefits (and costs), outputs (and inputs), cannot be valued in money terms. Where benefits which are not susceptible to market valuation are important, the efficiency criterion is not adequate to the task of project evaluation. To use the efficiency criterion, one must assume that the unvalued effects of the project, including income redistribution, are not significant or do not matter. Of course in practice this is rarely true; and in principle it is never true.

If valuation techniques could be improved so that all kinds of benefits could receive meaningful price tags, there would be no problem of multiple objectives. Or if some higher order measure of value (welfare?) could be developed, into which all dollar values could be converted, and in terms of which all otherwise noncommensurable benefits could be measured, there would be no problem. In short the multiple objective problem exists because of limitations in our techniques for assigning values; the multiple objective problem is a valuation problem.

This valuation problem cannot really be avoided, at least if choices are to be made on anything but an arbitrary or random basis. In the literature two techniques have been proposed for making choices where there are both benefits which are subject to market valuation and benefits which are not. These will be described below. It will be shown that any choice made using either of these techniques implies a particular set of relative values for the noncomparable benefits. When these techniques are used, choices determine values rather than values determining choices. Thus the problem of valuation cannot really be avoided. My thesis is that since valuation is unavoidable, it is better to confront the choice of values openly and explicitly.

---

[1] This paper does not deal with the welfare theoretic basis of public expenditure analysis. For a good introduction and survey of this area see Otto Eckstein, "A Survey of Public Expenditure Criteria," in *Public Finance: Needs, Sources, and Utilization*, Universities-National Bureau Committee on Economic Research (Princeton: Princeton University Press, 1961), pp. 439–504.

II. Kinds of Benefits and Costs

A *benefit* is anything resulting from the activity of the government which is favorable. To be more specific, a benefit is a government produced economic good, where good is defined broadly to include all things tangible and intangible (e.g. services) which individuals desire. Finally desire is shown by a willingness to pay some amount of money rather than do without the benefit entirely. Similarly *costs* are any unfavorable consequences of government action. To increase one benefit it is necessary to decrease some other benefit (which is a cost) or withdraw more resources from the private sector (which is also a cost). While the remainder of the paper is addressed to the problems of noncommensurate benefits, the same principles and conclusions apply to costs.

Table I

Type of benefit	*Examples*
1. Marketed:	
(*a*) No spillovers	Municipal water.
(*b*) Spillovers	Treatment services for industrial wastes; the services of toll highways.
2. Not marketed:	
(*a*) Divisible and measurable	Public recreation; the services of toll-free highways; flood control.
(*b*) Public good and measurable	The distribution of income; public health services; water and air pollution abatement.
(*c*) Public good and nonmeasurable	National defense; the quality of life; basic research.

Benefits can be further classified according to several attributes. Table I lists the several classes of benefits discussed here and provides examples of each. Some benefits are sold by the Government. Markets are used to allocate them to those individuals with the highest willingness to pay, or to those who placed the highest valuation on them. If there are no spillover effects, the market price is the value of that benefit. Where there are spillovers or externalities, prices are not equal to values; there is market failure. The external effects must be taken into account in determining the value.

The nonmarketed benefits can be divided into three groups on the basis of two attributes. First the benefits may or may not be measurable in units other than those of the inputs required to produce it. A benefit is measurable if units, standards, and techniques have been developed to indicate and to express the quantity of the benefit and changes in this quantity. National defense is an example of an unmeasurable benefit.[2] Also the benefit may or may not be divisible. Benefits are divisible if one person can have more without simultaneously increasing the amounts going to all other individuals. Nondivisible benefits are usually called public goods.*

[2] The remainder of the paper deals only with measurable benefits. Until a type of benefit can be measured, values cannot be assigned to varying quantities of it. The ommission of nonmeasurable benefits from the remaining discussion does not mean that they are unimportant. It does mean that I consider measurement and valuation to be distinctly different processes.

*Further discussion of this issue is found in the papers by Steiner and Arrow in this volume.

One class of nonmarketed benefits is divisible and measurable. An example is outdoor recreation which can be measured in user days or particular activities such as wilderness hiking, or flat water boating. This class of benefits could be marketed, but presumably the consequences of marketing are not desirable. For example there could be substantial spillovers or externalities, or as in the case of open access highways, the costs of enforcing the exclusion principle may be considered to be too high.

The nondivisible or public good class of benefits presents a different set of problems. Neither private nor publicly supplied markets would perform satisfactorily since once the benefit was provided to one individual it would be freely available to all. Examples include clean air, clean water, and certain public health services. Some of these benefits are stated in terms of attributes of the economic, political, or social system, for example the degree of equity in the distribution of income.[3] At least these attributes can be measured. In addition there are the nonmeasurable benefits such as aesthetics, cultural or political stability, and other even less specific indicators of the quality of life.[4]

All of the types of benefits described here have value, if by value we mean a willingness to pay some amount of money by individuals.* The problem is to discover this value. In the private sector of a market economy there is an outlet for expressing willingness to pay. Values are determined in markets. And where there are no spillover effects or externalities, value or willingness to pay at the margin is equal to the market price.

The basis for valuation of goods in private markets is individuals' preferences.[5] It is usually accepted that individuals' preferences should also govern the values placed on things produced by the public sector as well. Since direct expression of willingness to pay for benefits through markets is not possible for individuals, valuation must be based on indirect evidence garnered from observations of market related behavior of individuals.

The art of valuing the benefits of public investments in reclamation, flood control, hydroelectric power, etc. consists of just this. In fact considerable progress has been made in developing acceptable techniques for inferring willingness to pay or value for several types of benefits. These techniques are bound to be improved and extended over time.

Let us define *money valued benefits* as those benefits for which monetary values having a basis in individuals' preferences are directly

[3] The inclusion of income distribution as a benefit for which individuals are willing to pay calls for explanation. Under plausible assumptions about individuals' utility functions, namely interdependence, it can be shown that individuals can increase their utility by contributing to Government enforced tax and transfer systems. See for example, Harold M. Hochman and James D. Rogers, "Pareto-Optimal Redistribution," *American Economic Review* (forthcoming), Edgar O. Olsen, "A Welfare Economic Evaluation of Public Housing," an unpublished doctoral dissertation, Rice University, 1968, or A. Myrick Freeman, "Income Distribution and Social Choice: A Pragmatic Approach," *Public Choice* (forthcoming).

[4] The fourth possible combination of the two attributes, divisible but nonmeasurable, is, I believe, an empty box.

[5] Technically, individual preferences along with income determine demand. Prices and values are determined by the interaction of demand and supply.

*Further discussion of this issue is found in the paper by Margolis in this volume.

observable in markets or can be inferred indirectly from market information.[6]

It is the remaining *unvalued benefits* and costs which constitute the multiple objective problem. Solution of the multiple objective problem requires a two pronged program of:

(*a*) Stepping up efforts in the area of estimating money valued benefits from information on individual preferences and market behavior; and

(*b*) for those types of benefits where this does not seem to be potentially fruitful, developing alternative forms of public investment criteria where noncommensurate benefits and costs are involved.

It is to this latter possibility we now turn.

III. Public Expenditure Criteria

If there are only money valued benefits and costs from a program, the efficiency criterion is adequate as a public expenditure criterion. It calls for the selection of those projects with positive net benefits (money valued benefits less costs).[7] The criterion also provides rules for finding that design for a particular project which will maximize its net money valued benefits.

The analyst has to estimate the quantities of benefits forthcoming from a project under alternative designs, the quantities of things being used up or reduced, and to utilize available information generated from markets or the market oriented behavior of individuals to determine the prices to be used. The final step is to use the prices to convert the physical measures to values, add them all together, and determine net benefits. The process can be repeated for alternative designs until the best possible design is chosen.[8]

Let there now be some benefits (or costs) from this project which cannot be valued.[9] It must be assumed that they are measurable, although not in dollar terms, and that the amounts of both money valued and unvalued benefits and costs are known. The efficiency criterion, maximum net money valued benefits, is inadequate since some benefits (and costs) are not included in or captured by that criterion. What alternative criteria can be employed? Three approaches to this problem have been offered in the literature.

The first has been advocated by McKean.[10] He suggests that the decision makers be provided with a schedule showing the net money valued benefits as well as descriptions of the other benefits (and costs, if appropriate) for each alternative project design that was analyzed.

---

[6] I have not used the conventional distinction between efficiency and nonefficiency benefits and costs. In practice "efficiency benefits" has meant money valued benefits as I have defined them; but I think the latter is a more descriptive term. I would prefer to restore efficiency to its more conventional meaning, of getting the most out of what is put in.

[7] If the resources which can be committed to public projects are limited by a political decision, it may not be possible to undertake all those projects with positive net benefits. I assume here that there is no budget constraint so that all opportunities for increasing the total value of output by redirecting resources from private to public uses will be utilized. Also for simplicity I am assuming that all problems of discounting, risk, uncertainty, and measurement can be ignored.

[8] Of course there are mathematical techniques available which greatly simplify the search for an optimum design.

[9] All of the following arguments and conclusions are similar for the cases of unvalued costs. The remainder of the paper will deal with the case of unvalued benefits only.

[10] Roland N. McKean, *Efficiency in Government through Systems Analysis* (New York: John Wiley & Sons, Inc., 1958). See especially pp. 127, 206–208, 240–242. His suggestion has been reformulated to conform with my terminology and framework.

The decision maker is to choose that alternative which conforms best to his subjective evaluations or his interpretation of society's preferences for or valuation of money valued benefits vs. other benefits.[11]

TABLE II

	Net money valued benefits	Measurable unvalued benefits
Project A:		
Design No. 1	$9,000	4 units.
Design No. 2	10,000	6 units.
Design No. 3	12,000	5 units.
Design No. 4	13,000	4 units.
Design No. 5	14,000	3 units.

Table II is a presentation of data on a hypothetical project in the format recommended by McKean. Net money valued benefits have been estimated for each of five alternate designs for the project. In addition the impact of the project on some unvalued benefit has been estimated for each design.[12] Of course this other benefit must be measurable.

The first design is clearly inferior to Nos. 2 and 3, the latter having more of both valued and unvalued benefits. The choice among the remaining designs cannot be made without some knowledge of the relative desirability or valuations of valued and unvalued benefits. McKean's suggestion is to rely on the decision maker to make the choice himself. Whichever design is actually chosen will provide a clue as to the relative values of valued and unvalued benefits which are held by the decision maker. For example, if he chose No. 3, this implies that one unit of unvalued benefit is worth at least $1,000 since this was the amount of money valued benefit he was willing to forego to obtain the fifth unit. But the value is less than $2,000 since by not choosing design No. 2 he showed that he was unwilling to sacrifice this much to obtain the sixth unit.

In evaluating this approach to the decision problem with noncommensurate benefits, we must question the ability of the decisionmaker to infer the society's valuation for unvalued benefits and to apply it consistently to a number of similar choices over time. In McKean's system the decision maker is not encouraged to think in terms of an explicit relative value to be applied to the available data, but rather to rely on his judgment after viewing the range of alternatives. Since each choice implies a relative price or a trade-off ratio between valued and unvalued benefits, the question is whether or not this ratio will be relatively stable over time and over a number of choices. We will return to this question at a later point.

Marglin has proposed a somewhat different technique for incorporating multiple objectives in a single expression for determining the

[11] I interpret Arthur Maass' discussion of the political determination of trade-off ratios between efficiency and redistribution as falling in this category. See his "Benefit-Cost Analysis: Its Relevancy to Public Investment Decisions," *Quarterly Journal of Economics*, LXXIX (May, 1966) pp. 208–226, Maass describes a three-step process wherein the agencies present data on alternative mixes of efficiency and redistribution, the Executive proposes one such mix as the optimum, and Congress accepts, rejects, or modifies it. Maass both argues that this is an accurate description of the actual decision process in several instances and urges that this process be adopted in other areas where multiple objectives are involved.

[12] These must be net benefits. If there are associated unvalued costs, they would be displayed in the same way.

optimum project design.[13] Briefly he suggests that some minimum level of one benefit be established and that the project be designed to maximize the other benefit subject to meeting the constraint.[14] There is no established rule for deciding which variables are to be constrained and which are to be maximized in the analysis. The choice might be made on the basis of analytical convenience. In the example of table II suppose that the decision has been made that unvalued benefits must be no less than 4 units. Design No. 4 maximizes money valued benefits subject to achieving at least four units of unvalued benefits. The same design would be optimal if valued benefits were constrained to be at least $13,000 and the objective was to maximize unvalued benefits. The interesting thing is that just as in the case of McKean's procedure, any choice of a design implies a certain relative valuation or trade-off ration between valued and unvalued benefits.[15] The choice of the minimum value for the constrained benefit determines the resulting relative valuation of the two benefits. Once again choice determines values.

Any choice among alternative designs or projects whether made by Marglin's or McKean's procedure logically implies a valuation of one type of benefit in terms of the other. The question is whether choice should be allowed to determine value in this way or should value be determined independently of any particular project analysis and used to make the choice. It seems to me that a good case can be made for encouraging the conscious, deliberate selection of weights by politically responsive decisionmakers. Support for this idea can be found in the economics literature.

