INTERNATIONAL SERIES ON
APPLIED SYSTEMS ANALYSIS

PITFALLS OF ANALYSIS

Edited by
GIANDOMENICO MAJONE
EDWARD S. QUADE

International Institute for
Applied Systems Analysis

PITFALLS OF ANALYSIS

Edited by

Giandomenico Majone

and

Edward S. Quade

International Institute for Applied Systems Analysis

Systems Analysis—or Policy Analysis, as it is sometimes called when public problems are being addressed—aims to deepen our understanding of sociotechnical problems, and to bring about improved solutions to them.

However, the craft of systems analysis is a relatively new one, and its practitioners are still exploring ways to conduct such analyses properly, a task complicated by the necessary use of tools and knowledge from many disciplines and contexts.

Any craft learns both by what works and by what fails to work, and systems analysis is no exception. This book contributes to the craft of systems analysis by describing many pitfalls of analysis that will lead to failure, thus helping analysts and users to recognize and avoid them.

About the Editors–

Giandomenico Majone

Nino Majone obtained his first degree in Political Economy from Padua and his masters degree from the Carnegie Institute of Technology in Mathematics and his Ph.D. in Statistics from the University of California at Berkeley. He received a Fulbright fellowship at George Washington University in 1955. He has been at the Mathematics Institute of the University of Rome since 1964 and was Associate Professor with the Faculty of the Business Administration of the University of British Columbia in 1967–68. He was consultant for the National Planning Commission of Italy from 1969 to 1972 and is now on leave from the University of Rome at the International Institute for Applied Systems Analysis, Laxenburg, Austria.

8 International Series on
Applied Systems Analysis

Pitfalls of Analysis

Edited by

Giandomenico Majone

Edward S. Quade

International Institute for Applied Systems Analysis

A Wiley—Interscience Publication
International Institute for Applied Systems Analysis

JOHN WILEY & SONS
Chichester – New York – Brisbane – Toronto

British Library Cataloguing in Publication Data:

Pitfalls of analysis.—(International series on applied systems analysis; 8).
1. Systems analysis
I. Majone, Giandomenico III. Quade, Edward Schaumberg
003 QA402 79–41700

ISBN 0 471 27746 0

Text set in 11/12 pt Photon Times, printed and bound in Great Britain at The Pitman Press, Bath

Preface

Systems or policy analysis (the latter expression being used more frequently when public problems are addressed) is a form of applied research designed to acquire a deeper understanding of sociotechnical problems and to bring about better solutions. The questions systems analysts investigate are enormously complex, yet the data, tools, and theories at their disposal are often inadequate or untested. The multidisciplinary character of the systems approach is a source of strength, but it is also a source of delicate problems of communication and understanding. Consequently, it is easy for the analyst to lead himself and the policy makers for whom he works into conceptual errors or pitfalls. The purpose of this book is to point out some of these pitfalls so that eventually a well-marked path may be charted to avoid them.

The editors, and hence the format of this book, were influenced by discussions at a small workshop on the pitfalls of systems analysis held during August 1977 at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria. All but two of the contributors participated in that workshop, at which the decision to produce the book was made. The editors are indebted to the Institute for its sponsorship of the workshop, for its support in the production of the book, and especially for the stimulating intellectual atmosphere it has provided throughout our stay.

That this volume appears in the International Series on Applied Systems Analysis sponsored by the International Institute for Applied Systems Analysis reflects IIASA's interest in applying systems analysis to international and global problems and at the same time advancing the state of the art of systems analysis.

This volume is addressed to an audience of practicing systems analysts and of users of systems analysis. The international nature of the institution where

this book was conceived and took shape is apparent, we hope, not only from the affiliations of the contributors, but also from the effort to discuss problems that are important to practitioners and users of systems analysis in all countries.

Contents

1 Introduction

The quality of a systems analysis study[1] can be assessed in terms of two distinct but related sets of criteria: internal, process-oriented criteria of adequacy relating to the technical competence of the work; and external, outcome-oriented criteria of effectiveness relating to the impact of analysis on the policy process.

Internal criteria are generated within systems analysis itself, or related disciplines, and answer such questions as: Was uncertainty handled explicitly or assumed away? How strong is the evidence? Were the conclusions tested for sensitivity to simultaneous changes in the critical parameters, or was sensitivity testing done merely for changes in single parameters separately? Were extensive tests undertaken to establish the credibility of the models used?

External criteria are imposed from the outside, by the clients of the study, by the general public, and by objective limitations of time and resources. They answer questions like: Have the conclusions of the study been presented in a clear and persuasive manner, and in time to assist the decision maker? Was the study read by those for whom it was designed? Did they give any sign that it affected their reasoning? Have all legitimate interests been taken into consideration? Were the conclusions of the study found acceptable by the implementing organization?

Although both sets of criteria are needed to assess the overall quality of a study, there is a widespread tendency to raise one or the other to a position of unique importance. On one hand, analysts raised in the scientific tradition tend to assume that the high technical quality of a study is a sufficient guarantee of its practical effectiveness. This view does not explain why many good studies have had no visible impact on policies, while technically inadequate analyses have occasionally been quite effective in focusing public attention on certain problems and in stimulating suitable responses. On the other hand, the users of

analysis generally find outcome-oriented criteria intuitively very appealing. After all, the proof of the pudding is in the eating. If analysis does not demonstrably contribute to the success of a policy, what is the good of it? They forget that the process–outcome distinction, which is clear enough when results can be evaluated objectively and unambiguously, becomes blurred when we try to apply it to professional work. Even in the case of medicine, "[m]uch, perhaps most, of what a physician does must be categorized as process, and process not even calculated to affect outcome,"[2] and the same is true, in varying degrees, of the work of all professionals, including systems analysts. Moreover, the relationship between what the professional does and the final outcome is always indirect and uncertain, for there are always many factors that he or she cannot control or even foresee. In sum, tests of adequacy and tests of effectiveness are equally necessary, but not sufficient when used separately, to control the quality of a study, and in both types of testing the notion of pitfall plays a crucial role. Every problem has its characteristic pitfalls—its hidden dangers and snares—and no analysis of it can be of acceptable quality unless it recognizes and avoids them.

This volume contains extensive discussions of many conceptual and practical pitfalls that experience has shown to be particularly dangerous in the conduct and implementation of analytic studies. Of the chapters that compose the volume, some are primarily concerned with questions of adequacy, others with questions of social effectiveness.

Let us first briefly consider the question of adequacy. An analysis is adequate if it meets the particular tests that are appropriate to the nature, context, and characteristic pitfalls of the problem. But the process leading from the initial formulation to the implementation of a recommended solution is so complex that a global test is quite impossible. Rather, the analysis must be broken down into distinct components, each of which will be tested for adequacy.

An anatomy of the analytic process is described by Majone in the first chapter. Analysis, he observes, usually begins with a problem situation—an awareness that things are not as they should be, but without a clear idea of how the situation may be corrected. The problem situation has to be translated into a problem; this is the process of problem setting. Data must then be obtained and transformed into more useful and reliable information. The information is manipulated and interpreted by means of different tools and models and is used as evidence. The evidence is embedded in an argument supporting a particular conclusion. Finally, the conclusion is tested, communicated, and, it is hoped, implemented.

This stylized description is only a first approximation to the full complexity of the analytic process. It traces a simple linear path in an intricate network of interdependent operations. Yet even this simple path conceals a great variety of possible pitfalls. These must be detected, and safe paths around them clearly marked, if analysis is to meet minimal criteria of adequacy.

Critics of policy analysis cite as a reason for its frequent lack of influence the tendency of analysts to tackle questions of marginal interest to policy makers with models that produce unusable information. Serious abuses in problem identification and in building and using models, particularly large-scale multipurpose computer models, are discussed by E. S. Quade in his chapter, "Pitfalls in Formulation and Modeling." Quade first charts a path for avoiding the pitfalls encountered in turning a problem situation into something tractable and relevant—traps such as starting model building and calculation with undue haste, unquestioning acceptance of stated goals, or measuring what is easily measurable instead of what is important. In model building and model use, some pitfalls to avoid are viewing the models as more important than the question, adopting academic rather than policy goals, attempting to duplicate reality, and trying to substitute a model for the decision maker.

The pitfalls that await analysts who collect and process data and the policy makers who are the consumers of the information produced from that data are the subject of Stanley Altman's "Pitfalls of Data Analysis." He emphasizes that data analysis provides the user with evidence from which conclusions can be drawn, not with conclusions themselves.

Often, some of the most important information produced by an analytic study is information about costs. It is commonly imagined that cost estimation is both objective and simple, at least in comparison with the estimation of benefits. But if the correct measure of the costs of an alternative is the forgone benefits of the next best alternative, then, Robert Bickner argues in "Pitfalls in the Analysis of Costs," costs, like benefits, are inevitably a matter of conjecture. Hence good cost analysis requires imagination, as well as caution in using traditional concepts of cost.

Among the tools and methods used in analysis, optimization techniques occupy a position of special significance. The chapter by Martin Beale presents a classification of optimization techniques into five groups and an overview of their strengths and weaknesses. One serious pitfall of large-scale optimization models, Beale points out, is that the detailed formulation and implementation can easily become unintelligible to anyone except their authors, thus greatly reducing the value of the models as evidence. A first step away from this pitfall is the adoption of a systematic way of documenting models, and the author makes some positive suggestions in this direction.

The second group of chapters is concerned with issues relating to the effectiveness of analysis and external criteria of evaluation. Decision makers are often dissatisfied with the usefulness of analytic studies not so much because the work is incompetent as because it appears misguided and irrelevant. Laurence Lynn, Jr., introduces the user's perspective in the seventh chapter of this volume. After a short discussion of different conceptualizations of decision making in the public sector, Lynn proceeds to an identification of pitfalls in analysis from the perspective of a government official facing a complex policy

decision. The central conclusion from this survey is that analysis often does not matter in decision making because analysts do not know what matters to decision makers.

Good communication is an essential condition of social effectiveness. Communication is defined by Arnold Meltsner in his "Don't Slight Communication: Some Problems of Analytic Practice" as the act of gaining acceptance and asserting a claim on another person's beliefs. While accepting inadequate information is a serious problem for the client, a more familiar problem, and the author's chief concern, is the domination of the social aspects of the analytic process by the substantive: high-quality studies are rejected because communication is slighted. Meltsner discusses errors involved in three aspects of communicating analysis: the choice of audience (recipients), the argument (content and form), and the means (technique and media).

For policy effectiveness, the objectives, as listed by Eugene Bardach in his essay "On Designing Implementable Programs," are (a) to achieve the goals defined by the policy; (b) to hold delay to a reasonable level; and (c) to hold financial costs to a reasonable level. These goals are not easily attained, for we understand only imperfectly the limitations on those institutions and instruments we normally rely upon to implement policy. Bardach analyzes the most common obstacles to successful implementation and makes some positive suggestions about how they may be overcome.

Systems analysis is now increasingly employed in many countries for program evaluation, i.e., for assessing a program's past or current performance, usually to determine the impact or effectiveness of that program. Harry Hatry, in "Pitfalls of Evaluation," discusses the many pitfalls peculiar to this particular application of systems analysis. These include letting people whose work is being evaluated, in effect, evaluate themselves; not looking outside the program for other plausible explanations of the effects; and focusing only on the overall (average) results without considering the component effects.

In the final essay of the volume, Kathleen Archibald sketches some of the ambiguities and pitfalls of the language analysts use. Archibald shows how the meaning of analysis itself has changed as analysts increasingly face phenomena of culture and behavior that do not fit easily into the standard logical or empirical modes of explanation. Yet the language of systems analysis still derives largely from these two modes of explanation. By loosening up the working language, analysis becomes more capable of coping with today's "wicked" problems.

We have not attempted to treat all pitfalls, or even all classes of pitfalls, associated with systems analysis. There is, for instance, no thorough discussion of the pitfalls associated with the problem of choice. The difficulties with criteria, particularly ratio criteria and the use of wrong concepts of cost and gain, however, have been well treated elsewhere by Hitch[3] and Hitch and McKean.[4] Similarly, we think that pitfalls of statistical inference and

econometric analysis are sufficiently well covered in the literature and do not require detailed treatment in this volume.[5] Other omissions are the class of pitfalls associated with the treatment of uncertainty, with the use of suboptimization, and with managing a study. Good discussions of these topics can be found in Hitch and McKean,[4] Quade,[6] and Holling.[7] We have also included no discussion of the pitfalls associated with the use of analysis itself—delays in reaching decisions, the tendency for the recommendations to call for increased centralization and concentration of decision making in top-level staff, increased dependence on complex processes (for example, computerized information systems) that require continuous attention by expensive talent to work well, and even the elimination of inefficiency and redundancy that often serves to meet unexpected contingencies.