Otto Eckstein has argued that economists could make an important contribution to policy making if they would " * * * interpret the desires of the policy people * * * and express them in an analytical form as an objective function," and get on with " * * * the establishment of decision-models which will reveal explicitly what actions will maximize the achievement of specified objectives." [16] Eckstein is referring to those cases where net money valued benefits do not measure all of the favorable effects of an expenditure program. If the preferences of the policy people can be expressed in terms of explicit weights or values of one form of benefit in terms of the other, then the consequences of this particular set of weights for the design and selection of projects can be clearly seen. More importantly, if these consequences are thought to be desirable, then the same set of weights or objective function can be used to achieve a consistent set of decisions concerning all proposed projects. Usually broadening the design and selection criterion to include unvalued benefits will result in the

[13] In Arthur Maass, Maynard M. Hufschmidt, and others, *Design of Water Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning* (Cambridge: Harvard University Press, 1962), pp. 62–86.

[14] Where there are "n" groups of noncommensurate benefits, minimum levels must be set for (n–1) of them and the remaining one benefit maximized subject to these (n–1) constraints.

[15] In mathematical terms, this is a problem in the calculus of constrained optima. The solution of the problem includes a value for the LaGrangian multiplier which is the marginal cost measured in units of one benefit of obtaining one additional unit of the other benefit.

[16] Otto Eckstein, "A Survey" *op. cit.*, p. 445. His paper also includes an excellent discussion relating his suggestions to conventional welfare economics. Marglin also suggests an assignment of weights, to different forms of benefits or values as an alternative to specifying minimum achievement levels. See Marglin, *op. cit.*, pp. 78–84.

choice of a different design as the optimum. This is true so long as the monetary and nonmonetary benefits are not strictly correlated.[17]

The use of weights in defining an objective function and selecting the optimum project design is illustrated with the example of table II. Suppose that decision makers conclude that society's relative preferences for valued and unvalued benefits can be approximated by attaching a relative weight of 1,500 to each unit of unvalued benefits. This is equivalent to saying that each unit is worth 1,500 units of money value benefits or $1,500. The object of project design is to maximize the sum of valued benefits plus the assigned monetary value of unvalued benefits. The total benefit or contribution to welfare for each design has been computed. They are given in table III.

TABLE III

	Weighted total benefits
Design No. 1	$15,000
Design No. 2	19,000
Design No. 3	19,500
Design No. 4	19,000
Design No. 5	18,500

On the basis of the weighted total of both valued and unvalued benefits, Design No. 3 is the optimum and should be chosen. But this is not surprising since it was shown above in the discussion of McKean's suggestions, that if the decision maker chose Design No. 3, this implied that each unit of unvalued benefit was valued at between $1,000 and $2,000. When the procedure is reversed and a value between $1,000 and $2,000 is placed on unvalued benefits, Design No. 3 is the optimum.

In this section we have outlined three approaches to choosing among alternative projects when benefits are not commensurate. These are McKean's discretionary approach, Marglin's minimum target levels approach, and the system of explicit weighting of benefits. We have shown that given the necessity for making choices among alternatives, relative values must be determined. The process of valuation can be implicit and can be obscured behind the preferences of a decision maker or the apparent objectivity of a predetermined minimum target. But valuation cannot be avoided. In the first two approaches the valuation is implicit and may vary as decisionmakers make many choices over a period of time. It is also not obvious to the members of the body politic what relative values are actually implied by the set of choices made. Political review is made difficult by the lack of information. If weights are assigned and an explicit objective function is formulated, the valuations are obvious, and the implications of different weights for project design and selection can be worked out. With value made explicit, subject to political review, decisions are more likely to reflect the general preferences of society, and less likely to be influenced by the pressures from special interest groups.[18] Also they are likely to be more consistent over a wider range of choices and a longer period of time.

[17] This has been proven in the case where repayment of costs by beneficiaries is required and redistribution is the unvalued benefit. See A. Myrick Freeman, "Income Redistribution and Planning for Public Expenditure," *American Economic Review*, LVII, No. 3 (June, 1967), pp. 495–508.

[18] I am ignoring all of those problems of voting and the political process familiar to both political scientists and economists. I assume that most of the political decisions and choices which are made reflect and are responsive to individuals' preferences, although imperfectly so. It seems that this way of making choices produces reasonably tolerable results, or to put it another way, most people seem to think that the costs of changing the system outweigh any gains in the form of improved decisions.

IV. The Choice of Weights

If one accepts the notion that the systematic assignment of weights or values to unvalued benefits as a guide to choice is to be preferred to the piecemeal, *ex poste* determination of weights as a consequence of choice, there is still the question of how these weights are to be determined. This seems to be largely a political or administrative matter, since by the nature of the problem market oriented information is not adequate to the task.

One approach, first suggested by Eckstein, is to look for clues in the past decisions on resource allocation and taxation.[19] In a study of Corps of Engineers projects, Haveman, following this suggestion, used an explicit set of weights derived from the marginal effective tax rates on personal income to value incomes received by different individuals.[20] And more recently Weisbrod has made imaginative use of expenditure data on water resource projects to solve for the implied weights given to incomes (benefits) to different groups.[21]

The logic of this approach has already been outlined. If a choice concerning any one project implies a set of weights for the several kinds of noncommensurate benefits, similar information from a number of projects with different characteristics should enable the analyst to infer the weights used in different situations. Yet, Haveman and others, myself included, have serious doubts as to the efficacy of using expenditure data to infer the weighting functions. One cause for doubt is that it does not seem reasonable to assume, as one must, that those ultimately making the decision, Members of Congress, had full knowledge of the magnitude and composition of benefits of all forms when they made their decision. It seems more likely that decisions were actually based on an incomplete understanding of the full range of effects of the project, and also that the decisions were influenced by other considerations such as political advantage and the efforts of special interest groups.[22] Haveman also doubts that the choices made by Congress have exhibited a stable and consistent pattern of weights over time.[23] At issue is not only the stability of congressional choices over time but the internal consistency of the set of choices made in any period of time. In viewing the relative weights given to, for example, money valued benefits and redistribution or equity, has Congress employed the same set of collective values in approving the oil depletion allowance, the various farm price support programs, Upward Bound, and Aid to Dependent Children?

Even if these questions of the congressional history of making informed and consistent choices among alternative packages of noncommensurable benefits could be answered affirmatively, there is still the question of whether the technical conditions will be met for deriving the weights from observed choices. If a stable weighting function

---

[19] Eckstein, *op. cit.*, pp. 447–448. "He may choose to use a form of the [objective] function that has been implicitly produced by the political process."

[20] R. H. Haveman, *Water Resources Investment and the Public Interest* (Nashville: Vanderbilt University Press, 1965).

[21] B. A. Weisbrod, "Income Redistribution Effects and Benefit-Cost Analysis," in S. B. Chase, ed. *Problems in Public Expenditure Analysis* (Washington: The Brookings Institution, 1968).

[22] See the comments on Weisbrod's paper by Mack and Haveman in *ibid.*, pp. 209–222; and Haveman's "Comment" on the previously cited paper by Maass in the *Quarterly Journal of Economics*, LXXX (November 1967), pp. 695–699.

[23] He raises this issue in both of the Comments cited in the preceding footnote.

exists and all congressional choices have been consistent with it, its properties cannot be inferred from these choices alone. In addition we need to know the terms on which one benefit can be exchanged for another in moving from one alternative design to another. Turning again to the example of table II, knowledge that design No. 3 was chosen was not sufficient to infer the relative weights given to valued and unvalued benefits. We needed to know the amounts of benefits for designs No. 2 and No. 4 so that we could determine the costs of the fifth and sixth units of unvalued benefits in terms of valued benefits. It is not a common practice for agencies to draw up alternative designs for given projects. Usually they settle on a single design early in the decision process. Choices are usually limited to "yes" or "no" for any one project. The transformation conditions among different types of benefits for alternative designs are not likely to be known nor can they be readily inferred from the data usually available for use in project evaluation.

While the likelihood of extracting information on the social welfare function from observations on expenditures seems to be quite low at present, it has been suggested by Haveman and others that examination of tax information might be fruitful at least for the relative valuations of money valued and redistribution benefits.[24]

For one thing, in contrast to expenditures, the redistributive aspects of tax programs are not complicated by the existence of efficiency gains attributable to the correction of market failures. The redistribution consequences of tax programs are reasonably clear, at least in the case of personal income taxation, and they are given considerable attention in the debate and decision.

However, this approach runs up against the same problem as use of expenditure data, the necessity for knowledge of the transformation conditions before the social welfare function can be inferred from the tax rate schedules. If taxes on incomes and transfer payments to low income persons have any effect on incentives and the supply of effort, and if there are any administrative or deadweight costs to implementing a tax program, the level of aggregate income is not independent of the level and structure of tax and transfer rates. Using the tax system to achieve redistribution benefits may have a cost in terms of lower aggregate income.[25] Unless this cost, i.e. the transformation conditions between money valued and redistribution benefits, is known, the relative weights attached to redistribution and money valued benefits cannot be inferred from the observed tax system.

An alternative approach to the valuation problem is to provide decisionmakers with specified weighting functions determined after study of the stated objectives of the programs and discussion of the factors which are usually considered in making choices. The decisionmakers could accept or reject any weighting function depending on whether they thought it was representative of their objectives. Weighting functions could be given trial runs on sets of projects already decided upon, using the same information available to planners at the time of the decision. The object would be to see if things would have been different if the weighting function had been used, and if so, would the weighting function have improved the pattern of choice

[24] Haveman's "Comment" on Weisbrod, *op. cit.*, p. 210.
[25] This is a case of an unvalued benefit, redistribution, having a monetary cost.

in the eyes of the decisionmakers. An exact reproduction of past decisions would neither be expected nor necessary as a test of the acceptability of the weighting function. The dry run might produce a set of projects which decisionmakers would agree was better than the original set, which would indicate the potential advantage of a more systematic approach to valuation. Or the trial run might select a list of projects thought to be suboptimal by decisionmakers, in which case the weighting function would be rejected as not adequately reflecting policy objectives.

One very interesting application of this approach has been developed by two economists formerly with the Economic Development Administration.[26] They noted that grant-in-aid projects for regions which are administered through EDA are to be approved on the basis of the twin criteria of efficiency and need, the first criterion is reflected in a benefit-cost ratio which captures those benefits which can be valued in monetary terms. The need criterion reflects a concern for obtaining greater equity in the interregional and interpersonal distribution of income. The authors also noted that high unemployment rates and low incomes relative to national averages have been established by Congress as the basis for allocating grant-in-aid funds among applicants.

They then showed how these multiple objectives can be combined explicitly in a single expression which shows the total value placed on a package of noncommensurate benefits. The benefit-cost ratio for each project was weighted by a factor which combined the unemployment and low income situation in that locality into a single measure of need.

This exercise by McGuire and Garn is valuable for two reasons. First, it shows that legislative guidelines stated in nonquantitative terms may be translatable into quantitative counterparts which can be used in a weighting function. Second, it shows that thinking about how decisions are made, what is important, and how the important things are related, may reveal some heretofore unnoticed but quite reasonable ways of relating the quantitative counterparts.*

The most vexing problems of multiple objectives revolve around the relationships between money valued benefits and income distribution.** The efficiency objective simply adds up all increments to income irrespective of who receives them. One way to express the distribution objective is that increments to income are valued differently depending on who receives them. The two objectives can be combined if a decision can be made concerning the weights to be attached to additional income to different individuals. The total value of a project would be the *weighted* sum of its incomes. Here alternative weighting functions could be presented to decisionmakers along with tables illustrating the implications of the weights so that they could indicate which one most closely corresponded to their judgment of society's relative valuations. In table IV seven alternative weighting functions are presented along with the values placed on additions to incomes to persons at various income levels. The first column gives the expressions used to calculate the weights where I is current income. The other four

[26] Martin C. McGuire and Harvey A. Garn, "An Experiment in the Integration of Equity and Efficiency Criteria in Project Selection," (mimeo).

*Further discussion of this issue is found in the paper by McGuire in this volume.

**Further discussion of this issue is found in the papers by Weisbrod, Bonnen, and Schmid in this volume, and Feldman in vol. 3 of this collection.

columns show the weights to be attached to $1 of additional income to individuals with current income of $1,000, $10,000, $100,000, and $1,000,000 respectively. The weighting functions have been normalized to give a weight of 1 to $1 of income to persons with $10,000 current income.

The reader may wish to "play the game" by trying to decide which function comes closest to reflecting his own views as to the weights society should place on increments to income generated by public expenditure. Consider a project the costs of which are covered by personal income taxes. Recall that more than half of all personal tax collections come from individuals with incomes of less than $10,000. Since some persons' incomes are reduced to pay for the project, would you be indifferent as to whether the benefits accrued to millionaires, persons getting $100,000 a year, or persons with only $1,000? If so then weighting function No. 4 reflects your judgment that redistribution should get no weight in valuing project benefits. If you think that no value should be placed on publicly generated incomes to millionaires, then No. 5 may be for you. And if you think that income to poor ($1,000 per year) persons is at least 30 times more valuable than income to middle class ($10,000) persons, then No. 1 may come closest to reflecting your value judgment.

V. Some Recommendations

Our discussion has proceeded heretofore as if there were some Solomon-like decisionmakers who judged all alternatives on the basis of available evidence, and pronounced judgment, after which there was no appeal. It has ignored any of the realities of the political decision process in the Federal Government. The suggestion that decisionmakers be provided with *a priori* specified welfare and weighting functions and be directed to make future decisions in accordance with these functions should not be interpreted as a prescription for all agencies, or for Congress itself. It is neither feasible nor desirable to adopt such a plan across the board. But I think it is possible to draw from this paper some suggestions which recognize the political and administration realities, yet can assist in making more rational choices where multiple objectives are concerned.

TABLE IV

Form of social welfare function: Marginal social welfare (MSW) =	Value placed on 1 extra dollar of income to a man with income of—			
	I=$1,000	I=$10,000	I=$100,000	I=$1,000,000
(1) $I^{-1.5}$	$31.62	$1	$0.03	$0.001
(2) $I^{-1.0}$	10.00	1	.10	.01
(3) $I^{-0.5}$	3.16	1	.32	.10
(4) I^{0}	1.00	1	1.00	1.00
(5) $e^{-\left(\frac{I}{10,000}\right)}$	2.46	1	.0001	Nil
(6) $\left(\frac{I}{5,000}+1\right)^{-1}$	2.50	1	.14	.01
(7) $\left(\frac{I}{5,000}+1\right)^{-2}$	6.25	1	.02	.0002

First let me state what I think two of the important political realities are. Congress is not equipped for, nor is it likely to wish to involve itself in, consideration of a social welfare function or weighting function for particular programs. Also social welfare functions are not likely to be able to play any role in major one-shot decisions such as the shale oil development program, or the antiballistic missile system (thick or thin). This suggests that explorations of social welfare functions are more feasible and more likely in the executive branch and in areas where less is at stake on any one decision and decisions are more repetitive. The above cited work on EDA grant-in-aid programs is a case in point.

My recommendation is that executive agencies be encouraged to spell out the objectives of their own programs in terms of valued and unvalued benefits and to carry out studies similar to the McGuire-Garn work on EDA programs in an attempt to specify acceptable relative valuations for noncommensurate benefits. As their experience and self-awareness grows they should be encouraged to use their valuations as the basis for discretionary decisions. However, there should be continual review by a higher authority both of the weighting functions used and the decisions made. Approval of the Bureau of the Budget should be required before any such weighting system is implemented or is altered. This is necessary to assure that weighting functions and relative valuations are consistent among agencies rather than contradictory. Also it is necessary to curb a natural tendency on the part of agencies to assign relatively high weights to those things that it is best at producing, thus assuring a continued high level of authorizations and appropriations.[27] The Bureau of the Budget should take the initiative in establishing guidelines, and in selecting weighting functions for unvalued benefits produced by several agencies, such as redistribution.[28]

Congress itself could take the initiative when establishing a new program by stating its intent as clearly as possible so that this could be translated into a weighting function by the agency for its use in administering the program.[29]

This effort could parallel present efforts to implement PPBS and cost effectiveness analysis in more executive agencies. These new types of analysis force agencies to state their targets and to look carefully at alternative ways of achieving them. But it also naturally leads the agencies into consideration of the current set of targets itself, and the possibility that the current targets may not be consistent with the agencies' own estimates of the relative values to be placed on achievement of various targets. The implementation of PPBS may make it easier to begin explicit consideration of the implicit valuations placed on unvalued benefits and to gain acceptance of the principle that policy should be determined by one's objectives.

---

[27] I am indebted to Robert Haveman for emphasizing this point.

[28] This recommendation bears a strong family resemblance to Maass three-step description. See Maass "Benefit-Cost Analysis," *op. cit.* The main differences are that I think more progress is likely to be made in spelling out objective functions in existing programs and within executive agencies while Maass in his example suggests that Congress plays a larger role and that the procedure is primarily applicable to new programs.

[29] Again see McGuire and Garn, *op. cit.*

VI. Conclusions

The main points I have tried to make can be stated very briefly.

1. Choices among different kinds of benefits are unavoidable.

2. There is a one-to-one relationship between choices and relative values for different benefits. If values are known, choices are determined; if choices are made, values are determined.

3. Where values are not generated by market behavior, I have urged the systematic consideration of values in the form of explicit weighting or social welfare functions in order to obtain a consistent set of decisions on projects.

4. We are not likely to be able to infer weighting functions from observed choices among public expenditure programs, both because we cannot be sure that the same weighting function was used to make all the choices included in the sample, and because the required information is not likely to be available.

5. Therefore if weighting functions are to be used, they must be based on consideration of legislative intent, guidelines established in law and precedent, and ultimately subjective evaluations by politically responsive decision makers.

6. The procedures outlined here are most likely to be successful if the initiative comes from executive agencies, and in areas of established programs and repeated decisions, rather than if Congress is deeply involved in the process or if it is first applied to new programs or major one-shot decisions.

EFFECTIVE PUBLIC POLICY AND THE GOVERNMENT BUDGET: A UNIFORM TREATMENT OF PUBLIC EXPENDITURES AND PUBLIC RULES

BY A. ALLAN SCHMID*

A. Allan Schmid is Professor of Agricultural Economics at Michigan State University. He is currently on leave with the Systems Analysis Group of the Department of the Army (Civil Functions).

Professor Schmid discusses the relevance of benefit-cost type analysis to government rule-making decisions (e.g., "market bargaining and contract rules, permits, and licenses, zoning, prohibitions and quotas"). He points out that benefit-cost analysis is as necessary for the framing of government rules as in the evaluation of public expenditure decisions, for "the issuance of a rule also directs the use of resources which have alternative employment. * * * The test is the same for both budget outlays and rules—namely, is the value of the resources in a new use worth more than the alternative uses foregone?" Professor Schmid discusses both the efficiency impacts and the incidence of benefits and costs of public decisions and emphasizes the possible redistributive effects of rule as well as spending decisions.

Insofar as both rule-making and expenditure decisions have similar resource allocation and equity effects, there should be a common framework through which both kinds of impacts of both kinds of decisions can be evaluated. Professor Schmid proposes the formation of an economic budget which would display these relevant variables in a meaningful way. "Systematic treatment of the relationship between expenditure and rule-making decisions is one of the major unresolved issues and next steps in PPBS."

I. *Introduction*

In order to have an analytic system that considers the full range of alternative ways to get goods or service produced for the public, it is appropriate to look not only at the range of public spending alternatives, but also at the police power and rulemaking alternatives. In some cases these two sets of institutional arrangements are complementary to each other and in others they are substitutes. The rulemaking decisions referred to here are the broad category of Government action including market bargaining and contract rules, permits and licenses, zoning, prohibitions, and quotas.

This paper will discuss how spending and budget matters can be systematically related to rule and control matters. Involved in this is how the program and output of a department like the Justice Department which is a relatively low budget, rule administering agency can be related to those which are primarily spending agencies with big budgets. Focus on this issue is relevant not only for Federal policy decisions, but also for the connection between Federal and local governments, the latter possessing much of the police power. The Govern-

* The helpful comments of Jim Tozzi and Steven Dola are greatfully acknowledged.

ment is interested in improved resource allocation which will increase national income and this is treated in part II of this paper. Government is also interested in changing the distribution of the ownership of wealth which is treated in part III.

Therefore, this paper will make these objectives of income increase and its distribution explicit in analyzing spending and rule decisions. Systematic treatment of the relationship between expenditure and rulemaking decisions is one of the major unresolved issues and next steps in PPBS.

All are familiar with the presentation of spending alternatives in PPBS and benefit-cost analysis. A Government program such as flood control reservoirs or disease control employs resources which have alternative uses. The value of these alternative products foregone is the opportunity cost of the Government project to be compared with the value of the project output to the public. Government rulemaking is usually analyzed outside of the above formulations. Yet, the issuance of a rule also directs the use of resources which have alternative employment. Can we then conceive of a benefit-cost ratio for a rule change as well as for an item in the Federal budget?

As other papers have pointed out, the Government acts as agent for groups of people who cannot make their demands for the production of certain goods effective through market bargaining alone. The bid of the representative government for the output of a public project means that resources are allocated differently than they would have been in the absence of the bid. The bid based on tax money can be thought of as an order or command for the allocation of resources made legitimate by the public's representatives.

A rule also directs the use of resources. While it may short circuit the usual order contained in a money bid, the function is the same. A rule orders that resources be used in a certain way for the production of services which its users consider more valuable than the alternative foregone. Again this is made effective and legitimate by the public representatives. The test is the same for both budget outlays and rules—namely, is the value of the resources in a new use worth more than the alternative uses foregone.

II. Spending and Rules in Resource Allocation

Relation between spending and rules: flood control illustration

Perhaps an illustration would be useful at this point. Take the case of people subject to flood damage. One alternative is to build a reservoir to control the water. There may be problems for the potential users to express their demand for this service in the market. Therefore, we find flood control reservoirs in the Federal budget. Systems of analysis have been developed to compare various flood control projects and other water development projects.

There are, however, alternative ways to allocate resources to accomplish a similar service. For example, transportation resources can be substituted for location and protection in the flood plain. Economic activity could be organized outside the flood plain. Assume a simple case where a particular set of industries finds that transportation costs are cheapest with a river bank location. The cost of a nonflood plain location then is the extra transport cost associated with it.

The economics of the reservoir compared with alternative locations of industrial activity turn on their costs. We have the cost of the reservoir construction resources on the one hand and the cost of the transport resources on the other. If the cost of transport exceeds the cost of the reservoir the firms should locate in the flood plain and petition the Government to buy the reservoir for them with their tax money. This would result in a net profit for the firms and a gain for the economy.

But, just as there is a problem in expressing demand for certain goods and services such as those produced by reservoirs, there may be problems in organizing industry off the flood plain. Acting as individuals, the managers of these firms may not account for all costs. If they were one giant firm, the above economics of reservoir versus alternative industrial location should be clear. But acting alone, some firms, for example, may be unaware of the flood risk and thus locate next to the river. If enough do this, the whole economics of location changes and other firms will find it advantageous to locate next to them in spite of the flood risk. If all firms acted together at once this uneconomic dynamics would not develop. The method to institute this demand for the desired allocation of resources may therefore not be market bids, but rather take the form of a zoning law prohibiting anyone from locating in the flood plain. The economics of the zoning rule, however, turns on the same type of analysis used for evaluation of the reservoir. If the allocation of resources to overcoming the disadvantage of a nonriverbank location is less than protecting the flood plain location, the rule is superior to the reservoir and represents the optimum direction of resource use.

RELATION BETWEEN SPENDING AND RULES: WILD RIVER ILLUSTRATION

The above illustrates only one way in which a rule and spending are alternatives. Consider the use of a stream as a wild river. A proposed law to create a wild river is not now likely to be in the same analytic system as that for various water development projects. However, it is conceptually possible to put this rule into the benefit-cost framework. Perhaps some hypothetical numbers could make this clear. Assume that it costs a set of industries $1,000 more in transportation to locate out of the flood plain than in it and that this is the only relevant cost difference. This $1,000 then represents the potential benefit of a flood control project which would allow the industries to locate on the flood plain. Further, assume it is possible to obtain protection with some combination of reservoirs, levees, and flood proofing at a cost of $400. This would produce a ratio of benefit to cost for the Government spending project of $1,000/$400 or 2.5/1. The $400 project cost includes the cost of any reservoir land site. In this situation, the Government acting as a collection agent for the industries could afford to pay $400 for the flood control works and also up to $600 more for any other costs that might have to be covered.

Assume that the reservoir destroys certain uses of the river that could be made in its wild state and that this has a benefit worth $700 in total to all wild users. However, the consumers of this wild river product may have difficulty expressing their demand for this use in the market. Therefore, the agents of the industries may not

be able to see all the opportunities foregone by their reservoir. They may only see the $400 construction cost and the $1,000 benefit and conclude they have a sound project.

For simplicity, assume the land has no alternative uses except for reservoirs or wild uses. It was noted above that the industrial group through Government could afford to pay up to $600 for any damage or extra land needed to construct the reservoir. The wild users as a group would, through Government, be willing to make a bid of $700 for the affected land. This would then appear as a cost of the flood control project as follows: 1,000/400+700 and the benefit-cost ratio would be less than one and uneconomic.

Considering the wild river as a project, it would have benefits of $700 and costs equal to the industrial alternative foregone of $600 and thus a favorable benefit-cost ratio of $700/$600. Rules sometimes short circuit the Government bidding process and there may simply be a rule banning all reservoirs on the stream. The wild river users wouldn't have to pay the $700 they would be willing to pay for the benefits. This involves a redistribution of income in their favor which will be examined later. But, the economics of resource use remains the same with the wild river as the better "project."

Rules are often ambiguous and may or not be redistributive. The land in this case may already be owned by the users of the wild river. In that case, they have a property worth $700 to them for which the industrial users would pay only $600. Obviously, this ratio says they shouldn't sell. Yet, we may be back to our group demand problem. Some few individual owners may be tempted to sell their rights which would destroy the value of others' property. To prevent this, the owners may wish to zone the whole area as a wild river. In effect they are saying they can't imagine a total bid exceeding $700, the value to them in the wild state, and therefore as a group they reject all bids and other users need not bother making bids. The economics of the wild river rule still turns on the value of the benefit compared to the values of the alternative uses foregone. In this case, there is no transfer and no relevant cost to be charged to any public budget.