A discussion centering on pitfalls is necessarily critical of widespread methods and practices. This may leave a negative impression on the reader, tending to obscure the substantial accomplishments of systems analysis. To minimize this danger, we should perhaps add a few words about the positive uses to which such a discussion may be put. Because of the nature of the work, it is impossible to eliminate pitfalls from systems analysis. But the experience of mature disciplines and professions shows that pitfalls can be largely avoided once they are identified, and standard methods of avoiding them become part of the craft knowledge of practitioners. Also, carefully selected examples of pitfalls, like good counterexamples in the teaching of mathematics, are an effective way of introducing students of systems analysis to the more subtle aspects of the craft. More generally, as Ravetz points out,

> A recognition and systematic use of the phenomenon of pitfalls might be very effective in the teaching of those simple but essential craft skills which are involved in scientific, scholarly, or administrative work. An exposition of standard techniques in terms of the pitfalls they are designed to circumvent, with examples, could go far to make them meaningful and obviously worth mastering.[8]

Avoidance of pitfalls guarantees minimal standards of quality, nothing more. It does not imply originality, depth, or any other of those intangible qualities that distinguish the brilliant from the merely competent study; nor can it ensure the success of the proposed solution. And yet, even meeting these minimal standards is not an easy task in a world of increasing complexity, pervasive uncertainty, and fallible knowledge. This volume will have achieved its goal if it succeeds in suggesting some useful ways of fulfilling this task.

NOTES AND REFERENCES

1. In this volume, systems analysis and policy analysis are used as essentially synonymous terms for the same activity: research that attempts to bring modern science and technology to bear on society's problems, seeking ameliorative solutions. These analyses search for feasible

courses of action, generating information and marshalling evidence about the benefits, costs, and other consequences that would follow their adoption and implementation, in order that the most advantageous action may be chosen.

2. McDermott, W. 1977. Evaluating the physician and his technology, pp. 135–157. *In* Doing Better and Feeling Worse (J. H. Knowles, ed.). New York: Norton.
3. Hitch, C. J. 1958. Economics and military operations research. Review of Economics and Statistics 40. 199–209.
4. Hitch, C. J., and R. N. McKean. 1960. The Economics of Defense in the Nuclear Age. Cambridge, Massachusetts: Harvard University Press.
5. See, for instance, Box, G. E. P. 1966. Use and abuse of regression. Technometrics 8: 625–629.
   Johnston, J. 1973. Econometric Methods. New York: McGraw-Hill.
   Reichman, W. J. 1961. Use and Abuse of Statistics. Harmondsworth, Middlesex: Penguin.
   Streissler, E. W. 1970. Pitfalls in Econometric Forecasting, London: The Institute of Economic Affairs.
   Wallis, W. A., and H. V. Roberts. 1956. Statistics: A New Approach. Glencoe, Illinois: The Free Press.
6. Quade, E. S. 1975. Analysis for Public Decisions, New York: American Elsevier.
7. Holling, C. S. (ed.). 1978. Adaptive Environmental Assessment and Management. Chichester: Wiley.
8. Ravetz, J. R. 1973. Scientific Knowledge and Its Social Problems. Harmondsworth, Middlesex: Penguin, p. 100.

2 An Anatomy of Pitfalls

Giandomenico Majone

A pitfall is a conceptual error into which, because of its specious plausibility, people frequently and easily fall. It is "the taking of a false logical path"[1] that may lead the unwary to absurd conclusions, a hidden mistake capable of destroying the validity of an entire argument. Logicians distinguish between a fallacy—the violation of a logical principle—and a simple falsity or error of fact. A single statement may be false, but what is fallacious is the transition from inadequate premises to a conclusion. To fall into a fallacy is to make a mistake in reasoning, not in what is reasoned about.[2] Similarly, in systems analysis, pitfalls should not be confused with blunders or factual mistakes that may affect, say, the numerical value of a solution but do not vitiate the basic structure of the underlying argument.

The earlier literature of systems analysis has devoted considerable attention to the treatment of pitfalls, and some of these discussions have attained the status of minor classics of the field.[3] The present chapter attempts to extend previous work in two ways: by categorizing analytic pitfalls in terms of the detailed operations by which a problem is investigated and solved; and by examining in some depth their conceptual roots, particularly in the case of pitfalls whose newly recognized significance reflects the complexity of the tasks facing systems analysts today.

SYSTEMS ANALYSIS AS CRAFT WORK

The real significance of the category "pitfall" is best appreciated in connection with the craft aspects of analytic work. These aspects have been neglected in methodological discussions primarily concerned with establishing the scientific status of the systems approach, or its usefulness as an aid to decision making.

To justify the claims to scientific status, methodologists have stressed attributes, like quantification, objectivity, and formalization, that could be used

as criteria to demarcate systematic analysis from informal analysis by intuitive judgmental processes. Such distinctions may have served a useful role in the early stages of development of the field, but they overlook the fact that even the most theoretical forms of knowledge presuppose an extensive matrix of personal judgments and practical skills. In the words of Jerome Ravetz, "without an appreciation of the craft character of scientific work there is no possibility of resolving the paradox of the radical difference between the subjective, intensely personal activity of creative science, and the objective, impersonal knowledge which results from it."[4]

On the other hand, the role of systems analysis in assisting choice is generally discussed in a language that tends to obliterate the distinction between analysis and decision making. The same categories—objectives, alternatives, prediction, control, optimization, and so on—are used in both cases, thus setting up a one-to-one correspondence between the stages of the decision process and the stages of the analytic process. This approach has considerable intuitive appeal and economizes on terminology, but it has several disadvantages. First, it mixes description with prescription: what purports to be a description of analytic operations immediately becomes a prescription for rational decision making. Second, the decision-making paradigm is modeled on one particular mode of inquiry—normative prospective analysis. For other, equally legitimate uses of analysis—diagnosis, evaluation, and advocacy, for example—the categories of decision making are not particularly relevant. Finally, and most important for the present discussion, this paradigm emphasizes outcome at the expense of process. Practical decisions are evaluated in terms of actual results, not by their compliance with procedural or professional standards. Evaluation by results makes sense when outcomes can be unambiguously classified as either success or failure, as in the case of goods and services sold on the market. What counts in the market are the ultimate characteristics of the goods and services offered for sale, not the process by which they are produced. The situation is almost reversed in the case of professional work. Because success depends on many factors outside the control of the professional, and because the causal link between process and outcome is indirect and uncertain, process becomes intrinsically important. In health care, for example, "[m]uch, perhaps most, of what a physician does must be categorized as process, and process not even calculated to affect outcome."[5] Consequently, the criteria of evaluation used in all the mature professions rely heavily on procedural considerations and craft judgments of adequacy.

In the case of systems analysis, the problem of evaluation is, if anything, even more delicate. For here it is necessary not only to distinguish between the quality of an argument and the outcomes of a decision, but also to avoid the pitfall of relying on the same standards of quality that are used in some of the older academic disciplines, like physics or mathematics. Analytic arguments cannot be shown to be true in the sense of mathematical theorems or empirical

propositions, but only to be adequate in relation to the special features of the problem, the quality of the data, the limitations of the available tools, and the needs of the client. Such criteria of adequacy are derived from craft experience, not from abstract logical canons.

Craft judgments play a significant role at every stage of analysis. Formalistic accounts of analytic methodology may give an idea of the wealth of tools and techniques that are available today, but they have little to say about how choices are made among alternative assumptions, data, and methods of analysis. These crucial decisions are not guided by theory (since usually no theory is available), nor merely by intuition (which some methodologists use as the *deus ex machina* for anything that cannot be reduced to rule and algorithm), but by the craft knowledge of the analyst.

The notion of craft suggests an analogy between the work of the analyst and that of the traditional artisan. In both cases, successful performance depends on an intimate knowledge of materials, tools, and processes, and on a highly personal relationship between agent and task. Good analytic work cannot be produced mechanically any more than handicraft can be mass-produced. Style plays as big a role in determining the value and acceptability of the analytic product as it does in the results of the artisan's work. This analogy, like all analogies, cannot be pushed too far. There are obvious differences: the artisan uses concrete materials in order to produce objects having an appropriate shape and serving well-defined purposes; the analyst operates with data, technical tools, and conceptual constructs to produce arguments supporting certain conclusions. Despite these differences, the classical Aristotelian analysis of the craftsman's work in terms of four basic constituents (material, efficient, formal, and final causes, in Aristotle's terminology) may be usefully extended to intellectual crafts.[6] In the case of systems analysis this leads to identifying the material component of the analyst's task with the data, information, and other conceptual inputs used in formulating a problem. The technical tools and methods used by the analysts represent the efficient component, while the formal component is an argument in which evidence is cited and from which a conclusion is drawn. The final component is the conclusion itself, with the related activities of communication and implementation. Accordingly, in the following discussion, pitfalls of analysis will be grouped, in correspondence with the basic constituents of the analyst's work, as follows: (a) problem setting, (b) data and information, (c) tools and methods, (d) evidence and argument, and (e) conclusions, communication, and implementation.

PROBLEM SETTING

Analysis usually starts with something less structured than a problem, namely, a problem situation. This is an awareness that things are not as they should be,

but without a clear understanding of the nature of the difficulties and of how they might be solved. Problem setting is the intellectual process by which a problem situation is translated into a specific problem. Typical stages of this process are systems definition, systems diagnosis, and goal setting.

Since the questions asked and the methodological decisions taken in effective formulation of a problem shape the final conclusions, one can speak of a phenomenon of preselection—a term introduced by Gerald Holton in discussing Plato's formulation of the problem of planetary motion. In moving from problem to solution, Holton writes, "There are many constraints which have been placed from the beginning on the possible solution in an almost imperceptible manner; constraints on the 'facts,' on the hypothesis, and the correspondence between analogon and facts by which we are relieved of the feeling of puzzlement, by which we 'understand.' "[7]

Often these constraints are the most difficult ones to detect because they are deeply embedded in the cultural, institutional, and professional backgrounds of client and analyst. For example, it seems "natural" for an environmental agency working on a specific pollution problem to consider only alternatives falling within its jurisdiction. Even the competent analysts of the Delaware Estuary Comprehensive Study (DECS) chose to confine their inquiry to the Delaware estuary for which they were directly responsible instead of considering other and probably better alternatives for outdoor recreation.[8] This exemplifies the phenomenon of suboptimization and the related pitfall of assuming that the scope of analysis should coincide with the scope of decision-making authority.[9]

The next task facing the DECS analysts was to define the nature of the pollution problem on the Delaware River. Traditionally, the amount of dissolved oxygen (DO) in the water has served sanitary engineers as the principal indicator of water quality. It is therefore understandable that the DECS engineers would take recourse to the received wisdom of their profession. But this technical decision had major policy consequences. Once DO was accepted as the relevant indicator, the section of the river experiencing the most serious pollution problem and urgently requiring public action was automatically identified: it was the heavily industrialized region between Philadelphia and Wilmington that was suffering the most acute oxygen sag. Although there was no reason to believe that raising DO to economically feasible levels in this critical section of the river would make the Delaware more suitable for recreational and other uses, this became in fact the focus of the technical analyses. And because the economists responsible for the cost–benefit analyses accepted the engineers' formulation of the problem uncritically, the policy alternatives from which the agency had to choose came to be expressed in terms of different DO levels.[10] Incidentally, this instructive example reveals another pitfall whose importance has not been sufficiently recognized in the literature. It is often said that any analyst who accepts the client's initial formulation of the problem uncritically is

heading for disaster. The case of the DECS shows that it is an equally serious mistake to accept uncritically the initial problem formulations of technical experts.

The same lesson can be learned in many other policy areas. Thus, the assumption that medical needs can be determined objectively on medical grounds has often led health planners to approach health service decisions as problems of ascertaining and meeting arbitrarily predetermined needs of the community, rather than as problems of allocating scarce health resources among competing programs. The conception of sickness as an objective and absolute state is largely responsible for the absence of effective rationing mechanisms in the existing national health systems.[11] Similar pitfalls in military planning and systems analysis have been discussed at length by Hitch and McKean, who also point out other common errors in the selection of choice criteria.[12]

An important but often neglected aspect of problem setting is determining who constitutes the affected group with respect to the policy in question. A narrow focus on the direct impact of a proposed policy on the decision maker or the client organization leads to overlooking externalities (such as pollution and congestion) that may cause more serious problems than those the original decision attempted to solve.

In defining a system, only a few of the elements and interrelationships that could be considered can in fact be included. The necessity of preselection and, at the same time, its inherent dangers become particularly obvious when constraints are introduced in order to make the analytic problem more manageable. The reasoning behind such postulated constraints must be examined very carefully, for poorly chosen constraints can easily lead to unacceptable solutions or to dramatic increases in systems costs. There is, for example, the often-cited case of a weapon-systems study in which an arbitrary limitation on acceptable casualties led to solutions in which 100 million dollars was being spent, at the margin, to save a single life. Many more lives could probably have been saved with the same amount of resources: the opportunity cost of the stipulated constraint was high indeed. In some cases it is possible to calculate such implicit costs by varying the constraints parametrically. Even when formal calculations are not possible, a rough estimate of the implications of different feasibility conditions may be enough to avoid serious errors of judgment.

Because of the unstructured character of a problem situation, the production and interpretation of data at this stage of the analysis present peculiar problems. Thus, information obtained through opinion polls or attitude surveys is particularly sensitive to the way the questions are framed, which in turn may depend on how the problem is conceived. The distinction between facts and artifacts, always a tricky one when social and economic data are involved, becomes particularly difficult when one tries to find out what the problem is.

To say, for example, that 4.7% of the labor force is unemployed is only to say that the result of applying certain operations embodied in a questionnaire and the answers thereto gave 4.7%. Any economist knows that this number is sensitive to the questionnaire: a simple change in one question may produce a change of half a million in the number unemployed.[13]

Similarly, a reported drop of one or two percentage points in a country's gross national product used to be viewed as a most significant indication of a serious problem requiring immediate government action. In fact, national income figures are known only up to an error of ± 10 to ± 15 percent, and comparable or greater uncertainty is present in foreign trade, price, unemployment, and growth statistics.[14]

Finally, in defining a policy problem one should avoid formulations that are either too broad or too narrow. As some of the examples discussed above suggest, the second mistake is probably the more common one among analysts. One reason for this is the mistaken assumption that the amount of detail that is useful at the problem-setting stage must be the same as that needed in the phase of problem solving. Actually, even the styles of thinking differ somewhat in the two cases, with imagination, judgment, and analogical reasoning playing a larger role in formulating a problem than rigor and technical skills.