III. Spending and Rules in Redistribution of Income

Incidence of benefits and costs

Under current procedures, the benefits and costs of a flood plain zoning law and spending for a dam have quite different incidence, though they could be designed to be identical. If the dam is paid for by all taxpayers and extra transport costs of nonflood plain location are paid only by specific industrial firms, the incidence of costs are quite different. One of the reasons that these considerations are not now resolved and incorporatd into systematic analysis is their complexity. Government spending and rules represent not only the direction of resources use by their acknowledged owners, but also sometimes a redistribution of that ownership. Separation of these factors is difficult. Many Government spending projects are financed by taxes and by user charges. The spending decision correctly involves total cost regardless of source. However, it may not be clear if there is also a redistribution of wealth involved. This is difficult to tell on a project by project basis and often even in total. Therefore the spending project

may or may not involve a redistribution and, while the project analysis tells the return on the total investment, a separate analysis is needed to indicate how the public feels about transferring resources from present owners to project beneficiaries. There will be further discussion of this below.

Consider the rulemaking decision. It too can be double barreled. We have already noted that a rule such as zoning to achieve a wild river may, if the land is already owned and accessible by the wild river beneficiaries, simply involve a decision to retain or sell. If retained, the opportunity cost of refusing to entertain bids from other users is borne by the wild river users who own the land.

However, the wild river users may not own the relevant land or have hired its use. A zoning rule may make the owners of the land unable to get bids from potential industrial users and unable to be reimbursed by the wild river users though the latter would be willing to pay. This would involve a redistribution which in effect taxes the landowners of a portion of their potential values and gives it to the wild river users. A separate analysis is needed to determine the economics of this transfer. In short, spending projects and rules which involve reallocation and recombination of resources by their owners—whether collected by user fees or taxes—can be grouped and compared. On the other hand, spending projects and rules which involve a redistribution of ownership and one-way transfers must have their test made in a different framework.

SYSTEMATIC ANALYSIS OF REDISTRIBUTION*

A suggestion on what a systematic redistribution framework might look like can be seen if we start with a tentative public objective to make one-way transfers of ownership (income). This is what is usually meant by redistribution of income or it might also be called a grant. For simplicity, this discussion will consider only projects whose sole output benefits a target who may pay only a portion of the cost. From the viewpoint of the general taxpayer, redistribution is the difference between project cost and the amount paid by project target beneficiaries, whether in taxes or user charges. This assumes that, after netting out all payments and benefits of other public programs, target beneficiaries paid less than the cost of the projects under consideration.[1] However, a given transfer from the grantors may or may not be received as income by the target beneficiaries. This is affected by the productivity of the investment.

For example, the general taxpayers may contribute the total cost of a given project (say $100), but because of low productivity, the target beneficiaries receive only $85. The beneficiaries are better off by $85, but $15 of the $100 transfer is wasted. If this represents the best investment available, the beneficiaries would have been better off if the $100 transfer had been made in cash.

---

*Further discussion of this issue is found in the papers by Weisbrod, Bonnen, and Freeman, in this volume, and Feldman in vol. 3 of this collection.

[1] A technical note on this conception which differs from some of the literature on this subject is attached at end of this paper.

It is sometimes suggested that the transfer be measured by the difference between returns to beneficiaries and their contribution to cost. In the above case this would be $85-0=85. Yet, this would not call attention to the fact that the taxpayers transferred and gave up $100 and, while they intended it to go to the target beneficiaries, $15 of it was wasted.

A grant can be given as cash or in the form of a particular project. The alternatives are that the recipients might invest (or consume) it as they wish, or the Government may invest it for them in projects from reservoirs to urban renewal. The first thing decisionmakers need to know is the size of the grant involved, and then whether it generates as much income to the target group in the form of a given project as it would in other alternative projects or cash. If it does not, the grant-in-kind in the given project is wasteful (assuming indifference on the part of the recipient to the form).

While the objective may be to give a grant and to achieve the maximum resulting change in the target group income, this latter calculation in no way prices or evaluates the desirability of the grant itself. The size of the grant that taxpayers wish to make to a target group must be an expression of general public values communicated to public representatives. For example, the public may be willing to give grants over a period of years with the objective of raising all incomes in the United States to $3,000. Whatever budget is available for grants, a particular investment competes with other project alternatives in terms of the productivity of return.

The taxpayers don't want to spend more than they have to in achieving $3,000 for everyone. In fact until it is known just how much transfer is necessary to achieve this, the objective may be a tentative one, or at least the time schedule for its achievement would be uncertain.

In current policy terms the amount of Federal cost share of projects paid by taxpayers, who receive less in public investment than their share of tax payments, is a major vehicle of redistribution. The amount involved depends on general public policy and is not the function of a particular project analysis. The project analysis only indicates what the productivity of the grant is in the form of a particular investment. This knowledge would be masked if analysts or Congress attempted to put weights on the benefits received by target beneficiaries.

In short, there are no redistribution benefits to be added to other categories of project benefits. All that is relevant is for the public to ask itself if a dollar of its assets granted to others is what they want and to communicate this to their public representatives. The project analysis can then indicate which projects are most productive for investment of this grant or whether it should be transferred in cash. If a redistribution objective is adopted, it should be implemented systematically with all kinds of public projects and programs competing in terms of productivity to the specified target groups and not piecemeal on a hit-and-miss individual project basis. Even within a given product field, care must be taken that special ad hoc arguments are not made for a particular project because of its impact on a certain group of worthy beneficiaries when other projects may produce more for the target group.

IV. Recapitulation in Terms of Illustrative Accounts

The above discussion can be summarized and further illustrated in terms of an account or bookkeeping system. Consider a list of survey reports which show preliminary benefits and cost estimates for water development. Some suboptimization has already taken place. In this illustration the optimum development for each *site* is an item in the system for selecting new starts. So the array of "projects" and benefit-cost ratios might include some which are Federal spending projects and some which are rule projects. An account designed to show national income gain might look like this:

TABLE 1.—NATIONAL INCOME GAIN ACCOUNT

Project	Benefit/cost	Total cost	Cost relevant to Federal budget constraint	Cumulative cost relevant to Federal budget constraint
1. Reservoir on river A	4/1	$10,000	$10,000	$10,000
2. Zoning rule, river C	3/1	10,000	0	10,000
3. Reservoir on river W	2/1	10,000	10,000	20,000
4. Reservoir on river Y	1.5/1	10,000	10,000	30,000

Such an array would indicate first of all that preliminary survey money which produces a zoning project discovers opportunities for improving the economy just as those which result in public spending projects. In fact, the table shows that in retrospect if the survey budget were limited to two surveys, projects No. 1 and No. 2 (with No. 2 being a rule "project") should have had priority over projects No. 3 and No. 4. The practical problem at the current time is that construction agencies tend to prefer survey efforts which produce spending projects to surveys that do not.[2] In part, this is because they get no credit from anyone for nonspending recommendations. In fact, it may work the other way with the field offices of a given agency being commended if survey costs are low relative to construction spending, but criticized if planning costs are relatively high. This would be less likely to happen if accounts were kept as in the above table. There are many unresolved issues in systematic management of programs that require substantial project surveys, but the direction of improvement lies in the above approach.

Now consider the new-start analysis.[3] Table 1 shows that if the Federal budget constraint were $20,000, then the reservoir projects A and W and the rule project of river C should be recommended. Since the rule project has no Federal cost—although their is a private cost—it should be recommended for local government implementation even if the budget constraint were $10,000 since it is a net gain for the economy which is not limited by the Federal budget constraint that happens to be in force. It is assumed here that the costs are borne by the beneficiaries with no demand on any public treasury.

[2] Also the action of congressional committees that control rule changes may not be coordinated with those committees that control spending.

[3] For simplicity, the fact that some projects require a detailed planning and design stage and others do not is ignored. Also ignored is the fact that the rule change may also require detailed planning and design before it is ready to function as is the case for construction projects.

If we turn to the objective of redistribution, a separate analysis is relevant. This might be termed the "Grants" or "Redistribution" or "Transfer account." This account is aimed at some target beneficiaries determined by Congress. Therefore, only projects from table 1 that benefit the target group are included in the "Grants account" in table 2. (In the hypothetical cases here assume only reservoir project, river W from table 1 aids the target group.) For simplicity assume all costs are paid by the public and not by the target beneficiary who receives all benefits. In practice only a portion of most projects would be redistributive.

TABLE 2.—GRANTS ACCOUNT (REDISTRIBUTION AND TRANSFERS)

Project	Benefit to cost	Transfer and total cost	Cumulative cost
1. Health project X	5/1	$10,000	$10,000
2. Reservoir, river W	2/1	10,000	20,000
3. Wild river rule B (700/600)	1.1/1	600	20,600
4. Cash		(1)	[2] 40,000
5. Reservoir, river N	.8/1	10,000	

[1] Dollar for dollar.
[2] Or to limit of budget constraint.

Assume that the public acting through Congress has decided that an amount up to $40,000 would give the public who taxed itself more satisfaction if transferred to a specified group than if it is consumed or invested by its owners.[4]

If table 2 is the list of available projects that aid the target group, then the $40,000 grants budget should be spent for projects 1 through 4. $20,600 would be used for investments, and $19,400 would be transferred in cash since the next best investment project would transfer less to the target group than the grantors give up.

Wild river rule B would, therefore, be enacted. Assume the target beneficiaries do not now own the relevant lands and because of poverty have no effective demand. In effect, the rule forces the current landowners to forgo bids from industrial users of $600 (using portions of the case developed earlier in part II). This is the same as taxing them $600 and giving a gift to the wild river beneficiaries. This equivalent of a tax of $600 and transfer of $600 may not show in current Federal budgets, but the effect is as described nevertheless.[5] If the group that wants to make a transfer is in fact these landowners, this rule would be an efficient transfer.

If the landowners are not the relevant grantors, then the public can reimburse the private landowners for the foregone $600 bid from industry that wants to build flood control works. Since the benefits obtained for the noncontributing target group of wild river users are worth $700, this is an efficient transfer. In this case, the $600 investment would show in the Federal budget as now constituted. If the wild river users were not the objects of transfers, but wished to tax themselves to preserve the river, they would express their demand through Government. In that case, project B would show in the table 1, Na-

[4] To be perfectly symmetrical this entire grants budget might be regarded as a project and put on table 1 showing that it has a benefit-to-cost ratio at least greater than 1. But this is not too helpful in an *ex ante* decisionmaking sense.

[5] This calculation might be added to the Department of the Treasury's *Comparison of Budget Outlays and Tax Expenditures by Function*, presented to hearings before the Joint Economic Committee, Jan. 17, 1969, *The 1969 Economic Report of the President* (Washington), Government Printing Office, pp. 11–31.

tional Income Gain Account, with a Federal budget cost of $600 necessary to reimburse the landowners. Table 1 includes only projects for which there is effective demand backed by owned income.

The relationship between the national income gain account and the redistribution account can now be spelled out in more detail. It is possible that a project with a benefit-to-cost ratio greater than 1 in the table 1 national income account may not be built because of a budget constraint on the types of projects in the Government program represented by the table. This rejected project may have a better benefit-to-cost ratio than one accepted in the redistribution account. To illustrate this, assume that the budget of table 1 is limited for some reason to $10,000 and only the projects on rivers A and C were selected with W and Y rejected.

Further, assume that the public communicates to their representatives that the redistribution account budget constraint should be $20,600. This would mean that the project on river W would be built although rejected in terms of the national income account budget. Also, under the wild river rule, river B would be implemented because it helps the target group even though its benefit-to-cost ratio of 1.1 : 1 is less than the project on river Y which has a benefit-to-cost ratio of 1.5 : 1. River Y project was rejected in the national income account because of a budget constraint and not accepted in the redistribution account because it does not aid the target group.

If public decisionmakers could rely completely on these program analyses as a guide to budget size, they should expand the national income accounts budget to include river Y and then they would never miss a project that added more to national income than it cost.

This does not show, however, that the redistribution budget is too large relative to the national income budget. The public may wish to transfer wealth (and give up the opportunity represented by project W and other alternatives including consumption) because they derive more satisfaction through transferral than through their own consumption. The return of project Y given up will certainly be a factor in the public's opinion on whether they want to transfer part of their wealth to others, but there is nothing inherent in the productivity of project Y which automatically invalidates their demand to be charitable. However, if the benefits of project Y could be collected and given to the target group this would be superior to project B. However, in this illustration it was assumed that this was not the case and only project W from table 1 would benefit the target group.

V. Uses of a Systematic Combination of Expenditure and Public Rules Accounts*

Some further issues illuminated by systematic analysis

The water field can be used further to illustrate issues in the mix of Federal spending and rulemaking decisions. If we look forward to the day when all agencies affecting a given product or service are grouped together for budget analysis, it will be useful to have a PPB system that encompasses spending and rules. In addition to the reser-

*Further discussion of this issue is found in the papers by Schultze, and Mushkin & Cotton in this volume.

voir spending and zoning rules there is now flood insurance under the administration of HUD which is not one of the traditional water agencies. Insurance schemes also have their opportunity cost and benefits. Much work remains to incorporate this into PPB so that the complete effort in a given field can be totaled.

Next consider the 1966 Executive order directing increased attention to the location of Federal installations in the flood plain. The costs of implementing this will show up in the budgets of every Department that has extensive building facilities. The Federal investment in flood damage reduction will not be complete until this is systematically accounted for.

Another area of increasing interest is that of interstate compacts often involving the Federal Government as a partner. Negotiation of these compacts involves not only the traditional water agencies but also the Justice Department. The compacts are ratified in the Senate, for example, by the Judiciary Committee and not the Public Works Committee which examines spending projects. These compacts are going to have a great effect on the ability of non-Federal units to pay for improvements that have primarily a local effect. At the present time, the Federal Government pays for some projects of relatively local effect because there is no procedure for facilitating agreement among the local government units on their relative cost shares. So this is a case where a rule change may have a great impact on demands for the Federal budget.

State pollution control rules may affect whether an expected recreation benefit on a Federal reservoir will be realized. Here rules and spending are complements.

Various agencies administer licenses and permits. In the water field, for example, the FPC licenses hydrodams, AEC licenses nuclear powerplants, and the Corps of Engineers issues permits for private dredging. To take only one dimension, each of these can have an effect on water quality. Each of them directs resource use in a certain way which has benefits and opportunity costs. Each of them in some respect is an alternative to Federal spending such as that for municipal sewage treatment plant grants or for reservoirs that provide low-flood augmentation. A rational decisionmaking system must somehow encompass all of these or the Nation will be over investing in one area while there are cheaper substitutes or possibilities of greater output if certain complementary rules and projects are combined.

OTHER EXAMPLES OF SPENDING AND RULE ALTERNATIVES

Water examples have been chosen to illustrate the general problem of relating spending and rule decisions. To further illustrate the kinds of questions involved, several other fields will be briefly explored. Much of Government activity is concerned with formulating rules of market behavior. The Department of Justice spends $8.2 million to enforce competitive behavior.[6] To illustrate the potential relationships, consider the competitive rules for railroads and spending programs of the Department of Transportation. The public may want to obtain a certain performance in the railroad industry. In certain contexts this

[6] *Special Analyses, Budget of the United States, 1970* (Washington), Government Printing Office, 1969, p. 262.

might be secured by having the Department of Justice follow a certain antitrust and merger policy.[7] An alternative—or complement—might be a grant or subsidy program offered to encourage the same performance administered by the Department of Transportation. Somehow these alternatives must be related and shown together in an information system.

Another relationship of spending and rules in the regulatory field is the use of Government enterprise as a yardstick to stimulate private performance. The public spending project has not only commodity-producing benefits but it may test and demonstrate new efficiences possible for other firms. A direct regulatory order could accomplish the same thing but for various reasons it may not be possible. A decision system encompassing direct regulation and spending yardsticks would be useful, but many unresolved issues remain.

There is great interest now in rebuilding our cities. Among the many alternatives are such things as direct Federal spending for urban clearance and renewal. A nonspending alternative—or complement—is reform of the property tax rules. New tax systems could be designed to give greater encouragement to private owners to improve their properties. This is an illustration of how State-Federal relationships are important in relating spending and rule decisions. Property tax rules are matters of State and local control, but nevertheless, failure to incorporate these alternatives into systematic analysis has great impact on demands for the Federal budget.

CONCLUSIONS

1. The argument has been presented that both public spending and rulemaking decisions produce benefits and have opportunity costs, and thus can be compared and ranked together as alternatives in a PPB system. While unresolved problems remain, there appear to be strong possibilities for improved systematic analysis.

2. Both spending and rules may involve redistribution of the ownership of wealth. Care must be taken to determine whether a given spending or rule project is designed to increase national income or transfer ownership and to see that the appropriate analysis is made for each. The benefit-ond-cost incidence of expenditures and rules must be clearly spelled out so that decisions may be properly accounted for.

3. The display of spending and rule projects in a single information system gives credit to survey investigations which produce no Federal spending for construction. While availability of information does not insure its use, this is the first step in avoiding a construction bias.

4. Explicit and systematic consideration of spending and rule decisions illuminates some of the connection between Federal and State and local decisions. Federal spending and more local level rulemaking are often substitutes and in some cases if local rules don't complement the Federal spending project the potential benefits won't be realized.

5. Combination of spending and rulemaking activities in a single information system will facilitate the totaling of Federal activity in a given field regardless of the particular department it happens to be in.

[7] It is interesting to note that a research study on railroad mergers has recently been made, not by Justice, but by the Department of Transportation, *Western Railroad Mergers* (Washington), January, 1969.

Technical Notes on Redistribution

A. The computation of redistribution is slightly more complicated if the taxpayer who wishes to make a transfer also receives part of the benefits of an indivisible project. In that case, redistribution is the difference between the target beneficiaries' actual contribution to cost and the contribution they would pay if total costs were shared in the same proportion as benefits are shared. For example, assume a general taxpayer group (G) and a target beneficiary (B) and a project as follows:

$$\$60_G + 60_B = 120$$
$$\$60_G + 40_B = 100$$

If costs were shared in the same proportion as benefits, the target beneficiary would have paid one-half of the cost or $50, instead of the $40 actually paid. The difference, $50−$40=$10, is the amount of redistribution or transfer.

If the total benefit-cost ratio is less than one, the amount of the transfer plus the target beneficiaries' contribution will be more than the beneficiary receives in benefits and the project would be wasteful. The beneficiary would be better off to retain his own contribution and to take the general taxpayers' transfer in cash.

The U.S. national income accounts, as now kept, assume Government transfers have a benefit-to-cost ratio of unity. When this is the case, the definition of redistribution outlined above gives the same result as that sometimes defined in the literature as the difference between target group benefits and their contribution to costs. This latter concept is appropriate when considering redistribution from the recipient point of view, but it gives the wrong answer from the taxpayer-grantor point of view when the benefit-to-cost ratio is greater or less than unity.

B. There is a technical problem in valuing the contribution of the taxpayer who wishes to transfer income to a target group. Assume a project with a ratio of benefits to costs $150/$100 with all benefits going to a target group and all costs paid by the general taxpayer. In a sense, if an investment of $100 could earn $150, then those who own the $100 are really transferring an asset whose present worth at market rates is $150 in a certain investment. If the $150 is a project return to a certain target group, the payer of the $100 could have offered it as a loan and bargained for a share of the net profit and thus it might be said that the transfer is what the beneficiary received minus what he paid or $150—0=$150.

If the taxpayer consumes the $100, this means his consumption is worth more now than consumption of the returns of investment later. Yet, for bookkeeping purposes we can only observe that $100 is consumed. The U.S. national income accounts as now kept show only the dollars consumed. Similarly, if the taxpayer derives more satisfaction by transfer than by his own consumption or investment we can only observe the $100 being transferred. The $100 is not the full value of the wealth being transferred, but we cannot determine it through observation of market transactions.

Taxpayers are generally aware of investment opportunities in the private sector and can ask themselves if the transfer would give them more satisfaction than the consumption and later investment returns given up. They may not be aware of certain opportunities for the public to loan money to certain groups who cannot express their loan demand in the private sector. If this were quite high and known to the taxpayers, they might prefer to tap these returns rather than be charitable and transfer their wealth. This does not seem highly probable. For these reasons, it is preferred here to consider only the observable nominal value of the dollars transferred by the general taxpayer.

PROGRAM ANALYSIS AND REGIONAL ECONOMIC OBJECTIVES

BY MARTIN C. MCGUIRE*

Martin C. McGuire is Associate Professor of Economics at the University of Maryland.

All Federal expenditure decisions have an economic impact on the various regions of the country. In some cases, Federal programs are an important determinant of the pattern and extent of a region's growth and development. Since the National Government is concerned with the relative prosperity of its component regions, and has economic objectives pertinent to regional incomes, it is essential that the influence of public expenditures on regional economics be an integral part of the broader analysis of these expenditures.

Professor McGuire argues that "the first approach for policy analysis should be to consider means of 'regionalizing' the massive allocations now undertaken." In this paper, he sets forth an approach for incorporating the evaluation of regional impacts into the on-going process of policy analysis and offers his judgment on the key issues for deliberation in developing a comprehensive procedure for regional analysis. He presents a format for setting regional objectives and systematically appraising the regional impact of Federal actions and uses this format to analyze the goals of economic efficiency and distributive justice in appraising regional objectives.

Professor McGuire emphasizes the existence of a number of basic issues which must be resolved if Federal expenditure policy is to be coordinated in attaining regional objectives. These issues include the development of regional budgets for both agencies and the entire Federal Government, of criteria for defining appropriate regional areas, of methods for projecting regional trends, the development of a proper mix between migration and industrial location policies, evaluation of the various alternative policies for influencing regional development, and the need for systematic intelligence on regional economic and social patterns. With respect to the last issue, he notes that there is "at present no means of systematically comparing and anticipating the regional effects of various programs; hence, the regional distribution of programs tends to fall out of executive action as a by-product of decisions, rather than as the result of conscious choice."

I. *Introduction*

While the Economic Development Administration, Department of Commerce, [1] was established explicity to pursue economic development goals for declining regions and small areas of the country, all Federal expenditure programs have an effect on the regional pattern of national growth. Currently, however, no single agency or group of agencies exists within the executive with an effective mandate to appraise the regional impacts of expenditures programs or to determine regional development programs and policies.

[1] The Economic Development Administration created by the Public Works and Economic Development Act of 1965, operates in about 900 counties with grants to communities for public facilities construction and for development planning, with low interest loans to private business, and with technical assistance. Staffed by about 1,000 people, this agency's annual appropriations have averaged about $300 million in recent years (plus $100 million for Appalachian highways).

*I am indebted to Charles Schultze, Robert Haveman, and Robert Raynsford for comments on a draft of this paper.

It is arguable that no consensus exists for including regional objectives explicitly within the domain of public policy. It is not arguable, however, that Federal laws, policies, and programs have an enormous impact on regional development. This fact alone provides strong reason for analysis of ongoing Government operations and policies on regional development. Before proceeding to outline such analysis, however, it will prove useful to clarify a possible ambiguity in the term "regional"—to distinguish it from "subnational."

Particular Federal programs and policies exist for a great variety of sub-Federal problems and objectives; for example, Federal objectives exist for urban areas taken as a whole, or for rural areas taken as a whole, for all recreational areas, or for classes of low-income areas. Such objectives as these, although they may have "regional" implications, are not regional objectives. In this paper the term "regional" will be used with reference to an exhaustive geographic division of the country into sets of adjoining areas, and to our concern about the distribution of economic activity among those areas.

With this definition of "regional" understood we may ask what are the main features of the impact of Federal activities on regional development? Three characteristics stand out. The decisions which affect regional development span the entire Government—from defense programs, through the total range of domestic programs, to commerce and industry regulatory agencies, and tax policy. Second, the aggregate effect of Federal decisionmaking on regional development is no doubt enormous, although the extent to which Federal activity establishes regional trends—in contrast to following and thereby amplifying trends inherent in the private economy—is uncertain. Third, Federal policy while often directed toward various subnational objectives never articulates regional objectives as such. Policy analysis of the effects of this multitude of decisions on regional development are largely fragmented, with the result that no overall, Government-wide, coordinated strategy of regional economic development exists today. A few examples will illustrate the wide range of Government actions influencing the regional distribution of economic activity.

1. Federal installation decisions (with their implied continuing payroll expenditures), and Federal procurement and contract decisions obviously have enormous effects.

2. Tariff and trade policies have selective impacts on different regions (e.g., import duties on textiles).

3. Transportation price and route regulations as well as specific building programs (highway, airport, waterway, etc.) favor some parts of the country more than others.

4. Some "regions" of the country have their "own" economic development agencies known by other names, e.g., the Bureau of Reclamation or the Department of Agriculture.

5. Many grant-in-aid or transfer programs proceed on formula bases with so much weight given to population, so much to State area, so much to income, etc.

6. Tax policy may effectively favor one region more than another, the oil depletion allowance, for example, presumably being a significant resource to oil- and gas-producing States.

7. As a last example, there are agencies with the specific mission of economic development in selected parts of the country, the Economic Development Administration mentioned previously, the Appalachian

Regional Commission, the Upper Great Lakes Regional Commission, and others. For quantitative impact on regional development, however, these latter agencies are dwarfed by the larger forces of the Federal Establishment.