DATA AND INFORMATION

Once the policy problem has been specified, new types of data come into play, and with them come new types of possible pitfalls. Consider some of the data problems of econometric analysis. Since model builders do not usually collect the primary data but rely on figures from a variety of sources, problems of interpretation can be quite serious. Social and economic statistics are collected according to classification schemes that are often dictated not by logical or theoretical considerations, but by expediency—the availability of data, the feasibility of making estimates, the operating procedures of the data-gathering office. The *ad hoc* nature of much socioeconomic accounting is seldom publicized by the data collectors and is often overlooked by the data users. Professor Streissler tells the story of the econometrician who made the rather surprising discovery that in Austria profits rose, over a certain period, exactly as much as wages and salaries. He had apparently forgotten that in the early days of Austrian national income accounting, profits were *assumed*—for lack of data—to move in proportion to wages and salaries.[15]

Problems of interpretation are compounded by the fact that the data utilized by the analyst have already gone through several stages of manipulation. Alan Coddington has vividly described this process:

Economic statistics are the result of the bureaucratic compounding of enormous quantities of fragmentary and even ambiguous pieces of information: the components are

thrown up as a result (or even as a by-product) of the general administrative processes of society, from tax-returns, sample surveys, censuses, and so on; the components are assembled and aggregated by teams of statisticians who have not themselves collected the information.[16]

Sometimes the analyst is in the fortunate position of being able to collect the primary data, or at least of suggesting an appropriate sampling design. Even then, the original data are usually much too raw to be used in a model or in an analytic argument without first being refined into a more reliable and useful form. Such refining requires craft skills that are rather different from those used in problem setting and data collection. This new phase of analysis, the transformation of data into information, is exemplified by a number of familiar operations, such as calculating averages and other statistical parameters, constructing indices, or fitting a curve to a set of points. The operations performed on the original data may be technically involved or quite simple, but they always represent a crucial step. From here on, the primary data are no longer used, except for an occasional check, and analysis is carried out exclusively in terms of intellectual constructs—with all the advantages and potential pitfalls of abstract language.

Two basic craft judgments are involved in the production of information: that the equation or model fits the data sufficiently well, and that this particular transformation of the data, out of the many theoretically possible, is the significant one. There are, of course, standard statistical tests for judging goodness of fit, but a purely statistical validation provides weak evidence in the absence of good theoretical reasons for the choice. It is always possible to try empirical equations in succession until one is discovered that fits the data to any desired degree of accuracy. Alternatively, one can always improve the fit simply by adding to the model variables exhibiting the requisite type of behaviour—although the substantive significance of such variables may be quite doubtful. These possibilities explain the frequency of alleged discoveries of beautifully simple laws governing an amazing variety of social and natural phenomena. William Feller makes the point with his usual brilliance:

An unbelievably huge literature tried to establish a transcendental "law of logistic growth." ... Lengthy tables, complete with chi-square tests, supported this thesis for human populations, for bacterial colonies, development of railroads, etc. Both height *and* weight of plants and animals were found to follow the logistic law even though it is theoretically clear that these two variables cannot be subject to the same distribution. Laboratory experiments on bacteria showed that not even systematic disturbances can produce other results. Population theory relied on logistic extrapolations (even though they were demonstrably unreliable). The only trouble with the theory is that not only the logistic distribution but also the normal, the Cauchy, and other distributions can be fitted to the *same material with the same or better goodness of fit.* In this competition the logistic distribution plays no distinguished role whatever; most contradictory theoretical models can be supported by the same observational materials.[17]

The second kind of judgment, concerning the relevance of a particular piece of information, is more delicate than any goodness-of-fit test. Quade gives an entertaining example of a pitfall involving a judgment of relevance: the use, during World War I, of the arithmetic instead of the harmonic mean in computing turnarounds of troop and cargo ships of different speeds.[18] The mistake is admittedly rather trivial, though fairly common even today, but it is precisely its elementary character that shows how easy it is to stumble into pitfalls in even the simplest aggregation of data. As a more advanced example, one can mention the so-called "ecological fallacy," well known to statisticians and social scientists but less familiar to systems analysts despite its importance, particularly in regional and urban planning. The pitfall consists in using ecological correlations (i.e., statistical correlations in which the variables are properties of groups, such as percentages or means for each of the 50 states of the United States) as substitutes for the usual individual correlations. But ecological correlations are not appropriate substitutes for individual ones since, as Robinson has shown,[19] the two correlations in general are different (they may even differ in sign), and the values of ecological correlations depend strongly on the type of grouping used.

There are few formal tests of relevance comparable to the statistical tests of goodness of fit, and these few are of limited use in systems analysis.[20] Theoretical considerations can provide some guidance, but basically the judgment of whether a curve, an equation, or a model is of the right type depends on the craft experience of the analyst.

TOOLS AND METHODS

The tools of systems analysis may be roughly classified according to their function in the production, manipulation, and interpretation of data. Having discussed some of the pitfalls encountered in the production and manipulation of data, we can go on to consider the category of interpretive tools, which includes tool disciplines like mathematics, statistics, and economics that are constantly needed in analytic work.

This heavy dependence on tool disciplines creates special problems for the teaching as well as for the practice of systems analysis. Today, students of systems analysis are exposed to academic curricula that vary from school to school and that represent tentative compromises among different intellectual traditions. Sometimes the compromise is reached more at the level of language than of deep understanding. There results a sort of jargon that, because of its composite character and lack of depth, tends to mask ambiguities and subtle differences in meaning that are at the root of many pitfalls. Also, concepts and techniques removed from their broader disciplinary context quickly become stereotypes, and their limitations are not easily perceived by people primarily interested in immediate applicability.

The danger of misuse of the formal tools of analysis is made particularly acute by the "prevailing metaphysics,"[21] according to which the scientific character of a field is assumed to be in direct proportion to the degree of its mathematical or statistical formalization. In this perspective, quantification and algorithmic elegance almost become ends in themselves, not only at the cost of what cannot be easily quantified but also at the expense of a deeper understanding of the substance of the problem. A related point has been raised by Myron Tribus. He argues that the statistical methods routinely employed in the study of cloud-seeding experiments are unsatisfactory because they assume that all that counts are the experimental results while paying insufficient attention to the elucidation of the underlying physical mechanisms.[22] His conclusions have general relevance for systems analysis: when the system under study is complex and highly dependent on what happens at a particular branch point, one cannot learn much about systems behavior by merely looking at some simple measures of the outcome; statistical and other formal techniques of analysis cannot replace detailed investigations of the fine structure of the system.

In disciplines with a long intellectual tradition, the introduction of new tools usually opens up lines of research that were previously inaccessible. In newer fields, technical innovations often give rise to the phenomenon of "new-toolism," a disease to which systems analysts seem to be particularly predisposed. Those affected by this disease become "possessed of and by new tools (various forms of mathematical programming, vast air-battle simulation machine models, queuing models and the like), and they look earnestly for a problem to which one of these tools might conceivably apply."[23]

Formalism, new toolism, and mechanitis—"putting . . . machines to work as a substitute for hard thinking"[24]—are characteristic symptoms of a stage of disciplinary development in which generally accepted criteria of adequacy are still lacking. Powerful methods developed in different contexts and by other disciplines are adopted enthusiastically and uncritically, only to be abandoned after a few years when the hoped-for results fail to materialize.[25] In the effort of bringing the field toward maturity, the temptation to adopt the criteria of adequacy of successful academic disciplines, along with their methods, is understandably strong. But experience shows that such imitations are seldom fruitful and if carried far enough may lead to the production of vacuous results. The reason is that the criteria of adequacy of an intellectual craft are intimately related to the characteristic pitfalls of the problems investigated by its practitioners, and these pitfalls are very particular to the data, tools, and patterns of arguments used in that field.[26]

The composite nature of systems analysis, in which descriptive, prescriptive, evaluative, and advocacy elements are so intertwined as to be virtually inseparable, increases the difficulty of establishing useful criteria of adequacy; this is shown by the abortive attempts to define principles of professional practice that might provide guidance to analysts involved in policy arguments.[27]

EVIDENCE AND ARGUMENT

The argument is the link connecting data and information with the conclusions of a study. The structure of an analytic argument is typically a complex blend of methodological considerations, factual statements, interpretations, evaluations, and recommendations. Along with mathematical and logical deductions, it will include statistical and empirical inferences, guesses, references to previous studies and expert opinions, discussions of alternative criteria of choice, and caveats and provisos of different kinds. This unavoidable complexity rules out the possibility of any formal testing, and hence any possibility of proving or refuting the final conclusions. Whatever testing can be done at this stage of the work must rely on a variety of disciplinary standards, corresponding to the different techniques and methods used in the argument, on the plausibility and robustness of the results, and on the client's criteria of usefulness.

The nature of the evidence plays a crucial role in the process, since a wrong assessment of the strength and fit of evidence before it is included in the argument can lead to pitfalls in the drawing of conclusions. As the term is used here, evidence is not the same as data or information. Rather, it is information selected from the available stock and introduced at a specific point of the argument in order to persuade a particular audience of the truth or falsity of a statement. An inappropriate selection of data or models, their placement at a wrong point in the argument, a style of presentation inappropriate to the audience at which the argument is directed—any of these can destroy the effectiveness of information as evidence, regardless of its intrinsic cognitive value. Hence, criteria of assessment of evidence are different from those applicable to "facts." While facts can be evaluated in terms of general logical canons, the acceptability of evidence depends on a number of factors peculiar to a given situation, like the nature of the case, the type of audience, the prevailing "rules of evidence," and even the persuasiveness of the analyst.

The category of evidence has received little attention in the natural sciences, and, perhaps because of this, it has also been neglected in methodological discussions of systems analysis. But the two situations are different in this as in other respects, and what is justifiable in the one case is a serious pitfall in the other. Ravetz points out that neither descriptive nor theoretical natural sciences require highly developed skills in testing evidence, beyond the tests already involved in producing information, for here one usually has either a large mass of information with a relatively simple argument or a complex argument needing evidence at only a few points. However, there are fields where problems typically involve both complex arguments and large masses of data and where the reliability and relevance of information itself cannot automatically be trusted. Law and history are two such fields, and here evidence has been explicitly recognized as an autonomous conceptual category.

Although complexity of argument and large amounts of data of doubtful reliability are also characteristic features of systems studies, suitable rules of evidence have not yet been developed by the analytic profession. However, some of the preceding discussion is relevant in this context. For example, in judging the acceptable level of accuracy of a set of data used as evidence, one must keep in mind that different standards apply to different kinds of data. After all, even in the natural sciences an error of a billion years for estimates of the age of the earth is found acceptable, while a basic physical constant may be quoted with a possible error of, say, 10^{-14}.

Again, the contemporary fashion of using mathematical arguments on every possible occasion produces a tendency to accept numerical results as facts rather than evidence. The sensitivity of those results to changes in the data, in the model, or in the estimation procedures is easily forgotten. Worse still, we do not even know which of these changes may have a greater effect on the results. Thus, one study comparing two estimation procedures for a five-equation econometric model using two sets of data concludes that "[t]he variances due to the estimation method are greater than those due to the data revisions, which indicates that the choice of estimating procedure has more effect on the parameter estimates than the choice of data."[28] On the other hand, the authors of a study also comparing ordinary least squares with two-stage least squares, but for a much larger model and using three sets of data, find that "[v]ariations in parameter estimates are generally much greater between different sets of data than between different methods of estimation, at least for the model and methods used in this paper."[29]

The value of large-scale models as evidence is often very doubtful. Poor documentation makes it almost impossible for anyone but the modelers themselves to reproduce the results. On the other hand, a complete documentation would be too voluminous to be properly digested even by an expert evaluator. Sometimes a model is kept proprietary for commercial reasons and the customer is allowed to see the results but not the assumptions.[30] Summing up the state of the art as it stood a few years ago, Lee writes:

> Large models are not simply constructed and operated; they must be "massaged" into being, first to make them operate at all and then to get sensible output. Inevitably this requires numerous special features in the computer program that keep the model from going out of bounds, so that output described as "reasonable" is a self-fulfilling tautology. The model produces reasonable results because its builders imposed constraints on the model's operation that prevented it from producing anything else. Because the models contain large but unknown amounts of error and they are too complex, and there are no evaluation measures, modelers have little choice except to fudge the models into shape.[31]

Like legal arguments, analytic arguments cannot be properly evaluated unless they are set out in a sequence of steps conforming to certain basic rules of procedure. The evaluation of evidence is an activity that always involves for-

malities. The legal scholar asks: What different sorts of propositions are uttered in the course of a law case, and in what different ways can such propositions and the supporting evidence bear on the soundness of a legal claim? Without the large number of distinctions recognized by legal doctrine (statements of claims, different kinds of evidence, testimony, proof of facts, presumptions, and so on) it is impossible to understand the nature of the legal process.[32] To adapt the legal analogy to our case, reliance on mathematical modeling and other analytic techniques is not a guarantee that an argument is "in proper form." A mathematical style of presentation is not incompatible with a black box approach, and may even encourage it. No amount of technical virtuosity can compensate for carelessness in drawing the necessary distinctions between assumptions, data, facts, evidence, and conclusion, or for a lack of sophistication in understanding their respective roles in the overall economy of an argument.