Of course, the impact of Federal policy on regional development does not end here (even supposing we had drawn up a complete list of Federal actions affecting regional economies). We have mentioned only first round or direct effects of such Federal actions. Added to them must be the indirect, induced effects resulting from the Federal impact on the private economy.[2]

Why is it that (excepting EDA and the regional commissions) so many decisions with far-reaching regional implications are taken without systematic analysis of those effects and often, presumably, without explicit recognition of those effects? First, it should be emphasized that neither are the regional effects of particular programs, nor the consolidated regional effects of all programs and policies, taken together, ignored out of negligence. Rather the fundamental causes of this lack of analysis are two: The first is that agencies with programs having *multiple effects* are organized to pursue *single objectives at least cost.* This principle of organization naturally excludes examination of such questions as how much economic growth in rural America could be bought for a one-tenth of 1 percent inflation of defense procurement costs (about $25 million in fiscal year 1968), since such nondefense purposes are excluded from Department of Defense objectives. It could turn out, however, that the $25 million increase in defense costs would be more effective than an equal appropriation increase to another agency with a more direct interest in development or poverty. At least one could not know until the question is analyzed; right now it is not even being asked. As a corollary of this first reason, a second reason for disregard of systematic analysis of the regional effects of various programs is that such analysis would be necessarily inter- or supra-agency, and neither the executive machinery nor a body of informal practice and tradition exists for coordinating regional detail of programs and budgets.

The very magnitude of the Federal impact on the regional distribution of economic activity, and the fact that its impact is relatively uncoordinated suggests that the first approach for policy analysis should be to consider means of "regionalizing" the massive allocations now undertaken. As we shall argue later in this paper, such policy analysis should consider not only the impact of Federal policy on demand for capital by region, i.e., regional effects on industrial location and activity level; equally important is the hitherto largely neglected question of Federal impact on the regional distribution of population and labor force, i.e., of the supplies of labor. In short, regional policy analysis which omits the possibility of national migration policies is necessarily truncated. I happen to think that a confluence of pressures must sooner or later result in greater explicit regionalization of Federal planning and programing than now exists. The point of this paper is not merely to advocate that outcome. The paper also sets forth an approach to regional analysis as well as the key issues for deliberation as I see them.

---

[2] Foremost among efforts to measure such induced effects by region is the work of W. W. Leontief, "The Economic Impact—Industrial and Regional—of an Arms Cut", *The Review of Economics and Statistics,* vol. XLVII, No. 3. August 1965, pp. 217–241. See also the paper by Haveman in this volume.

II. Key Elements in the Analysis of Regional Objectives

A FRAMEWORK FOR REGIONAL ANALYSIS

The range of issues called to mind by the term "regional development" is indeed wide. Ultimately all economic activity including its spatial distribution is explicable as an interplay among resources (including human resources), technology, human wants, time, and money. On a more practical level, something concrete is needed to define and limit the reach of the subject for policy and planning purposes. Table I will be used as a format for relating the various issues to be discussed to each other within single unified framework.

TABLE I

	Regions				
	1	2	——	N	National totals
Population					
Work force					
Employment and average					
Wage per employee:					
Total					
Industry 1					
Industry 2					
Industry N					
Personal income					

Table I suggests a format for displaying the major indicators of regional *economic development.* An important component of policy analysis may be usefully thought of as the systematic appraisal of the impact of Federal actions on the numbers in such displays. A single table, such as table I will display the distribution of population, work force, and employment by region: from these figures the unemployment rate in each region can be derived. Also shown is the structure of industry by region; this structure together with knowledge of each industry's regional wage pattern allows one to calculate incomes and hence per capita income differences among regions. (The table could be expanded to include more detail, e.g., occupational breakdowns). It should be observed that the table is not explicit as to the numbers of industries or of regions to be included, nor as to the definition of either. These questions themselves are important policy questions requiring analysis.

If one were to examine a sequence of such displays for various years, one's attention would turn to certain key changes in the indicators of regional economic activity—changes in population brought about by birth, death, and migration; changes in total employment reflecting each region's participation in and contribution to national employment growth; changes in the industrial structure of each region; changes in relative wages among industries and regions, and the resulting growth or decline in incomes available to various localities.

Such displays of data are also useful for constructing preferred regional policies. In this context, a systematic appraisal of regional alternatives will feature the following broad categories of analysis: (1) an analysis of objectives—resulting in decisions as to the *relative desirability* of various regional distributions of jobs, income, population, etc.; (2) an analysis of the tools or instruments appropriate for achiev-

ing alternative objects; (3) an analysis of costs, with an appraisal of whether the objectives desired are worth those costs. Naturally, analysis of the forces at work in the economy which lead to current regional configurations, as well as the means and costs of altering their direction, cannot all be reduced to a single format. Such displays as table I nevertheless provide a useful framework for discussion of these major policy issues.

VARIETIES OF REGIONAL OBJECTIVES

As we stated above, the evaluation of what constitutes a good pattern of regional development ranks high in the priority of issues for policy analysis. An advantage of the format just described is that it helps one visualize the major alternative objectives for regional development policy.

One primary difference between competing regional objectives would distinguish an *ex post* remedial, or corrective, conception of Government purposes, from an anticipatory, preventive, more directive view. Some might wish to restrict Federal objectives to building up (or phasing out!) economic activity in regions already observed to be in distress, to ease the differences in impact of the business cycle on differing regions of the country, or to ease the impact of technological change on various regions. Such alternatives would attempt not so much to direct geographical patterns of development as to ease the transitions within any region, occasioned by major alterations in demand conditions, in resource availabilities or technology. Alternatively, others might support Government efforts to anticipate and deliberately alter the geographic pattern of development. In this case, policy analysis would be directed to planning for a (moving) future time horizon, whereas in the former case, regional development policies would lag economic events by at least the interval required for data collection and reporting.

Although the philosophy of Government between these two conceptions of regional objectives may differ substantially, the practical implications of the two will diverge less. Inevitably the policy analyst will be drawn toward anticipating regional trends, since the effectiveness of remedial efforts in any particular region will depend on the larger trends in the private economy. Nevertheless, there would be significant differences between the two outlooks. The degree of commitment implicit in regional policies, the information requirements for analysis, and the appropriate policy tools will differ substantially between regional objectives which anticipate economic events and try to determine desired outcomes, in contrast to objectives of correcting, reversing, or compensating for undesirable outcomes already in existence.

Whichever of the alternatives stemming from anticipatory vs. remedial outlooks one prefers, some further major alternatives arise as to the criteria for judging any one configuration of regional development to be better than any other. Three general criteria seem plausible. I list them in decreasing order of general acceptability:

1. *Regional equalization of income and employment relative to population and labor force.*—By this criterion, a regional configuration is better the less the dispersion in unemployment *rates*, or *per capita* (or

per family) income it contains. To a major extent this is the underlying criterion of the Public Works and Economic Development Act of 1965.

2. *Regional equalization of income-employment-population densities.*—Sometimes arguments are made that (aside from or in addition to employment and income objectives relative to population) the *absolute* amounts of population and industry are maldistributed geographically; that too many people and too much industry are located in metropolitan areas, for example; that new cities and towns should develop in rural America. By this criterion the desirability of one regional configuration over another would be determined by comparing the absolute numbers of people, jobs, etc., relative to land area or some other measure of resource base in various situations. Such arguments are in fact less arbitrary than they might first appear, depending as they do on the economic case against congestion, soon to be discussed.

3. *Regional equalization of industrial structure.*—Lastly, arguments sometimes (more rarely) are made that the structure of industry varies too much from one region to another, that more high-wage industry should locate in the rural South or that a particular State should have "its share" of Government contracts.

ECONOMIC EFFICIENCY AND DISTRIBUTIONAL JUSTICE IN THE APPRAISAL OF OBJECTIVES

No one should suppose that a single one of these criteria can govern our choice of objectives in all cases; obviously the underlying rationale for economic development of Alaska and Harlem may differ. Since the choice between objectives and of the degree to which each is pursued must ultimately express a balance of judgments on the part of various regions or groups, political judgments and personal preferences will figure importantly in those choices. Nevertheless the analyst can fashion some structure for comparison of alternative choices. In fact, since legislative guidelines for executive policies and programs tend to direct general objectives and intents, the policy analyst will typically be required, (1) to specify in detail a broad range of options allowed by legislation, (2) to identify and reject those options which are inferior on all criteria, and (3) to present advantages and disadvantages of the remaining options. To assist in this process, the economist can identify certain major economic benchmarks relevant to a comparison of the three objectives.

On the one hand, there may be a case in economic efficiency to be made for altering geographic patterns of economic activity. Two examples of such efficiency arguments come to mind. It has been observed that when national average unemployment rates are very low, prices rise; the lower the average unemployment rate the greater the rate of price increase. It is also known that, whatever the national average rate of unemployment, there is a considerable dispersion of regional unemployment rates about that average. If tools could be found which increased demands primarily on the unemployed resources of lagging regions, but not on the more fully occupied resources of prosperous regions, the national aggregate output might thereby be increased and national average unemployment rate reduced, with less upward pres-

sure on prices than would result from indiscriminate demands on national resources.[3]

As a second example of the economic efficiency criterion for comparing various objectives, it can be argued that, even though real incomes in large urban centers are larger than in rural areas, population concentrations beyond a certain size result in congestion, in pervasive external diseconomies among producers and consumers, and in diseconomies of scale in production, administration, and control of public services. Intuitively, this proposition has appeal and a limited amount of empirical work exists to confirm that the *per capita* costs of providing public services first decline for increasing size of metropolitan area, reach a minimum, and then for even greater size cities increase.[4] Implicit in arguments to retain the rural population in rural areas is just this idea that the evils of congestion, environmental pollution, and isolation from the natural environment, which accompany very large population concentrations, outweigh the advantages of economies from the very specialized division of labor common to our largest cities.

The net result of such factors is to suggest an economic hierarchy of urban areas; presumably a large number of fairly large separated urban complexes can satisfy consumer wants and needs better than can a smaller number of immense urban agglomerates. One might also conjecture that vast urban areas have grown to uneconomical proportions because new arrivals (or established residents) in such areas can enjoy the advantages of urban life, and avoid paying for the disutility their presence imposes on established residents (new arrivals) in the same area.

In terms of table I, this efficiency criterion would recommend those changes in the tableau which result in greater national totals—that is, changes which increase the total employment, income, or product of the national economy. This may seem a simple enough criterion until it is realized that a net increase in national totals may result if the additions in one region exceed the reductions in others. This "improvement" in efficiency of the national economy might command less than unanimous support; yet, if benefits to one region, no matter how great, could not be attained because of small sacrifices imposed upon other regions, Government effectiveness in regional development would be paralyzed. In fact, such Government action is not paralyzed for a number of reasons: first, the true net regional effects of the vast array of Government decisions are not known; this may be advantageous, since large conscious, explicit regional actions tend to be more difficult than the more or less unconscious use of functional programs for regional objectives. Second, in a growing economy all regions may gain absolutely over time. And third, a vague consensus exists that some regional redistribution of prosperity is bearable, provided the gains to the gainers are relatively great, and the losers are not too great.

When money is allocated and plans and policies are established, however, this vague distributional ethic takes on a sharp edge. Actual choices imply just how much is taken away from one set of regions and given to another—choices which necessarily introduce a second major category of argument for purposeful Government intervention in the

[3] This argument is essentially static. Subsidization of regions with high unemployment in a dynamic context may redirect investment so as to lower the growth of productivity increases, and hence reduce future real incomes and thereby possibly restore some pressure on price levels.

[4] Unpublishd paper by Professor Roger Noll, California Institute of Technology.

geographic location of economic development—namely, distributional justice or equity.* Equity arguments may assume a variety of forms. One variation is that unemployment rates should be equalized across regions even if national average unemployment were not thereby decreased. Similarly, it can be argued that having the spread in regional per capita incomes narrowed is worth while, aside from changes in the national overall total—or distribution—of incomes per capita. This is basically an ethical judgment of good and bad, that having a given amount of employment and income spread around the country is better than having that same amount concentrated in a relatively few areas. In terms of table I, this asserts that some regional configurations would be preferred to others even though they allow for *lower* national totals in income, employment, and output. One might be prepared to reduce total national prosperity somewhat—that is, pay certain costs—to secure a more equitable distribution of prosperity—that is, gain certain benefits.

An ethical judgment of this nature, although difficult to make, is not unique to the problem of the regional distribution of unemployment or income. A similar question arises with regard to the distribution of unemployment among occupational groups—or by skill level, educational level or other classification—namely, how much unemployment at the low—or disadvantaged—end of the spectrum are we, as a nation, willing to tolerate in the interests of economic efficiency or rapid economic growth. In most cases one can observe a tendency to establish a threshold beyond which unemployment is considered intolerable. In the Public Works and Economic Development Act of 1965, for example, 6 percent unemployment is designated as the threshold for county-sized labor market areas. Presumably the larger the labor area the lower will be such a threshold.

One particular implication of a pure equity argument may seem questionable to many. Since this ethic would hold that merely narrowing the differences in unemployment rate or per capita income statistics between various regions is good in itself, it might recommend transferring unemployed people from one region to another without necessarily employing them at their destination, since this procedure would, in fact, reduce statistical variations in unemployment—and income—between regions. This proposition may not be so questionable, however, if one considers a third category of argument for positive regional economic objectives—namely, a particular mixture of equity and efficiency considered in a dynamic context. Assume that a region can handle only so much poverty and unemployment, that beyond some critical threshold measured as a rate of unemployment, or proportion of people in poverty, these unemployment rates and low incomes become self-perpetuating, grow beyond the capability of local government to correct, and cause external diseconomies in the form of social misbehavior.[5] It is generally agreed that self-perpetuating cycles of poverty occur and if the threshold of concentration of poverty and/or

[5] Implicit in this form of argument is a somewhat organic concept of a regional economy. This assumption finds support in Glen L. Johnson, "Supply Functions, Some Facts and Notions" in Heady *et. al., Agricultural Adjustment Problems in a Growing Economy* (Ames, Iowa State Press, 1958), and in Rufus B. Hughes, "Interregional Income Differences: Self Perpetuation," *Southern Economic Journal,* 28; 41–45, July 1961.