CONCLUSIONS, COMMUNICATION, IMPLEMENTATION

The conclusion of an analytic study may be a forecast, the clarification of an issue, a recommendation, an assessment of ongoing policies, a new idea, or a different perspective on an old policy problem. Whatever its nature, a conclusion is not concerned with directly perceived "facts," but with the intellectual constructs that have served as the objects of the argument. A different conceptualization of the problem, a different choice of analytic tools, a few different assumptions made at crucial points in the argument can easily lead to quite different conclusions. Thus the contact with the external world is always indirect and elusive. This is true for any kind of intellectual inquiry, including the natural sciences. But in science the pitfalls encountered when a conceptual system makes contact with reality can be detected, before too much harm is done, by various means, including controlled experiments, that reduce the abruptness of the impact. In systems analysis, where the possibilities of experimentation are severely limited, the lack of certainty of the conclusions lies at the root of the deeper problems of communication and implementation.

The current emphasis on implementation problems is a useful counterbalance to the excessive reliance of many analysts on normative decision models and on the pure logic of choice, on analysis rather than on synthesis and design. The decision theorist does not need to distinguish between decision and action: if the decision does not lead to the corresponding act, it is because something occurred to prevent it, and a new decision problem arises.[33] But this is a purely formal solution, of no help whatsoever for designing policies with fewer things that can go wrong. As a stage of analysis, implementation may be thought of as a gedankenexperiment performed by the analyst in an attempt to recognize in advance the many pitfalls and obstacles on the path that leads from decisions and plans to deeds and results.[34]

The contact between a conceptual system—the conclusions of an analytic study—and external reality is also problematic in another sense that corresponds to the second meaning "implementation" has in systems analysis: the ability of an investigation to improve systems behavior by redirecting or reinforcing the beliefs and attitudes of policy makers, managers, operators, and other relevant audiences. Here we are faced with a dilemma that, although it turns out to be more apparent than real, can easily confuse the analyst and thus represents a serious potential pitfall. Rational arguments are only partial determinants of attitudes: an element of persuasion is involved in any attempt to suggest one rather than another view, judgment, or course of action. Thus in order to be effective, an analyst must also be an advocate. But he is also a firm believer in the virtues of the scientific method, and this belief is generally associated with a distaste for problems of communication and persuasion, while favoring the pitfall that Kahn and Mann[35] have called hermitism.

One way to defuse the conflict between practical effectiveness and scientific integrity is to note that many outstanding scientists have not been loath to use persuasion when the situation seemed to require it. Thus, eminent historians of science like Duhem and Koyré have likened the work of Galileo to propaganda.

> But propaganda of this kind is not a marginal affair that may or may not be added to allegedly more substantial means of defence, and that should perhaps be avoided by the "professionally honest scientist." In the circumstances we are considering now, *propaganda is of the essence.* It is of the essence because interest must be created at a time when the usual methodological prescriptions have no point of attack; and because this interest must be maintained . . . until new reasons arrive.[36]

Discussing Adam Smith's principles of division of labor and free exchange, two well-known economists write:

> It is interesting that Smith's book did not contain a logically correct exposition; instead it contained a masterfully persuasive statement of the results of free exchange. It was Robert Torrens, who some forty years after the idea had been "sold," demonstrated its logical validity. Possibly, had Smith tried to give a logically air-tight demonstration, instead of a suggestive plausible interpretation, he would never have made his "point" popular.[37]

George Stigler adds Mill, Jevons, and Böhm-Bawerk to the list of famous economists who "have employed the techniques of the huckster."[38]

In systems analysis, as in science and in everyday reasoning, few arguments are purely rational or purely persuasive. A careful blend of reason and persuasion is usually more effective than exclusive reliance on one or the other. Style, elegance of expression, novel modes of communication (like Galileo's brilliant use of Italian instead of Latin) are often important means of winning support for a novel idea and overcoming preconceived hostility and institutional inertia. The practical question, therefore, is not whether to use persuasion, but which form of persuasion to use and when. It has been suggested that analysis should

be done in two stages: a first stage to find out what one wants to recommend, and a second stage to make the recommendation convincing "even to a hostile and disbelieving, but intelligent audience."[39] This is sound advice, as long as it is not interpreted as implying that communication and implementation are discrete and separable parts of analysis, rather than integral components of a continuous process. Of systems analysis, more perhaps than of any other intellectual activity, it can truly be said that "creation of a *thing*, and creation plus full understanding of a *correct idea* of the thing, *are very often parts of one and the same indivisible process* and cannot be separated without bringing the process to a stop."[40] For this reason, the conceptual and institutional separation of thinking from doing, of planning from deciding and implementing, must be considered the most serious pitfall of applied systems analysis.

NOTES AND REFERENCES

1. Koopman, B. O. 1956. Fallacies in operations research. Operations Research 4(4): 422–426.
2. Stebbing, L. S. 1959. Thinking to Some Purpose. Harmondsworth, Middlesex, England: Penguin, p, 157.
3. Koopman.[1]
   Hitch, C. J. 1958. Economics and military operations research. The Review of Economics and Statistics, 199–209.
   Kahn, H. and I. Mann. 1957. Ten Common Pitfalls, RM-1937. Santa Monica, California: The Rand Corporation.
   Quade, E. S. 1968. Pitfalls and limitations. *In* Systems Analysis and Policy Planning (E. S. Quade and W. I. Boucher, eds.). New York: American Elsevier.
4. Ravetz, J. R. 1973. Scientific Knowledge and Its Social Problems. Harmondsworth, Middlesex, England: Penguin, p. 75.
5. McDermott, W. 1977. Evaluating the physician and his technology. *In* Doing Better and Feeling Worse (J. H. Knowles, ed.). New York: Norton, pp. 135–157, 138.
6. On science as craft work, see Ravetz.[4]
7. Holton, G. 1973. Thematic Origins of Scientific Thought. Cambridge, Massachusetts: Harvard University Press, p. 22.
8. Ackerman, B. A., S. R. Ackerman, J. W. Sawyer, Jr., and D. W. Henderson. 1974. The Uncertain Search for Environmental Quality. New York: The Free Press.
9. Hitch, C. J., and R. N. McKean. 1967. The Economics of Defense in the Nuclear Age, New York: Atheneum. pp. 161–164.
10. For a fuller discussion of this example, see Ackerman *et al.*,[8] especially chapters 3 and 4.
11. For the British experience see, for example,
    Hagenbuch, W. 1958. Social Economics, Cambridge: Cambridge University Press, pp. 275–287.
    Feldstein, M. 1963. Economic analysis, operational research, and the national health services. Oxford Economic Papers (March): 19–31.
12. Hitch and McKean,[9] pp. 175–186.
13. Deming, W. E. 1969. Boundaries of statistical inference. *In* New Developments in Survey Sampling. (N. L. Johnson and H. Smith, Jr., ed.). New York: Wiley–Interscience, p. 656.
14. Morgenstern, O. 1963. On the Accuracy of Economic Observations. 2nd ed. Princeton, N.J.: Princeton University Press.

15. Streissler, E. W. 1970. Pitfalls in Econometric Forecasting. London: The Institute of Economic Affairs, p. 28.
16. Coddington, A. 1969. Are statistics vital? The Listener (11 December): 822–823.
17. Feller, W. 1966. An Introduction to Probability Theory and Its Applications. Vol. II. New York: Wiley, p. 52. I am indebted to Professor E. Tufte of Yale for this reference.
18. Quade, E. S. Analysis for Public Decisions. 1975. New York: American Elsevier, p. 299.
19. Robinson, W. S. 1950. Ecological Correlations and the Behavior of Individuals. American Sociological Review 15: 351–357.
For a brief discussion of the ecological fallacy in large-scale urban modeling, see
Lee, D. B., Jr. 1973. Requiem for large-scale models. American Institute of Planners Journal, May 1973, pp. 163–178.
20. I am thinking of such examples as Stevens' theory of "appropriate statistics," and the technical notion of "sufficient statistic." According to Stevens, operations on measurements on a given scale are appropriate if they are invariant under topologically permissible transformations of the scale. For further developments, see
Adams, E. W., R. F. Fagot, and R. E. Robinson. 1965. A theory of appropriate statistics. Psychometrika, 30(2): 99–127.
A clear understanding of the concept of sufficient statistic is useful for clarifying the nature of some common pitfalls of data analysis, like the use of the arithmetic instead of the harmonic mean in the example mentioned above.
21. Ravetz,[4] p. 122.
22. Tribus, M. 1970. A physical view of cloud seeding. Science 168: 201–211.
23. Wohlstetter, A. 1970. Analysis and design of conflict systems. *In* Analysis for Military Decisions (E. S. Quade, ed.). Amsterdam: North-Holland, p. 106.
24. Hitch.[3]
25. For a good discussion of the "rise and fall of the modeling movement" in urban planning, see Lee.[19] Similar examples can easily be quoted for practically all the subfields of systems analysis.
26. Ravetz,[4] pp. 148–159.
27. Operations Research Society of America Ad Hoc Committee on Professional Standards. 1971. Guidelines for the practice of operations research. Operations Research 19: 1123–1158.
Hugh Miser has made the important point that although the guidelines were drafted in 1970, they reflect experience gained in a military context, during and immediately after World War II. Their weaker parts are precisely those more directly relevant to the current practice of systems analysis. Compare
Miser, H. J. 1973. The scientist as adviser: The relevance of the early operations research experience. Minerva XI: 95–108.
28. Holden, K. 1969. The effect of revision of data on two econometric studies. The Manchester School, pp. 23–27 (quoted by Ravetz[4]).
The two estimation methods compared by this study are ordinary least squares and limited-information maximum-likelihood two-stage least squares.
29. Denton, F. J. and Y. Kuiper. 1965. The effect of measurement errors on parameter estimates and forecasts, a case study based on the Canadian preliminary national accounts. Review of Economics and Statistics 47: 198–206.
30. Greenberger, M., M. A. Grenson, and B. L. Crissey. 1976. Models in the Policy Process. New York: Russell Sage Foundation.
31. Lee,[19] p. 167.
32. Toulmin, S. 1964. The Uses of Argument. Cambridge: Cambridge University Press.
33. See, for example,
Lindley, D. 1971. Making Decisions. London: Wiley-Interscience.

34. See the chapter by Eugene Bardach in this volume (Chapter 9).
35. Kahn, H., and I. Mann. 1957. Ten Common Pitfalls. RM-1937. Santa Monica, California: The Rand Corporation.
36. Feyerabend, P. 1975. Against Method. London: New Left Books, p. 141.
37. Alchian, A. A., and W. R. Allen. 1974. University Economics. London: Prentice-Hall International, p. 200.
38. Stigler, G. J. 1955. The nature and role of originality in scientific progress. *In* Essays in the History of Economics. Chicago: The University of Chicago Press.
39. Kahn, H., and I. Mann. 1956. Techniques of Systems Analysis. RM-1829. Santa Monica, California: The Rand Corporation.
40. Feyerabend,[36] p. 26.

3 Pitfalls in Formulation and Modeling

E. S. Quade

Problem formulation, or problem setting, and modeling, particularly for predicting the consequences of a possible action, are closely related activities in systems analysis. Problem formulation, involving as it does the simplification of a subset of the real world selected for study, may, in fact, be considered part of the modeling process because the assumptions about the problem setting affect the predictions of the consequences, with the model being the analyst's conception of the factors and relationships significant to his investigation. Craft judgment, much more than theory, governs both activities.

The purpose of this chapter is to call attention to some of the pitfalls to be found in problem formulation and in modeling and to offer comments that may help others to recognize and avoid them.

FORMULATION

The initiation of a systems or policy analysis usually means that someone is dissatisfied with the current or projected state of affairs. He may, however, have only the vaguest ideas of where the "problem" lies. An analyst, seeking to help find ways to improve the situation, particularly when the issue is a public one, almost always finds himself facing not a "problem" but something between a cluster of problems (what the French call a *problématique*) and a mess.[1] To help, the analyst must identify the problem to be studied and define its scope in such a way that he has some hope of finding an acceptable and implementable solution within the economic, political, technological, and other constraints that exist, including the limitations imposed by the policy maker's span of control and the time available for decision.

Problem formulation is concerned with such things as determining the goals

or objectives to be achieved by a solution, setting boundaries on what is to be investigated, making assumptions about the context, identifying the target groups, and selecting the initial approach the analysis is to take. Much depends on the judgment of the analyst. Checkland (1978) points out one major reason for this:

> Problems . . . in the world are unbounded, in the sense that any factors assumed to be part of the problem will be inseparably linked to many other factors. Thus an apparently technical problem of transportation becomes a land-use problem, which is seen to be part of a wider environment-conservation problem, which is itself a political problem. Do these then become part of the original problem? Can any boundary be drawn? How can the analyst justify the limits that practicality forces him to impose?

But there are other difficulties and thus numerous opportunities for the analyst to go astray.

INSUFFICIENT ATTENTION TO FORMULATION

One great pitfall is to be in such a hurry to get started on the "real work" of analysis that one pays insufficient attention to formulation and ends up working on a problem that has little relation to the real issue, or on which no real progress can be made with the time and resources available, or in which no one has any interest. As Brewer (1975) remarks: "The simple question, 'What's the problem?' is often never asked at all by the analyst; it is assumed."

First, proper formulation is important. A large share of the effort of the leaders of a project (up to 40 percent is an estimate by a number of analysts; more, if we are to believe the old saying that "a problem well-put is half solved") needs to be invested in thinking about what problem is the right problem, exploring its proper breadth, and trying to discover what the objectives should be. Wellington (1887), a forerunner of modern systems analysis, clearly recognized the importance of problem formulation:

> As the correct solution of any problem depends primarily on a true understanding of what the problem really is, and wherein lies its difficulty, we may profitably pause upon the threshold of our subject to consider first, in a more general way, its real nature—the causes which impede sound practice; the conditions on which success or failure depends; the directions in which error is most to be feared. Thus we shall more fully attain that great prerequisite for success in any work—a clear mental perspective, saving us from confusing the obvious with the important and the obscure and remote with the unimportant.

Many analysts, trained in the quantitative techniques of operations research rather than in the craft aspects of policy analysis, have the feeling that until models are constructed and calculations are started nothing really analytic has been done. But for what purposes are the calculations to be carried out, and with what models? It is of very little use for analysts to compare alternatives that the policy makers cannot adopt because they involve action outside the

policy makers' sphere of influence or because they have features that any political observer could see to be unacceptable. The analyst, to be helpful, must select a formulation that is useful to the decision maker.

How the problem should be formulated depends on the use to which the analysis is to be put. For identifying policies that are feasible and desirable, one must be able to screen many variations rapidly and inexpensively (in terms of computer and human costs) in ways that are responsive to suggestions. If, on the other hand, the purpose is implementation, say of a particular means of transportation once the general type has been selected, the analysis must, and can, be detailed and concrete since the alternatives are ordinarily relatively similar and differ in detail rather than in concept. The pitfall is to take the latter approach when the former is wanted. If the decision makers really want the analysis as an excuse for inaction and delay or as a facade for promoting their preconceived ideas or if they do not care what is produced "... as long as they, the research sponsors, get credit for having been modern, management-oriented, and scientific ..." (Brewer, 1975), they may lead the analyst into this trap.

UNQUESTIONING ACCEPTANCE OF STATED GOALS AND CONSTRAINTS

Another pitfall is uncritical acceptance of the client's or policy maker's statement of his objectives and of the constraints on the actions to be considered.

(Incidentally, the opposite action would be an equally serious or even greater error—to pay no attention to the client's statement about what he wants and how he wants it. But it cannot be considered a pitfall, for it is not a trap into which an analyst could fall blindly; at the start of an analysis there are too many circumstances that keep the analyst's attention on the policy maker's words. Later, of course, memory may weaken, particularly if the policy maker loses contact with the study effort. Indeed, almost every pitfall we mention has an opposite number, which may, or may not, also be a pitfall, for to concentrate on any aspect of the analysis to excess is an error. Thus, to delay too long in starting a study would be an error, but it is not something an analyst is likely to do inadvertently and without good reason. Sponsors seldom want to allow time for a thorough job to be done anyway; hence there is always pressure to get started.)

To compound the difficulties with formulation, the clearly identified decision maker of the early days of military systems analysis, who was assumed (sometimes erroneously even in those days) to be a goal-oriented individual who actually made decisions that stuck by taking into consideration the probable consequences of his possible courses of action, has largely vanished. Instead, with today's spread of systems analysis into less disciplined areas, the policy maker or client is usually no more than a key participant in a process that involves numerous decision-making bodies, using the analysis for argu-

ment and evidence. Decisions may thus be taken several times or not at all. Public opinion has emerged as an intermediary between analytic results and the policy makers. As a consequence, packaging is often proving more significant than the product.

In fact, the analysis needs to be carried out and presented in such a way that its results are accepted by the multitude of decision makers, and in such a way that their decision is implemented in the way called for. To be fully successful, from the analyst's point of view, the analysis must provide help on how to do what should be done.

To help a decision maker, it is crucial to know what he wants. To be "rational" in his choice of action, the chooser has to know what he wants to achieve. A policy maker will usually have an idea of what he wants done, but his objectives may not be very well thought out; not only may they be so vague that no way exists to tell in the end whether they have been achieved or not, but they may also be inconsistent, encompassing clearly conflicting goals. Also, until considerable analysis has been done, the decision maker may not be aware that the implications of his objectives include things he does not want or ought not to want.

The point is that objectives cannot be taken as given but must be investigated. Means are sometimes confused with ends: a decision maker may say his problem is where to place a new comprehensive health center in his district, but his real goal may be to improve health services in his community. In that case, a better way to do so may be to provide several smaller neighborhood health centers or to provide services through other mechanisms (e.g., hospital outpatient clinics, group practices). Perhaps programs focusing on maternal and child health services should be considered, or attempts to screen apparently well people to discover heart defects. If, in these circumstances, the analyst sets out merely to find the optimum location for the health center, he is not using the full potential of analysis, and other studies, or advice from other sources, or even the mere idea that a better alternative may exist, are likely to push his work aside.

An illustration of an analysis that avoided the means/end trap is found in the New York City Rand Institute Fire Department Deployment Study. Here the analysts succeeded in switching their primary effort from the decision makers' desire for a computerized communications system designed to relieve the workload problem to a direct attack on the workload problem itself, finding in the process that computerizing the communications system would bring negligible improvement.[2]

It is a pitfall to believe that goals should, or can, be set independently of the means to obtain them. There is considerable evidence that, on the contrary, operationally significant objectives are often the result of opportunities offered by means that become available rather than the source of their discovery. The goal of landing a man on the moon was not set until it became feasible to do so.

The reason: it is usually impossible to select satisfactory objectives without some idea of the cost and difficulty of attaining them. Further, not all of the consequences of the various actions that may be necessary to attain an objective can be anticipated before analysis; there may be unpleasant side effects that render these actions, and thus their goals, completely unacceptable.

The constraints imposed by the client or the decision maker are in the same category as goals, for it is doubtful if any real distinction can be made between goals and constraints of this type (Simon, 1964), for goals are little more than constraints selected for special attention because they motivate the decision maker or because it is convenient for the analysis. (There are, of course, other constraints imposed by nature or sometimes by higher-level decisions that are unambiguous and irremovable.)

A client's constraints can sometimes be arbitrary and ill-founded. In a case cited by White (1975), "the senior executive, ... refused to consider certain alternatives, probably not because they were not worthwhile, but probably because someone else thought they were worthwhile, and he wished the study to emanate from his own ideas."

The translation of a decision maker's desires into something operational —that is, into measurable objectives or into a set of constraints—is a path full of pitfalls. For example, once a constraint, a limit value, is established by, or in concurrence with, the decision maker, it can distort the results significantly. Majone (1978, p. 20) points out:

> ... the opportunity cost of a proposed policy constraint must be carefully considered before the constraint becomes firmly embedded in the analytic structure. As Hitch and McKean (1960, p. 196) write, "casually selected or arbitrary constraints can easily increase system cost or degrade system performance manyfold, and lead to solutions that would be unacceptable to the person who set the constraints in the first place." They cite the example of a weapon-systems study, where a constraint on acceptable casualties led to solutions in which 100 million dollars was being spent, at the margin, to save a single life. Many more lives could have been saved with the same resources. Had the policymakers realized the opportunity cost of their safety requirements, they would probably have set them at a lower level. Or, like good utilitarians, they may have chosen to treat the risk factor as a subgoal, to be optimized compatibly with other systems' requirements and the available resources.

MEASURING ACHIEVEMENT BY PROXY

Worthy and apposite objectives are necessary, but for the purposes of analysis it is desirable that they also be clear and measurable, or that good surrogates or proxies exist. A good proxy is one that tells us with acceptable accuracy how well the real objective is being attained. To measure the health status of a population, mortality rates are often used (Grosse, 1971, p. 19). This, of course, is in the right direction; the pitfall is that a mortality rate ignores many factors that are important to good health.

The mark of a good measure of effectiveness is that it closely reflects the objectives. Unfortunately, there are a number of pitfalls that lie in the path of the analyst who tries to find such a measure.

One crude scheme is to use input to measure output, as, for example, in comparing the quality of primary-school education in various districts in terms of the expenditure per pupil. A second is to use workload or efficiency measures to compare quality of output. Consider a single unambiguous goal, say to improve garbage collection. To compare alternatives, we need a scale on which to measure their effectiveness. But, as there is no obvious scale to measure better garbage collection, we have to use a proxy—a measure such as the percentage of city blocks remaining without health hazards, or the reduction in the number of fires involving uncollected solid waste, or in rodent bites, or in valid citizen complaints. Unfortunately, all of these treat just one or a few aspects, not the full value of better garbage collection. In practice, people often use even less satisfactory scales, for instance, an efficiency measure—number of tons collected per man-hour—or a workload measure—tons of waste collected—that indicate nothing about the quality of the work.

A third approach, where there are a number of attributes that should be considered, is to devise a composite index of worth, say an ordinal or cardinal utility function. The possible danger here is that the function is to a large extent the product of the analyst's judgment of the decision makers' values and their weights and not that of the decision makers themselves. The decision makers, if they were willing to spend the time, could work out their own set of weights (with guidance from the analyst) but even here the analyst's influence is powerful. Hatry (1970) comments:

> There is no doubt that the job of decisionmaker would be easier if a single effectiveness measure could appropriately be used. However, I contend that such procedures place the analyst in the position of making a considerable number of value judgments that rightfully should be made in the political decisionmaking process, and not by the analyst. Such value judgments are buried in the procedures used by the analysts and are seldom revealed to, or understood by, the decisionmakers.
>
> Such hocus pocus in the long run tends to discredit analysis and distract significantly from what would be its principal role.

One would like measures of effectiveness to reflect in a measurable way the ultimate good effect intended. For example, in a school health program aimed at detecting and correcting medical conditions that handicap a child's social adjustment or academic achievement, the ideal measure is the increase in the number of such children who, as a result of the program, engage in less socially aberrant behavior and make better grades. Unfortunately, improvement in behavior or even in grades due to a health program may be so difficult or expensive to measure that measurement of some intermediate goal may be operationally more significant. Thus, the number of health examinations given, or the number of children with handicapping conditions who are detected, or

the number of those detected with handicapping conditions who receive appropriate medical treatment, or the number of those who receive appropriate treatment who have their handicap corrected, or even measurement of the number of handicapped children who are missed or not treated may have to serve as a proxy to indicate the success of the program. Actually, even our ideal measure is defective, for we should be talking about learning, not grades, if we remember what's important rather than what's measurable.

MISJUDGING THE DIFFICULTIES

Another pitfall lies in the misjudgment by the analyst of the difficulties of problem solving in the practical world of affairs as contrasted with the laboratory or classroom. In the latter the analyst or scientist can decide for himself where to bound his problem. (The formulation of scientific problems is no less important; as has often been said, scientific reputations are made in choosing the problem.) The systems analyst, on the other hand, must tackle what is there, and problems in the real world tend to proliferate. An analyst can set out to tackle far more than he can handle with the time and resources available. He can trap himself if he does not realize that he is not going to give a complete solution to the decision maker's problem. At best, he will produce evidence that other analysts will find hard to refute. As Holling (1978, p. 170) puts it:

> Any useful analysis is based on an abstraction of reality. Such analyses therefore will always be incomplete. Attempts to "include everything" result in ambiguity, confusion, and intractability. The irony is that the more rigorous and organized the attempt to abstract a useful portion of reality for analysis, the more tempting it is to presume that those features left out of the analysis are unimportant. The more effectively the known is analyzed, the more likely it is that decisions will be based upon the analysis. But the unknown cannot be ignored, and any attempt to do so is bound to end in unpleasant surprises and policy failures. . . .

An analyst, of course, must be just as careful to tell his client what he has not done and not included as he is to describe his results, for otherwise he may trap his client into the belief his results mean more than they do.

Analyses, except those directly requested by the decision maker or his immediate staff, require considerable time to penetrate to the decision maker; unless there is vigorous lobbying, it may take several years. If analysts want their work to affect policy, they must allow time for it to reach the policy maker before the moment for decision is past, as well as time to do the study.

BIAS

Finally, the analyst must guard against the biases introduced during formulation, a pitfall found elsewhere in the process, but to a lesser degree. Some biases

are well known, such as deliberately giving preference to aspects that are easy to quantify and neglecting the intangible, unquantifiable ones. Others are more subtle, introduced, say, because the analyst unconsciously acquires the goals and preferences of the organization for which he works. As a naval analyst, for example, he may consider only "scenarios" that are ideally suited for employment of the weapons he advocates. Wilson (1973) finds these latter biases to be almost universal, observing that "[a]ll policy interventions in social problems produce the intended effect—if the research is carried out by those implementing the policy or their friends." If, however, the research is carried out by those skeptical of the policy, the opposite holds.

Wilson goes on to say:

> Rarely does anyone deliberately fudge the results of a study to conform to pre-existing opinions. What is frequently done is to apply very different standards of evidence and method. [Agency studies] accept an agency's own data about what it is doing and with what effect, adopt a time frame (long or short) that maximizes the probability of observing the desired effect, and minimize the search for other variables that might account for the effect observed. [Independent studies] will gather data independently of the agency, adopt a short time frame that either minimizes the chance for the desired effect to appear or, if it does appear, permits one to argue that the results are "temporary" and probably due to the operation of the "Hawthorne Effect" (i.e., the reaction of the subjects to the fact that they are part of an experiment), and maximize the search for other variables that might explain the effects observed.