*Further discussion of this issue is found in the papers by Weisbrod, Bonnen, Freeman, and Schmid, in this volume, and Feldman in vol. 3 of this collection.

unemployment can be estimated, then these facts argue for evening out the incidence of economic distress among regions even if no national increase in income or employment results from such action.

Still another possible rationale for regional development may be purely esthetic or ethical preferences for certain configurations of population and industry. While mere opinions do not justify certain regional objectives the way the foregoing three approaches do, when value judgments are widely held in society, that very fact constitutes good reason for considering policies to carry them out. In this case, the requirements for analysis are no less necessary than in the efficiency-equity context; in particular the analyst must examine the feasibility and the true costs of following such preferences.

In summary, policy analysis and the analyst must confront the fact that a wide range of possibilities exist for the regional distribution of economic activity, that legislative or executive guidelines may narrow that range somewhat, but that a crucial task of analysis must be to make explicit the possible trade-offs between efficiency, equity, and noneconomic considerations.

III. Some Basic Issues for Policy Analysis

To this point the discussion has concentrated on the larger question of establishing a framework for policy analysis of economic development. Such a framework which can encompass the fragmented influences of the Federal Government on regional development is essential to conscious and efficient employment of the instruments of Government to social on economic goals. Simply to capture and display the net regional impacts of Government policies and decisions is an enormous task. It is a prerequisite to effective legislative and executive control of Government operations; yet this framework only initiates policy analysis. We therefore will turn next to discuss the more important particular issues for policy analysis within this larger framework.

REGIONALIZATION OF FEDERAL PROGRAMS

Foremost among these particular policy issues is the question of explicit regionalization of Federal agencies and the Federal budget. The possibilities for such regionalization range from (1) making regional consolidations of the Federal budget within the Bureau of the Budget, (2) to the creation of regional governments with the constitutional and political changes such innovations would imply. Since these options are discussed elsewhere in this volume,* and since the issues they call forth reach far beyond economics, they will not be discussed in detail here. It is important to note, however, that the pressures for such regionalization are largely economic. The complexity of our social and economic problems seem to demand a national policy on interregional development, on urban growth and development, and on environmental control. A grander national strategy than now exists is needed to channel the migration of people and of resources toward social and economic betterment. The half-conscious and inefficient use of current functional programs to influence regional imbalances is becoming inadequate to the country's needs.

*Further discussion of this issue is found in the paper by Schultze in this volume.

A larger perspective with explicit choices made public is required. Included in this larger strategy must be the following central elements.[7]

1. Industrial location policy.
2. Immigration policy.
3. Urban size and location.
4. New cities policy.
5. Depressed areas policy.
6. Internal structure of cities.

CRITERIA FOR ESTABLISHMENT OF REGIONS

We have used the term "region" repeatedly throughout this paper without definition. This has been deliberate, since determination of what, for policymaking purposes, constitutes a region is itself an important question for analysis. The question is complicated. On the one hand, one may search for an existing hierarchy of regions based on natural boundaries—e.g., river basins—on economies of scale in production and local government operation, and on spillover effects of costs and benefits between subregions. Whether or not such a national hierarchy of regions has been discovered, or defined, the further question remains of how small an area or how few people must live in a locality for the U.S. Government to lose effective operational interest in the economy of a locality as such (not its people). This question is essential since distributional or equity considerations should be built into regional policy from the program design stage forward—e.g., regional transportation systems design.

It should be understood that regional prosperity for its own sake is not a Government objective. The economic prosperity of people, rather, is the proper objective. This being the case, the fineness or degree of detail of Government interest in local or regional economic prosperity and development would seem to stem from two sources. First, to secure economic prosperity for the citizens, the creation of a viable local economic base and hence, of self-sustaining local or regional prosperity may be adequate *substitute* for a variety of individually oriented programs (and very economical in terms of information required for administration and control). (Imagine the "local economy" to be an individual. If the individual earns $100,000 per year the Government is probably justified in being unconcerned about his housing, medical, transportation, education/training, etc., standards. His income serves as an adequate proxy for all these particular elements of his living standard.) By this standard Government should find itself inclined toward defining a "region" as a small area. As smaller areas are considered local economic prosperity becomes a better proxy for individual prosperity,[8] and more individual-oriented programs can be included under the general rubric of economic development. Second, the Government will have an interest in any region which people are unwilling or unable to abandon when its local economy fails or declines. This

[7] These elements are discussed in detail in a paper of Charles T. Stewart, Jr., "A National Policy for Regional Development."

[8] There may be cases in which this conjecture is false. Imagine a large region with two populations, one very rich, the other very poor and *each evenly distributed* over the terrain of the region. No matter how finely one divides the region into subregions, the subregional average income will mask a bi-modal distribution.

argument would tend to restrict regions to fairly large areas since the larger the area the lesser the proportion of its people could escape from economic depression. By either of these accounts, to repeat, the Government is not interested in regional economic development for the sake of the region but rather because of the people who occupy the region.

Of course, the foregoing account of what determines a "region" implies that this issue itself can only be resolved with the help of analysis. Specifically the size cut off below which it is not worth the Government's effort to be concerned with economic development must depend in part at least on the alternative costs of handling the human problems in such areas by direct means—such as welfare, or negative income taxes, manpower, health, housing programs, etc. This in turn will depend on the prospects for organizing a pyramidal structure of regions and subregions and on the possibility of creating incentives for sub-Federal authorities to carry out larger Federal purposes. Simultaneously to be determined with the question of size of region are the exact boundaries of regions (of the "proper" size). There is a substantial body of literature on this subject, of how to carve a region up into x subregions when the value of x is known.[9]

Whatever definition of "region" one selects, the discussion which follows table I concerning alternative objectives makes clear that regions should be chosen to include the entire country. A regional organization which includes (say) only half the country precludes, or invalidates, the comparison of regional alternatives on efficiency and equity grounds which should form the core of policy analysis. (Such comparisons would be invalidated since excluding part of the national economy from the regions would result in mistaken estimates of costs and benefits of various regional configurations.)

FORECASTING REGIONAL TRENDS

Thus far in this paper we have been concerned with the questions surrounding what is desirable. This and succeeding sections of the paper are concerned with what is feasible, and economic.

Let us assume that a structure of our table I format has been decided, at least tentatively, so that planning can proceed on the basis of a known regional breakdown. Within this framework alternative regional objectives would be articulated as specific regional configurations of (1) population, (2) of industry, (3) of industrial structure, (4) consequently of industry wages/salaries, and therefore, (5) of family (or per capita) incomes, which appear feasible and desirable. What is a feasible or desirable regional configuration depends on three factors: (1) the basic objectives of regional economic development policy or policies (alternatives discussed above), (2) the future course of regional developments, and (3) the costs, economic, social, and political, of attaining various configurations.

For determining the feasibility of achieving a desired configuration, whatever the objectives contemplated, analysis can proceed only from an understanding of the forces at work in the economy which have led to present regional patterns. A theoretical structure is required to

[9] For example, the 1967 report of Brian L. Berry "Functional Economic Areas and Consolidated Urban Regions of the United States," sponsored by the Social Science Research Council.

allow the analyst to estimate future locational trends under various assumptions as to underlying trends in consumer taste and technology, and hence to gage the impact of alternative Federal policies and programs. The outcomes of these projections of location trends can be very usefully cast into formats like table I. Such projections will form a base line from which various policies may be evaluated.

Recent advances in national and regional forecasting both within the Government and outside [10] give hope that with a few years of sustained effort (concentrating on data collection and econometric modeling) quite reliable projections of regional economies will be attainable. *Such projections will always depend upon certain underlying assumptions.* It is mistaken therefore to think of these projections as predictions of regional development; rather, they are a basic tool for management and policy analysis, for estimating the sensitivity of regional development to underlying economic forces and to alternative government actions. As such, projections of this type can provide only the roughest of guidelines to policy makers as to the viability of alternative development strategies. Pilot efforts at this type of projection have already been made. Table II below gives an example for four industry groups and five regions. The table shows expected percent changes in total employment, industrial structure, and of income in each region and in the nation as a whole *under one particular set of assumptions.* In practice, such projections may be made at a finer level of detail with as many as 50 industries and at a geographic level as small as a county (naturally the finer the level of detail the less reliable the projection for particular "industry and region" cell.) Given the assumptions on which it is based, table II estimates the demand for labor in five major regions (chosen purely for illustrative purposes). Other information on regional wage trends by industry allowed for projections of the total income available to each region.

TABLE II.—FORECASTS OF PERCENTAGE CHANGE IN EMPLOYMENT AND INCOME, 1975/1960, WITH ASSUMED NATIONAL UNEMPLOYMENT RATE OF 4 PERCENT THROUGHOUT THE PERIOD 1965–75 [1]

(In percent)

	U.S. total	Regions				
		North-east	Mid-west	South	South-west	North-west
Total employment	+29	+16	+16	+37	+57	+51
Agriculture, mining	−34	−34	−42	−48	−25	−38
Construction	28	21	11	30	48	39
Manufacturing, utilities, trade, transportation	20	3	16	38	47	47
Services, finance, government	60	45	44	65	92	82
Personal income	85	71	63	93	123	106

[1] The 5 regions include States as follows: Northeast—Vermont, Rhode Island, Pennsylvania, Ohio, New York, New Jersey, New Hampshire, Michigan, Maine, Massachusetts, Indiana, Illinois, Connecticut; Midwest—Wisconsin, South Dakota, Oklahoma, North Dakota, Nebraska, Missouri, Minnesota, Kansas, Iowa; South—West Virginia, Virginia, Tennessee, South Carolina, North Carolina, Mississippi, Maryland, Louisiana, Kentucky, Georgia, Florida, District of Columbia, Delaware, Arkansas, Alabama; Southwest—Texas, New Mexico, California, Arizona; Northwest—Wyoming, Washington, Utah, Oregon, Nevada, Montana, Idaho, Hawaii, Colorado, Alaska.

The projections are based on a set of assumptions as to national growth patterns from 1965 to 1975, and on a set of structural equations (derived from 1940–65 data) representing the geographic distribution of employment. These procedures are explained in detail in McGuire and Harris, op. cit. The figures in table II are projected from the 1965 actual employment distribution.

[10] Curtis C. Harris, Jr., *State and Local Projections,* (Bureau of Business and Economic Research, University of Maryland, 1969).
Clopper Almon, Jr., *The American Economy to 1975,* (Harper, 1967).
Ira S. Lowery, *Migration and Metropolitan Growth* (Chandler Publishing Co., 1969).
M. C. McGuire & Curtis C. Harris, "Planning Techniques for Regional Development Policy," *Journal of Human Resources,* forthcoming.

One serious shortcoming in regional forecasting techniques presently employed is that no general model to explain or project *both* population movements and industrial shifts *simultaneously* yet exists. Hence the analyst can only fall back on a comparison of various population migration and industry projections which have been derived independently. Nevertheless, when matched with such independent estimates of future population and work force by region, the information in table II will give the analyst some clues as to expected welfare indicators such as the regional unemployment rate and the per capita or per family income given the sets of assumptions on which the population and industry location projections are based. As an example of this procedure, table III below shows three 1975 population estimates on the same regional breakdown as table II.

TABLE III.—REGIONAL SUMMARY OF ALTERNATIVE POPULATION PROJECTIONS, 1960–75

[In millions]

	Northeast	Midwest	South	Southwest	Northwest	Total
1. 1960 population	77.0	21.7	43.2	27.8	10.0	180.0
2. 1975 population if no interregion migration occurred (from 1965–75)	93.1	25.3	54.5	37.8	12.9	223.6
3. 1975 population if 1950–60 net migration trends continued between 1965–75	91.4	24.5	54.7	40.1	12.9	223.6
4. 1975 population which would insure 4-percent unemployment in each region given the job of projections of table II	89.3	24.3	57.6	39.3	13.1	223.6

Table III indicates that, roughly speaking, a continuation of past migration patterns will tend toward equalization of regional unemployment rates—given the employment projections of table II. This can be seen from the progression of lines 2, 3, and 4. A continuation of past trends will reduce population in regions where more population is projected than can be sustained with 4 percent unemployment—e.g., Northeast. Similarly where the population base is inadequate to the demand for labor, past trends show an inflow of migration—e.g., the Southwest.

Tables II and III also are useful in illustrating a potential conflict between employment and income equalization objectives. Compare the regions labeled "Northeast" and "South." In the former, natural population growth exceeds industrial growth; in the South, the reverse is true. The resulting excess supply of labor in the North and excess demand in the South should tend to reduce the differential in wage rates (and hence in per family incomes) between the two regions. As this gap is closed by migration of people, however, the wage differentials may reappear. Hence, nonmarket, Government-sustained incentives may be necessary to achieve greater parity in both measures of regional welfare simultaneously.