Parochialism, the "cherished belief," or adherence to a "party line" has been called by Herman Kahn and Irwin Mann (1957) "... the single, most important reason for the tremendous miscalculations made in foreseeing and preparing advances or changes in the strategic situation." Today, the introduction of bias by the analyst is well recognized, as Lesourne (1977, p. 75) notes:

> ... Ten years ago, there was a tendency to consider that a model produced by a team did depend on the decision studied and on the policymaker's objectives, but was not affected by the team itself, except by the team's intellectual abilities. Now we are beginning to realize, under the influence of new ideologies, that, as with the reconstruction of history, what OR professionals produce is partly conditioned by their goals, their values, the culture to which they belong.

Lesourne goes on to say that the presence of such bias does not mean systems analysis must be abandoned, "... but it introduces relativity into a field in which some of us have been trained in over-positivist attitudes." Even though most analysts strive to be objective, the knowledge that bias is often there underlines the need for competing studies of the same issue, with their different sets of biases.

The decision maker, as client or problem giver, can also trap the analyst by forcing assumptions on him or merely by expressing such strong prejudices against certain alternatives that the analyst feels their investigation is not worthwhile. The natural tendency of the decision maker is to keep the range of

alternatives narrow, but he may pay a high price, as some of the excluded alternatives may be far superior to those remaining.

PITFALLS IN MODELING

Let us turn now to models and modeling. Models, in my definition, include any procedure that is meant to help the policy or decision maker and his staff predict the consequences of proposed courses of action. Models may be implicit or conceptual, existing only within the mind of the decision maker (and possibly rife with pitfalls), but we will confine our attention to explicit models designed by others. These latter may take many forms—physical, natural language, mathematical equations, or a computer program, for instance—and the method of prediction may be judgment, physical manipulation, numerical approximation, or simulation, among others. The role of the policy model is to tell or predict what would happen were the decision maker to adopt a particular course of action, and in some cases it is also to indicate the one he should (on some explicit basis) prefer—processes obviously full of difficulties if not pitfalls.

EQUATING MODELING WITH ANALYSIS

Analysis without the use of some sort of model, explicit or otherwise, is impossible. Many analysts, unfortunately, tend to view modeling as identical with analysis and even with policy making itself. But models are only one ingredient and modeling only one step in policy analysis; searching out the right problem, designing better alternatives for consideration, and skillfully interpreting the computations from the model and relating them to the decision maker's problem are equally significant. In Bruce Goeller's (1977) analogy, models are major tools in the analysis of policy, just as brushes and palette are major tools in painting. A model is a means to an end, not an end in itself. Forgetting that precept can trap policy analysts into acting as if the model they set out to build were more important than the question they set out to help answer. That is, they begin to tailor their model to simulate, to greater and greater perfection, larger and larger segments of the situation of interest, knowing full well that much must be left out. Often this is done without the specific problem in mind, with no thought to the decision the model is to guide, or even to who the real decision maker, with his preferences and constraints, will be. For this last reason alone, it is no wonder that decision makers so frequently find model results irrelevant and unusable.

Analysts, in addition, become wedded to a particular modeling approach—say, optimization, systems dynamics, or input–output—and then tend to adapt the problem to fit the model type they know best rather than the

other way around. But it is the problem that is important; adapting it to the demands of a particular model form can change it. Moreover, even if that were not pitfall enough, each method has its characteristic traps that modelers must constantly guard against.[3]

For example, computer simulation may be selected as the modeling technique because it is well understood, very easy to set up, and seems to be the least costly. But it is only inexpensive in the model-building stage (Bowen, 1978). In fact, a simulation is likely to be the least desirable of models. It offers low insight, since it does not provide an explanation of the observed results, and it may involve a long and expensive analysis phase. Nevertheless, it may be the correct choice as the model, if only because no other choice is available.

For certain purposes, a model may indeed be an end in itself, and merely to demonstrate that some process can be modeled on a computer may be a worthy accomplishment. It is a pitfall for policy analysis, however, for the analyst to become more interested in the model than in the problem he set out to help solve. This trap is easy to fall into, for modeling can be a fascinating business. Technical people with training, knowledge, and capabilities like to use their talents to the utmost. It is not hard for them to focus their attention on data analysis, or the relationships between various factors, or how to approximate a difficult relationship, attempting to find a more representative fit to the situation. In doing so, they may neglect facets of the study that cannot be easily incorporated into the model, a frequent accusation of critics of policy analysis. They also like to operate with the model and to explore its capabilities. Thus they may find out a great deal about the inferences that can be drawn from the model, but not necessarily very much about the question they should be trying to answer.

IMPROPER TREATMENT OF UNCERTAINTIES

No aspect of systems analysis is more difficult to handle properly than uncertainty. Uncertainties (the unpredictabilities in factors that affect the outcome of a course of action) are the *sine qua non* of the analytic process; actions can be taken to counter their presence, but they cannot be eliminated. Uncertainties about parameters whose distributions are known or can be inferred with some confidence and with which calculations can be performed tend to absorb the attention of analysts out of all proportion to their importance as compared with the more serious uncertainties about which little is known. To take the calculable uncertainties into account properly can be an attractive challenge to the analyst's ingenuity. The danger lies in a corresponding neglect of the uncertainties that are not stochastic in nature. There is a huge critical literature on the treatment of uncertainty in decision making; for example, see Raiffa (1968), Schlaifer (1969), or Madansky (1968).

Sensitivity analysis is a major technique for compensating for uncertainty. There are, however, several misconceptions about its use. Holling (1978, p. 103), for example, points out that sensitivity testing in which sensitivity to a series of parameters is tested by changing them one by one and determining the result will not produce correct results. Simultaneous variation of the parameters is necessary to give reliable sensitivity testing. He cites a study of the Meadows world model by Scolnik (1973). One-by-one variation had shown the model's prediction of boom and collapse to be stable to small variations in many parameters. But when several parameters were varied simultaneously over ranges of less than 10 percent, the results changed dramatically. Bowen (1978) cites a second misconception, the belief that sensitivity analysis can overcome the omission of the stochastic nature of the real world system when "expected value" models are used.

Along these same lines, there is a precept, also pointed out by Bowen (1978), often followed that, in a situation in which two alternatives are being compared, assumptions, aggregation, and simplifications in modeling that apply "equally" to both sides produce usable relative solutions. This is far from true even if the situation is fairly symmetrical.

ATTEMPTING TO REALLY SIMULATE REALITY

There are two common ways to go about constructing a model. One is to design the model with relevance uppermost, i.e., with the issue or problem as a guide. The other is to keep realism uppermost, attempting to capture the real world by simulating the situation to perfection and thus, if all goes well, to be in a position to answer not just a single question, but other questions about the situation or subject as well. Using this latter approach, there have been, for example, attempts to model the behavior of a city, its population, and its government so realistically that the model would predict faithfully what will happen to growth, population movement, industrial development, and so on.

What can happen in the attempt to simulate reality is that the analyst finds himself criticized because the model he is building has left out various aspects of the situation or issue under investigation that the critic feels are important. The analyst is vulnerable to these criticisms unless he heeds the principle that the problem—the question being asked—as well as the process being modeled, determines what should be in the model. Without reference to the question, he has no real guide to what to accept or reject, because he has no way to decide what is important or relevant. He can answer such criticism only by making the model more inclusive and more detailed. Doing so may not stop the criticism, for a great many factors must still be left out, but the result may be to make the model complex and difficult to handle, increasing the number of operations to a point where the total number is determined not by what is relevant but by the capacity of the available computer. Indeed, as Kenneth Bowen

(1978) remarks, "some simulations are not appreciably less complicated than the real world. We need models to understand such models."

Analysts will deny that they believe that their model is equivalent to reality, yet they may still convey this impression to the unwitting decision makers. For instance, Kenneth Kraemer (1973, p. 74) in his description of the San Francisco Housing Market Simulation Model remarks: "For example, Wolfe and Ernst speak of the San Francisco model as reproducing in an abstracted form the *actual* [italics Kraemer's] interrelationships that occur, or might occur, in the market for land and building space in the city!"

BELIEF THAT A MODEL CAN BE PROVED CORRECT

A particularly dangerous myth is the belief that a policy model can be fully validated—that is, proved correct. Such models can, at best, be invalidated. Holling and his colleagues in their perceptive work on adaptive environmental assessment (1978, p. 95) say

> In fact, it is the central tenet of modern scientific method that hypotheses, including models, can never be proved right; they can only be proved wrong (Popper, 1959). This is why the frequent claim of—and demands for—"valid" models . . . are so unsound. Provisional acceptance of any model implies not certainty, but rather a sufficient degree of belief to justify further action. In practice, the problem is one of model invalidation—of setting the model at risk so as to suggest the limits of its credibility. . . .

Thus the aim of the validation (or rather, as Holling suggests, invalidation) attempts is to increase the degree of confidence that the events inferred from the model will, in fact, occur under the conditions assumed. When you have tried all the reasonable invalidation procedures you can think of, you will not, of course, have a valid model (and you may not have a model at all). You will, however, have a good understanding of the strength and weaknesses of the model, and you will be able to meet criticisms of omissions by being able to say why something was left out and what difference including it would have made. Knowing the limits of the model's predictive capabilities will enable you to express proper confidence in the results obtained from it.

Models of the same situation built by different analysts will not be the same and will usually produce different results. If they are not the same, then which is the right one? "Our answer . . . is that there is no right one. A model is only one piece of evidence . . ." (Holling, 1978, p. 79).

One must also check the data. Holling (1978, p. 97) cites an example from a study to find ways to lessen damage to North American forests from the spruce budworm, a defoliating insect, where the model predicted that forest volume would decline independently of insect damage, while it was "common knowledge" that volume would remain high if insects were controlled.

We spent 2 months checking the model for errors when we should have been spending 2 days looking at available raw data on forest volume. When we belatedly took this obvious step, the model was vindicated and common knowledge was shown to be at variance with the data on which it should have been based. We suspect that this is not a rare occurrence.

NEGLECTING THE BY-PRODUCTS OF MODELING

It is a pitfall to believe that the only valuable output of the modeling process is the results of the computation and that, as a consequence, quantitative or mathematical models are really the only useful kind. In fact, for problems where behavioral, political, and social factors dominate, although we may be unable to construct any model that can produce useful computations, building a model may still be valuable because of what one learns about the problem during the process and because of the guidance even a crude model can provide to the judgment and intuition of analyst and policy maker.

Judgment and intuition permeate every aspect of policy making and of analysis to support it. Policy analysis, like science, is a craft activity, as Majone argues in Chapter 1. Judgment must be used by the analyst to limit the extent of his inquiry, to decide what approach to take, to select what factors to include, and at a host of other stages before he applies it to interpreting his results. A great virtue of the modeling process is that it provides a systematic way to focus the required judgment and intuition.

An explicit model, quantitative or not, introduces structure into a problem, providing a means of breaking a complex multidisciplinary question into constituent parts that can be handled one at a time, often by experts within a single discipline. In using and building models, analysts and the experts on whom they must call for practical knowledge and experience are compelled to use a precise terminology and to develop their ideas and to exercise their judgment and intuition in a well-defined context, one that puts their opinions in proper relation to those of the other participants. Moreover, if they disagree initially, they must reach an acceptable compromise. The model thus provides an effective means of communication. It is a pitfall, however, to accept uncritically the contributions of the experts (Ackerman *et al.*, 1974).

Because features other than the calculations can be so important, it is a pitfall to make such an effort to keep a model quantitative that factors important to the question have to be left out. One cannot, as Helmer (1978) shows, advance in an uncertain context of any complexity solely by formal theoretical or mathematical means. Overemphasis on quantification can result in a dangerous underemphasis on or even total neglect of considerations that resist quantification; this is a source of much criticism of systems analysis for neglecting ecological and other "fragile" elements (Tribe, 1972).

OVERAMBITION

While the usefulness of simple policy-relevant models is unquestioned, there has been, almost since large-scale multipurpose computer modeling began, sometime in the mid-fifties, controversy about the value of such modeling —particularly for use in public, or societal, decision making—to give advice about what to do about such issues as unemployment, the environment, or law enforcement. Greenberger and his colleagues (1976) and Brewer (1973) offer considerable evidence that analysis dependent on computer-based policy models has been a disappointment. In fact, with respect to urban problems, Brewer (p. 233) is rather pessimistic: "... results from real contexts to date have been so checkered that even the eventual realization of the mythical policy-assisting model is in doubt."

Large-scale models designed for many purposes are likely to serve none in the end. Lee (1973), in cataloging the sins of their designers, found that the limitations or unintended constraints resulting from the structure of most such models are almost impossible to perceive. He gives an example (p. 165):

> Trip distribution functions (the heart of any gravity model, which includes most of the land-use models) are fitted to observed trip frequencies for different classes of workers and (sometimes) different classes of households. While valid at the scale of a metropolitan area, the gravity model has no statistical explanatory power at the neighborhood level. This is a near-classic case of imputing individual (or census tract or neighborhood) behavior from aggregate relationships—the ecological fallacy. In addition, the purely descriptive nature of the gravity model implicitly assumes that the existing land-use patterns served by existing transportation, with existing prices, qualities, and service levels, will be maintained in the future. To then ask the model what will be the impact of a change in the type of transportation service provided is pointless; the model lacks the structural information to be able to trace out the consequences.