The comparisons illustrated by tables II and III have other uses. As techniques improve, they should give good indication of which particular areas of the country one may expect to decline and, therefore, to need assistance as depressed areas and which not; they may allow advance identification of the symptoms of regional economic distress, and of the source of the difficulties; e.g., projected technical

changes in certain industry groups. One might even hope that such projections and comparisons of employment, industrial structure and wages, population and labor force will serve to narrow the range of potential disagreement over objectives of regional development. Explicit estimates of expected locations of jobs, people, and incomes should at least spell out the quantitative differences between different objectives and assist in communicating policy decisions and their rationale.

Lastly, such comparisons may suggest alternative concrete policies and programs for regional development. For example, one may be led to consider policies to change not merely the gross direction of industrial development, but, also, to change the composition of industry from region to region. (If you contemplate raising the average income of a low-income region, you might wish to consider changing the proportion of low- to high-paying jobs in the region by encouraging high-wage industries to locate here.) *At the very least, an examination of such statistics and a comparison among regions may lead one to consider the alternatives of attempting to redirect industrial development versus attempting to influence the direction of migration of people.*

Such techniques, it should be stressed, remain planning tools and nothing more, even with the most sophisticated refinements. A complex, multiphased problem such as regional development cannot be compressed into a few equations. Still, the projection technique can be most valuable for emphasizing structural relations among regions and for analysis of major effects of alternative Government policies.

NATIONAL MIGRATION POLICIES OR INDUSTRIAL LOCATION POLICIES

Migration versus industry location is a major question for policy analysis if only because no serious national attention has yet been given to population location as a possible objective of Government policy. Yet, it is a crucial element in regional development. An examination of tables I, II, or II suggests that, for achieving regional balance in unemployment rate, *some* flexibility or choice will exist between trying to influence the location of industry and trying to influence the location of people. The proper mix between these two objectives as a major policy issue is at the core of overall regional development policy. The question is not merely that of the best location of the present population. More critical is the question of where will the 100 million or more Americans to be born over the next generation live and work.

If one were unconcerned about the absolute amount of employment and product generated in each region, attempts to influence the location of industry and of people should be viewed as two equally effective means of changing the distribution of welfare—unemployment *rates* and *per capita* incomes—among regions, both deserving consideration. If, on the other hand, the numbers of people and density of industrialization in particular regions are of concern, obviously a population location policy is more than an alternative; it becomes a necessity. In either case, the novelty of attempts directly to influence where people live and work becomes less startling if it is realized that, when projected into the future, major shifts in the location of industry and population are going to take place in any event. Hence, policies to guide location decisions need not interfere with the decision

to relocate itself, but probably can concentrate on the question of the direction of relocation. These problems, possibilities, and techniques of influencing the migration patterns of people are only beginning to receive serious study.

THE APPROPRIATE POLICY INSTRUMENTS FOR INFLUENCING REGIONAL DEVELOPMENT

There is a vast array of tools available to the Federal Government for providing incentives to private and public corporations and individuals, at all levels of the socioeconomic system. In contrast to offering incentives to people, groups, and governments, our Federal Government might possess or obtain powers to regulate economic development by decree—without providing incentives. Such actions, as establishing quotas for population and industry by region, would fall under this latter category. In addition to being politically intolerable, these latter alternatives are known to be economically inefficient. Hence, we will dismiss the possibility of creating a "command economy" as both undesirable and useless. This still leaves an enormous variety of instruments open to Government, among which the following might be considered:

1. Grants to local authorities.
2. Government loans to business enterprise or to individuals.
3. Interest, wage, or rent subsidies or tax rebates to private business or individuals.
4. Relocation and/or travel subsidies to individuals or business.
5. Government facility location and employment practices.
6. Set-asides of Government contracts for specific regions.
7. Higher than competitive prices accepted on Government contracts from specified areas.
8. Planning and technical assistance to local authorities or businesses.
9. Differential welfare, social security, or negative income tax payments.

Just which of these and other alternative instruments ought to be preferred and in what circumstances employed by Federal authorities can only be determined in the light of detailed analysis of the costs and estimated effects of the various alternatives. Our remarks in this paper must be restricted to a few general observations. By and large, these potential tools for influencing the geographic patterns of economic activity are neither fully exploited nor is their use coordinated between various agencies. Nevertheless, even without detailed knowledge of the benefits and costs of such alternative instruments some general evaluations are possible.

First, one can infer that, for effective regional development policies, a combination or mix of instruments is required. This follows from the fact that changes in the regional employment-income-population balance will call for creating an industrial structure, supplying adequate manpower, and providing public facilities and other infrastructure items approximately at the same time. It is illusory to search for a single item which will unlock or stimulate economic development for regions at small Federal expense. The incentives for wishing that

regional development can be bought cheap—e.g., by building water and sewer systems—no doubt are great, and there may be some cases in which a single bottleneck or obstacle blocks economic development. But on the whole the investments required for regional development will surely be comparable, at best, to the costs of growth in the national economy at large; more probably the developmental investment to stimulate growth in lagging economies will be greater than the national average—which is the very reason the lagging economies are that way.

Another useful distinction to be made between the possible policy instruments is to distinguish those which can be applied on a discriminatory basis to small areas from those which could only be applied more broadly for larger classes of areas. This distinction more or less coincides with the difference between instruments which would be effected as standing formula-type programs or regulations, and instruments which are effected on a project-by-project or discretionary basis. Policy instruments such as differential subsidies to costs of various factors of production, e.g., land, labor, and capital, are examples of instruments with very broad consequences across the entire economy. (One sorely needed type of information for determining the advisability of such subsidies is the effect of interest, land, and wage subsidies on employment, outputs, and profits of the various industries by the various regions of the country.) I conjecture that it would be difficult to write broad formulas discriminating finely between counties—for example, between adjacent counties. Such discrimination is easily accomplished, however, when discrete project decisions are made by Federal administrators. This suggests that formula-based programs should be directed toward quite broad regional objectives valid for large classes of counties or cities. From the viewpoint of economy of administration, minimal interference in the minutiae of private decisions, and economic efficiency in redirecting regional development, these general formula variables—which in effect alter the prices faced by private decisionmakers—have much to recommend them. Once promulaged, however, they would be more difficult to change in response to changing circumstances than would piecemeal lower-level instruments.

Of particular interest are incentives for altering population patterns. The instruments of wage, interest, or land subsidies might be constructed not only to alter the location of industry to bring it into closer balance with population, but also to relocate population to bring it more into balance with industry. Such incentives to relocate also should be directed toward large regional movements. One such instrument meriting consideration is an information program to inform those people contemplating moves where new jobs will be located. (The opportunities for Government/industry cooperation in this realm are obvious.) Once one considers policies beyond the relatively simple information program to a more directive program involving Federal payments, analysis would presumably have to address the questions of whether incentives—positive or negative—should be provided for people to leave certain areas, or to migrate to other areas, or both; what types of subsidies, tax allowances, or penalties to institute; what the effects of such actions would be on public facility demand and supply, on employment demands and so on, and how to integrate pop-

ulation policies with industrial development and public facility programs. More searching analysis than now exists is necessary to determine the desirability of these options. Policy analysis of possible migration programs deserves top priority on the menu of regional development issues.

To discriminate among small units such as counties one should look to less powerful instruments than those which operate on or influence the investment production and imgration decisions of the entire economy. To the extent that one desires to effect specific economic conditions at the local, country, or labor market level, therefore, some other more discretionary tools mentioned above would seem preferable. Federal installation location policy could, for example, be highly effective in influencing conditions in specific small areas. Procurement and contract set-asides may be effective provided the production capabilities exist in the target regions, and provided some cost markup is allowed. Defense, NASA, and other agency procurement contracts negotiated on specific contract-by-contract basis could be used as an instrument of low-level development. Of course, specific area development instruments are available and others can be created to encourage particular firms to locate in specific areas—such instruments as specialized interest, rent, or wage subsidies or low-cost loans, or even grants to private firms or public authorities to locate in specific areas—provided that the administration of such policy tools proceeds on case-by-case bases.

CONCENTRATION VS. FRAGMENTATION OF FEDERAL EFFORTS AND THE QUALITY OF ECONOMIC DEVELOPMENT INVESTMENTS

It is always unpleasant to face the fact that resources are limited. In the case of regional development policies the difficulties inherent in resource limitation are doubly troublesome. This would remain true even if the entirety of Government programs and policies were regionalized. First, since it is theoretically possible that in a growing economy all regions could share in that growth, it is tempting to conclude that all regions and areas *should* share in that growth. Second, the decision to allocate funds selectively amounts to the selection of prosperity for some geographic areas and decline for others; and the more such choices are made explicit, the greater the tendency to parcel some funds out to everyone. Where the politics of regionalizing Federal activity speaks for fair treatment for *all* communities, economics will argue for greater concentration. Evidence and reflection suggest that some communities and areas are not economically viable, and some are not capable of further growth. One major challenge for regional policy may be to find ways to phase out some communities and areas with minimal hardships imposed on their residents. The evidence for this is presented in table IV. Projections of past trends—based on the pilot study mentioned above—suggest that a substantial number of counties (394) will find themselves with fewer jobs in 1975 than they had in 1960—again for one particular set of assumptions as to technology consumer tastes, resource availabilities, etc. As shown in the second part of the table, by and large, these counties are concentrated among the very largest and the very smallest—in population—counties of the country. The message of these figures is clear: not all development planning can be for economic growth.

Some areas are bound to level off and others are bound to decline. There are other arguments for selectivity in Federal sponsorship of economic growth in addition to that against trying to buck the momentum of a trillion dollar economy. For one, it is reasonable to suppose that a certain scale and mix of total investment in any locality is necessary for any single investment to pay off. These conditions would call for some concentration of Federal effort. Second, insofar as Government stimulates or subsidizes private industrial investment, the objective should be to implant average American firms—that is to say, strong ones—rather than weak or marginal business with risky ventures undertaken only because the risk of loss is reduced by the Government. A strategy of locating branches of large firms in underdeveloped areas also calls for a concentration of funds, inasmuch as bigger firms tend to make bigger investment.

TABLE IV.—FAST- AND SLOW-GROWING COUNTIES, 1960–75

Percentage increase in employment or income, 1960–75	Distribution of employment growth	
	Number of counties	Percent of total 1960 U.S. population
Over 200	9	2
100 to 199	100	6
50 to 99	419	19
25 to 49	786	32
None to 24	1,362	34
Absolute decline	394	7
Total	3,070	100

RELATION BETWEEN COUNTY SIZE AND EMPLOYMENT GROWTH RATE

1960 county population of—	Number of counties	Percent increase in employment, 1960–75
Less than 10,000	822	21.0
10,000 to 50,000	1,652	25.3
50,000 to 100,000	292	37.1
100,000 to 500,000	239	39.3
500,000 to 1,000,000	49	32.8
Over 1,000,000	16	15.6
National total	3,070	30.0

To implement a strategy of selective concentrated regional development, decisionmakers will find a particular need for analysis of the appropriate balances between prospective benefit, costs, and, *need* of the benefiting population. Once the idea of selectivity is accepted it becomes paramount to formulate rules and guidelines for deciding systematically which regions are to receive most attention. Unlike traditional benefit-cost analysis, this calls for explicit comparisons of the value of subsidizing various groups. This problem receives attention elsewhere in this volume.*

*Further discussion of this issue is found in the paper by Freeman in this volume.

THE NEED FOR SYSTEMATIC REGIONAL ECONOMIC AND SOCIAL INTELLIGENCE

I have been impressed by the fact that in support of the foreign and defense activities of the U.S. Government enormous efforts are made to project or forecast developments in foreign countries and to evaluate and analyze trends so as to allow decisionmakers to anticipate require- domestic side of Government there is, in short no unified intelligence assumptions, projections, and analysis, a structure known by the foreign affairs community within which decisions and plans are made. In contrast to this, the domestically oriented agencies do not operate from a common range of assumptions, analyses, and projections. On the domestic side of Government there is in short, no unified intelligence base. One result of this is that public facility investments (for example) for any particular locality, made by HUD, HEW, Agriculture, Interior, etc., may be made on the basis of different expectations as to the physical environment in the locality—e.g., different population forecasts, if forecasts are used at all. With no common substratum of information, comparison of the impacts of various programs for the purpose of determining the proper balance among programs can proceed on only a case-by-case basis if at all. There is at present no means of systematically comparing and anticipating the regional effects of various programs; hence, the regional distribution of programs tends to fall out of executive action as a byproduct of decisions, rather than as the result of conscious choice.

RECOMMENDATION

The establishment of a domestic economic and social intelligence center in the Government, oriented toward providing information for program planning and analysis, would not in itself result in more rational consolidated regional decisions; but some intelligence substratum to unify the information base and relate it to program planning is a necessary condition for advance. The creation of such a system would probably be cheap, since the need is not for new staff but for a new orientation in existing staffs (such as the Bureau of the Census and the Office of Business Economics of the Department of Commerce, and the Bureau of Labor Statistics), to provide information specifically coordinated and arranged for program planning and decision as opposed to general information for the public.