For economic–demographic models, Arthur and McNicoll (1975, p. 262) find that the aspiration to be multipurpose "... makes them large, structurally inflexible, overdressed in side issues, and beyond proper validation...." They find (p. 257) that this can introduce false complexity:

> Take, for example, the question of fertility reduction. A recent TEMPO study of 19 different countries finds in every case that "slower population growth, produced by declining fertility, translates directly into a more rapid growth of GNP per capita. This conclusion is extremely robust in the sense that it is relatively invariant under the diverse socioeconomic conditions encountered in the different countries studied and under widely differing parameter values for the key equations in the models." (Brown, 1974). This appears to be astonishing testimony to the universal need for swift fertility reduction. The TEMPO models, however, make one key, central assumption. Economic output takes place according to a Cobb–Douglas production function, a function that enjoys the property that, no matter what the parameters, the derivative of per capita output with respect to labor is always negative. An increase in population can therefore never pay its way. One assumption, questionable in theory, therefore governs the outcome; the result has been settled in advance and built in....

A pitfall for those making use of the results from large computer models, noted by Majone (1977), is that modelers, to make their models produce answers that agree with what they or their clients think those answers should be, sometimes arbitrarily "fudge" the parameters. Holling (1978, p. 78) makes a similar point:

> Users often adjust values in the input interaction matrix in order to give "reasonable" output: i.e., data are adjusted to fit preconceived notions of what should happen—obviously not useful in the context of environmental impact assessment.

Many professional modelers argue that if large-scale multipurpose modeling has failed, it is because policy makers have misunderstood or misused the results of their models. While this may be a contributing factor, much of the negative reaction is due to the modelers themselves and the way they practice their craft, particularly in allowing themselves to fall into the next two pitfalls discussed.

SEEKING ACADEMIC RATHER THAN POLICY GOALS

Modelers ordinarily come from the academic world, and so do their rewards. This leads many analysts, as pointed out by Ansoff and Hayes (1973), particularly those still holding academic positions, into the pitfall of building their models to satisfy academic criteria. Professors of operations research are likely to evaluate models devised by their students to solve a textbook problem by a set of criteria based on quality. That is, they would like the models or methodologies to be nontrivial (permitting inferences not perceivable by direct observation), powerful (offering a large number of nontrivial inferences), elegant (using a minimum of carefully selected analytic tools to produce a model of great power), cost-effective (producing the answer with a minimal number of machine operations), valid (capable of reproducing results based on historical data), and so on. Unfortunately, their students, and often they themselves, on becoming practicing analysts, may unconsciously overemphasize the same "quality" in the policy models they build.

Decision makers, on the other hand, are interested in a largely different set of criteria based on the applicability of the model to the segment of the world for which they have responsibility. They want a model to be relevant (to treat a problem of importance to their domain), valid (to give a high degree of confidence that the inferences obtained from it will, in fact, occur in the real world), usable (for there to be a good chance that the model can be accepted as a basis for action), cost-effective (for the benefits to them and to their organization that result from its use to exceed the expense of developing and applying the model), and so on. They are interested in the costs of modeling—not only the cost of the resources required but other costs as well, particularly the cost of delaying

action because of the time needed for the modeling processes. They have little direct interest in power and elegance. They would, of course, like the model to be nontrivial—to give them or others new insight into, or ideas about, issues of interest. It is unreasonable, however, for them to expect the model and the accompanying analyses to determine a completely acceptable and unquestionable solution to the problem; a model produces evidence at most, not proof. The most one can ask is that the analysis be relevant and have some validity, for then it may provide the perspective that may lead, through judgment, to a satisfactory resolution of the problem.

Unlike the policy maker, the analyst is not ordinarily held responsible if a particularly pressing public problem does not receive the attention it deserves. The policy maker, on the other hand, may be penalized for failing to focus attention on a critical problem or for failing to take action that improves the situation.

This difference in responsibility between analyst and decision maker can lead the analyst, in choosing what sort of model to construct to help with the decision maker's problem, to select a type that will offer the greatest latitude for the application of modeling skills and knowledge and that will produce, as a consequence, the greatest recognition and reward. This appears to be found in highly nontrivial, powerful, and elegant models—mostly large computer simulations. The analyst may, in fact, seek to work only on problems that offer opportunities for models of high quality. In any given problem, these considerations can become the criteria for choosing how to tackle the problem and what kind of model to build.

The hold of the professional community on the modeler, the source of the pitfalls in this section, is strong. As House and McLeod (1977, p. 4) explain it, because "... one's professional career is predicated on peer group status, the modeling community usually tends to be more loyal to the demands of the professional community than to the needs of the policymaker."

INTERNALIZING THE POLICY MAKER

Many modelers subscribe to what, in my view, is almost a complete myth, namely, that a policy model can be made so comprehensive that it satisfactorily internalizes the policy-making process itself. They believe that the model can capture the preferences of and constraints, both real and imagined, on the policy makers, so that it can weigh and trade off the numerous factors involved in making a policy decision and designate the best alternative in a way that will be both credible and acceptable to the policy makers. In brief, they fall into the pitfall of thinking that one can substitute a model for the decision maker, something possible in only a very narrowly prescribed situation. We do not have, and do not expect to have, models that will simply allow the public

official to "plug in the numbers" and get back the correct decision with regard to housing, transportation, public welfare, or education.

In most policy studies—for instance, a study to assist in a choice of transportation alternatives—the many impacts or consequences predicted by the model or models—the environmental, economic, financial, service, social (i.e., distributional) and other outcomes—must in some way be presented for consideration to the policy makers. The usual presentation method (e.g., using multiattribute utility theory or cost–benefit analysis) combines the many different impacts into a single measure.

Any aggregated approach to evaluation has two serious defects. First, a great deal of information is suppressed by the aggregation; the fact that alternative *A* has environmental problems that can cause serious political repercussions whereas alternative *B* requires developmental work that may not pan out may escape attention. Second, as pointed out earlier in the discussion of proxies, any single measure of value depends on the relative weights assigned by the analyst to the various factors that enter and on the assumptions he used to get them into commensurate units. To assign these weights requires a great deal of arbitrary quantification, often using means and arguments that are, to say the least, questionable. To do so requires judgments by the analyst, or by experts on whom he depends, about probabilities and values that may have no solid basis in fact or theory. More important, where value estimates are required (which is almost everywhere), the results are often based on judgments by the wrong people—by the analysts, not by the responsible decision makers.

The most promising approach of the aggregative type, that of decision theory, attempts to model the value system or preference structure of the policy maker (or policy makers) so as to be able to predict what the decision would be were he (or they) to be presented with the alternatives and information about their consequences under various assumptions about the future. Somewhat similar schemes have sometimes worked satisfactorily when the decisions were of a repetitive type and for preliminary design and screening of alternatives (where the value judgments need not be as accurate).

In contrast, the natural disaggregative approach in which the alternatives and their full spectrum of consequences are presented in a tabular array to allow the decision makers to impose their own value judgments retains all the information and avoids interposing the analyst's judgment. There are, however, problems with the proliferation of information. The refinement of this approach by Goeller (1977, p. 217), the "scorecard" with shading or color coding to provide an effective "gestalt," manages to overcome many of them.

NOT KEEPING THE MODEL RELEVANT

An important factor in the success of analysis and modeling is the interest and attention of the policy maker in the work being done (Walker, 1978;

Greenberger *et al.*, 1976, p. 326; Holling, 1978, p. 51). A sure way to lose his interest is to allow the model to become irrelevant to his problem.

To keep the model relevant, Goeller (1976) suggests that it be considered initially as a "black box," with first attention being given to what knobs, representing policy variables under the control of the policy makers, and what dials, representing consequences of interest to them, should be put on the front of the box. Only then should attention be given to the design of the contents of the box.

It is important, however that the model not remain a 'black box.' To quote Lee (1973, p. 175) once more:

> Probably the most important attribute any model should have is transparency. It should be readily understandable to any potential user with a reasonable investment of effort. "Black box" models will never have an impact on policy other than possibly through mystique, and this will be short lived and self-defeating. A transparent model is still about as likely to be wrong, but at least concerned persons can investigate the points at which they disagree. By achieving a consensus on assumptions, opposing parties may find they actually agree on conclusions. A good deal of effort will be required to get people to think this way, because it is a trait of modelers and nonmodelers alike to throw in everything at once when they do not understand a problem; odds are that somewhere in the jumble are the parts to the explanation. The trouble is, nothing is learned from such an exercise, and future problems receive the same treatment. Patience and good will are necessary in this communication and education process; big "black box" models are not.
>
> The practice common to many policymakers of assigning one of his staff to monitor the study once a contract is approved can also lead to irrelevance and thus present a pitfall to the analyst. The staff monitor may be competent technically but he cannot pass on what the policymaker really has to have to satisfy the needs, interests, and goals of his constituency and what motivates his personal and political life. The policymaker loses interest if his needs are slighted, while the analyst, without the motivation the policymaker would provide, may turn to modeling for modeling's sake.

NOT KEEPING THE MODEL SIMPLE

A pitfall in allowing the model to become more complicated than it need be is that eventually it has to be explained to and understood by somebody else. For Greenberger *et al.*, (1976), the most important reason why so many policy analyses fail as "instruments for the clarification of policy issues and the enlightenment of policymakers" is that "the inner workings of a policy model are seldom understood by anyone but the builders of the model (and not always by them). This is a weak foundation for gaining the reliance and trust of policymakers."

One should not expect the model to be so simple and transparent that the client himself can understand its inner workings, but his analyst should be able to do so. It would be a rare client who would want to or could use this kind of detailed analytic knowledge. The analyst, however, should be able to exhibit in somewhat simplified form the key structural elements of the model and their

most important relationships so that the client can not only understand how the results came about, but also incorporate the logic of the analysis into his continuing thinking about the problem.

For a policy maker to convince himself that a model is conceptually and theoretically sound is not easy. He can ask that mathematical statements (a few) be translated into verbal statements to see if they make sense and seem plausible. He can inspect the list of variables that are selected as representative of reality. He can ask if various symbols have empirical referents and if degradation or fudge factors are present in the model. He may be able to check in simple cases whether the model output agrees with his intuition. The mere fact that a model can produce a close approximation to a historical time series is no guarantee that it has much value as a predictor. As is well known, almost any model that reasonably approximates the situation with a few free parameters can be made to produce results that closely fit any given curve (Bush and Mosteller, 1966). But the simpler the model, the easier it is to understand.

CAPTURE OF THE USER BY THE MODELER

A pitfall for the user (Brewer, 1975, p. 6) is to allow himself to be captured by the modeler. This can sometimes happen with models that are built for repeated use (e.g., a model to assign police patrol cars to beats). It is in the interests of the model builder to keep its details from competitors and to see that the ideas embodied in it enhance his own academic credit. For this and other reasons, documentation may not be done well. Thus, when the user, because of changed circumstances, wants to ask a slightly different question, he has to go back to the original builder—who, for a follow-on contract, will undertake to make the change. This may present difficulties, for often the computer specialist who actually put the model on the computer and made it work has moved elsewhere. To protect himself, the user should insist on full enough documentation to allow a second modeler to reproduce the model.

CONCLUSION

The pitfalls discussed in this chapter are only a small sample of those that await analysts and decision makers. Many other pitfalls associated with formulation and modeling have been pointed out in the literature, although not always under the heading of pitfalls. Collections are found in papers by Kahn and Mann (1957), Lee (1973), Koopman (1956), Quade and Boucher (1968), Quade (1975), and Brewer (1975), with Strauch (1974) and Webber and Rittel (1973) giving particular attention to formulation.

The primary purpose of systems analysis is to improve decision making

rather than to advance knowledge or to improve forecasting *per se*. Therefore, one way to avoid pitfalls is to remember that, from start to finish, problem formulation and modeling must be oriented toward the decision-making problem.

NOTES

1. To Webber and Rittell (1973) the situation may also be "wicked"; to Strauch (1974), "squishy." "Mess" is used here in the sense of Ackoff (1974).
2. For a discussion, see "Putting Out Fires in the Fire Department," Chapter 8 in Greenberger *et al.*, 1976.
3. For a discussion of problems associated with various model types, see *The Electronic Oracle: Computer Models and Social Decisions*, by D. H. Meadows and J. M. Robinson, John Wiley & Sons, Ltd. Chichester, England. Forthcoming.

REFERENCES

Ackerman, B. A., S. Rose-Ackerman, J. W. Sawyer, Jr., and D. W. Henderson. 1974. The Uncertain Search for Environmental Quality. New York: The Free Press.

Ackoff, R. L. 1974. Redesigning the Future. New York: Wiley.

Ansoff, H. I., and R. L. Hayes. 1973. Role of models in corporation decision making. *In* M. Ross, (ed.). Operational Research '72. New York: American Elsevier, pp. 131–162.

Arthur, W. B., and G. McNicoll. 1975. Population and development models. Population and Development Review 1(2): 251–265.

Bowen, K. 1978. Analysis of models. Paper submitted for the *Handbook of Applied Systems Analysis*. Laxenburg, Austria: Survey Project, Institute for Applied Systems Analysis.

Brewer, G. D. 1973. Politicians, Bureaucrats, and the Consultant: A Critique of Urban Problem Solving. New York: Basic Books.

Brewer, G. D. 1975. An Analyst's View of the Uses and Abuses of Modeling for Decision Making. P-5395. Santa Monica, California: The Rand Corporation.

Brown, R. A. 1974. Survey of TEMPO Economic–Demographic Studies. Washington, D.C.: General Electric Company—TEMPO.

Bush, R. P., and F. Mosteller. 1966. A comparison of eight models, p. 335. *In* Readings in Mathematical Social Science (P. Lazarfeld and N. W. Henry, ed.). Chicago: Science Research Associates.

Checkland, P. B. 1978. Formulation of problems for systems analysis. Paper submitted for the *Handbook of Applied Systems Analysis*. Laxenburg, Austria: Survey Project, Institute for Applied Systems Analysis.

Goeller, B. F. 1976. System Impact Assessment: A Comprehensive Approach to Public Policy Decisions. R-1446-RC. Santa Monica, California: The Rand Corporation.

Goeller, B. F. 1977. The San Diego–Rand methodology package, p. 217. *In* Large-Scale Policy Modeling (P. House and J. McLeod, ed.). New York: Wiley.

Greenberger, M., M. A. Crenson, and B. L. Crissey. 1976. Models in the Policy Process: Public Decision Making in the Computer Era. New York: Russell Sage Foundation.

Grosse, R. N. 1971. Cost–benefit analysis in disease control programs. *In* Cost–Benefit Analysis (M. G. Kendall, ed.). London: The English Universities Press.

Hatry, H. P. 1970. Measuring the effectiveness of non-defense public programs. Operations Research 18(5): 774.

Helmer, O. 1978. Role-playing and judgmental techniques. Paper submitted for the *Handbook of Applied Systems Analysis*. Laxenburg, Austria: Survey Project, Institute for Applied Systems Analysis.

Hitch, C. J., and R. McKean. 1960. The Economics of Defense in the Nuclear Age. Cambridge, Massachusetts: Harvard University Press.

Holling, C. S. (ed.). 1978. Adaptive Environmental Assessment and Management. Chichester, England: Wiley.

House, P., and J. McLeod. 1977. Large-Scale Policy Modeling. New York: Wiley.

Kahn, H., and I. Mann. 1957. Ten Common Pitfalls. RM-1937. Santa Monica, California: The Rand Corporation.

Koopman, B. O. 1956. Fallacies in operations research. Operations Research 4(4): 422–426.

Kraemer, K. L. 1973. Policy Analysis in Local Government. Washington, D.C.: International City Management Association.

Lee, D. B., Jr. 1973. Requiem for large-scale models. Journal of the American Institute of Planners, May 1973, p. 163–178.

Lesourne, J. 1977. A prospective look at the OR system. European Journal of Operational Research 1: 73–78.

Madansky, A. 1968. Uncertainty. *In* Systems Analysis and Policy Planning (E. S. Quade and W. I. Boucher, ed.). New York: American Elsevier.

Majone, G. 1977. Pitfalls of Analysis and the Analysis of Pitfalls. RM-77-1. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Majone, G. 1978. The ABC's of Constraint Analysis. Working Paper No. 2. New York: Russell Sage Foundation.

Popper, K. R. 1959. The Logic of Scientific Discovery. New York: Basic Books.

Quade, E. S. 1968. Pitfalls and limitations, pp. 345–363. *In* Systems Analysis and Policy Planning (E. S. Quade and W. I. Boucher, ed.). New York: American Elsevier.

Quade, E. S. 1975. Analysis for Public Decisions. New York: American Elsevier.

Quade, E. S., and W. I. Boucher (ed.). 1968. Systems Analysis and Policy Planning. New York: American Elsevier.

Raiffa, H. 1968. Decision Analysis. Reading, Massachusetts: Addison–Wesley.

Schlaifer, R. 1968. Analysis of Decisions under Uncertainty. New York: McGraw-Hill.

Scolnik, H. D. 1973. On a Methodological Criticism of the Meadows World 3 Model. Tech. Rep. Dept. Math. Argentina: Fundacion Bariloche.

Simon, H. 1964. On the concept of organizational goal. Administrative Science Quarterly 9(1): 1–22.

Strauch, R. E. 1974. A Critical Assessment of Quantitative Methodology as a Policy Analysis Tool. P-5282. Santa Monica, California: The Rand Corporation.

Tribe, L. 1972. Policy science: analysis or ideology? Philosophy and Public Affairs. 12: 66–110.

Walker, W. E. 1978. Public Policy Analysis: A Partnership between Analysts and Policymakers. P-6-74. Santa Monica, California: The Rand Corporation.

Webber, M. M., and H. W. J. Rittel. 1973. Dilemmas in the general theory of planning. Policy Sciences 4(2): 155–169.

Wellington, A. M. 1887. Economic Theory of the Location of Railroads. New York: Wiley.

White, D. J. 1975. Decision Methodology. London: Wiley.

Wilson, J. Q. 1973. On Pettigrew and Armor: an afterword. The Public Interest 30: 132–135.

4 Pitfalls of Data Analysis

Stanley M. Altman

We live in an age of continuous confrontation with decision-making situations. Rarely, if ever, are decisions made where all possible choices and their resultant consequences are known. Decision making takes place under conditions of uncertainty—uncertainty about options available, sometimes uncertainty about the goals used to judge alternative options, and uncertainty about the eventual consequence of each option. Uncertainty can be reduced only by obtaining information, although it has been shown that there are limits to the degree to which uncertainty can be eliminated. Generally, more information can be acquired only by incurring increased costs. Therefore, the utility of acquiring additional information must be measured against the costs so incurred.

Within existing boundaries for reducing uncertainty, there is generally room for improving the information available to a decision maker. When a decision-forcing situation arises, an analysis of the phenomena involved is performed to obtain information. The phenomena may involve the operation of a government agency's service delivery system; they may involve the operation of a national government's social service program; or they may involve the design and evaluation of a public policy. Whatever the phenomena are, the process of information gathering involves collecting and analyzing data. The pitfalls that confront both those who analyze data and the decision makers who are consumers of the resultant information are the subject of this chapter.

AN OVERVIEW OF DATA ANALYSIS

A working definition of analysis is

> . . . the resolution or separation of the reference system (object, system, process being described) into parts to illuminate their nature and the interrelations and to determine general principles of behavior.[1]

In data analysis, the reference system (object, system, process) is described by data. Underlying the description is the data analyst's hypothesis regarding which particular characteristics of the reference system are important and must be accounted for, and which characteristics can be ignored. Data analysts use methods of analysis on data to discover relationships between the different variables for which data have been collected. These relationships are tentatively established by descriptive models that relate output (or dependent) variables to input (or independent) variables. The model developed represents the data analyst's hypothesis about the phenomena. Inferential statistical tests are used to obtain evidence to support the hypothesis. Statistical testing is one of the many points at which inferential statistics and data analysis overlap.[2]

There are numerous techniques used by data analysts to build models from data. Data analysts use techniques from descriptive statistics for reducing data into summary forms that are more illuminating and easier to work with than the original data.[3] Modeling techniques based on probabilistic and optimization arguments, such as regression,[4] are used, as are other modeling techniques that do not involve probabilistic arguments, such as two-way tables. The development and codification of this latter group of techniques has been pioneered by Tukey under the name "Exploratory Data Analysis."[5]

The process of fitting a model to data is the first step in a larger process of model building and model verification. Model verification involves using data that are independent from the data used to construct the model in order to test the model's predictive powers. The process of verification helps to determine whether the model is only a good fit to the original data or if it is more generally applicable as a predictive tool for dealing with the phenomena being studied.

The use of models in the larger quest to reduce the uncertainty that surrounds decision making and decision-forcing situations is a subject broader than the defined objective of this chapter. This chapter is concerned with the pitfalls that confront analysts working with data as part of the larger process of using models to reduce uncertainty. In particular, the chapter is divided into five sections, each presenting pitfalls common to the practice of data analysis. Each of the pitfalls discussed represents a potential danger to the effective implementation of any data analysis effort. Therefore, rather than arrange this discussion by assigning different degrees of importance to different pitfalls, the discussion is organized around the data analysis process. This order of presentation emphasizes the major steps in problem solving, generally described by the scientific method.

TRANSLATING A PROBLEM STATEMENT INTO QUANTITATIVE TERMS

The starting point for any analysis is a definition of the problem to be explored in terms of the list of variables (both dependent and independent) to be included in the analysis. One should not expect this list to be complete, but developing a list at the beginning of the analysis serves two purposes. First, it forms the framework for deciding what data are required. Second, it is an implicit statement of the analyst's belief in the applicability of underlying theory for describing the phenomena in a given problem.

An incomplete specification of the list of variables to be included in the analysis can lead to an incomplete analysis, or to incorrect conclusions being drawn. Allocating too little time to problem definition, focusing on the symptoms rather than the causes of the problem, being unwilling to exert the effort necessary to understand the underlying theory applicable to the problem at hand are all *pitfalls of incomplete specification*. The very nature of problem definition almost guarantees that the problem specification will be incomplete. But the pitfall consists of not knowing what was left out and its possible significance for the final results of the analysis. The pitfall of incomplete specification starts the entire analysis off on the wrong foot. Relationships that may exist between the specified list of variables and those excluded will not be accounted for in the analysis.

COLLECTING DATA AND USING COLLECTED DATA

Even if the list of variables pertinent to the problem being studied is completely defined, there are additional pitfalls that the data analyst must watch out for during the process of acquiring data. For a variety of reasons, these pitfalls can occur throughout the process of acquiring data.

One pitfall that can appear during the data acquisition process is the *pitfall of inappropriate proxy variables*. This pitfall can occur when a variable cannot be measured directly. It is common practice in such situations to measure a different variable that is assumed to be related to the original variable. This may be the only thing a data analyst can do when information on the desired variable cannot be measured directly. In such situations, the data analyst must recognize the potential limitation this places on the entire analysis and attempt, when possible, to determine the magnitude of the errors that can occur. The data analyst must be cautious in taking this step, because doing so implies an underlying belief about the relationship between these variables that may not hold. In addition, it is difficult to prove that this relationship does hold because the original variable cannot be measured. This pitfall can result from an incorrect conclusion drawn from commonsense arguments or a misapplication of experience.

G. D. N. Worswick cites the following example of this type of pitfall:

> Many quantities which appear in economic literature, such as numbers employed, or tons of steel, look very like familiar physical facts, but there is a subtle and all-important difference. The physical quantities of steel, or labour, are only proxies for economic "values" or "utilities." This distinction is vital: failure to make it can have detrimental consequences, both for policy, as for example in the discussion of the relation between wages and productivity, and for theory, as in the measurement of capital.[6]

Hatry further illustrates this pitfall:

> Appropriate criteria for evaluation of manpower training programs would not be the number of persons entering training—analysts do want this information, but not for evaluation of overall effectiveness—but rather the additional number of persons who obtain employment and the number of families who thereby get their standard of living raised above, say, poverty levels.[7]

Another pitfall of data acquisition is the *pitfall of biased measurement errors*. This pitfall can occur when the assumption is made that errors of measurement are randomly distributed about the variable's true value, when, in fact, this is not true. As a result, the data analyst has fallen into the pitfall of biased measurement error and has also set the stage for the systems analyst, who makes use of the data, to fall into the same pitfall.

The various causes of bias in measurement error can be summarized as follows: lack of designed experiments; hiding or lying about information; use of untrained observers; reliance on questionnaires; lack of universally agreed-upon definitions and classifications; errors of instruments; and the fact that observations may have to span a long period of time. This means that data analysts often have to deal with data gathered by a widely separated group of untrained "observers" who have vague if not conflicting notions of basic policy classifications and who have to elicit information from people who may be less than candid. Hanson points out:

> Errors are thus inevitable; all the available evidence indicates that there is no benevolent god who makes these errors cancel out. Given this situation, the best we can do is to be sensitive to the many sources of error, and, whenever possible express numerically the magnitude of probable error.[8]

Data acquisition in public policy fields has been modeled on the approaches used in the natural sciences. But the nature of the data is much different. Nowhere is this difference more obvious than in the problem of hiding information or just plain lying. The natural sciences may have a problem in getting good answers to questions, but at least it can be safely assumed that nature will not deliberately lie. This is not the case with the subjects of policy research. For example, in the United States, much of the CIA's budget is hidden in other agencies' budgets; this falsifies not only information on the CIA, but also on the other agencies. However, it is not difficult to understand the origins of such a situation. Game theory suggests that bluffing is a sound move in strategic

behavior. In other words, giving the wrong impression to an opponent, real or potential, is an essential feature of rational strategies.

Data are also suppressed, or statements of needs, resources, or impacts are exaggerated or understated. In the early 1970s, it came to light that the New York City Transit Police was falsifying crime statistics to make the deployment of transit police on subway trains appear more effective than it was. Crimes that occurred shortly after the shift of increased transit police began were being reported as though they occurred before the shift began. Similarly, crimes committed just before the shift ended were reported as though they occurred just after the shift ended. These statistics were being used to argue for more resources to increase the number of transit police deployed on subway trains.[9]

Other data may be unreliable, even though there was no intention to lie. Crime statistics provide such an example. It is difficult to even know if the crime statistics reported are reliable because they only refer to reported crimes and not to total crimes committed.

> As an example, if more policemen are assigned to patrol duty, their augmented force may cause them to observe a larger fraction of all crimes than they did before the increased manning, thus showing an apparent increase in crime.[10]

Morgenstern[11] discusses other types of data acquisition errors that the data analyst should be aware of. His discussion of errors associated with questionnaires, untrained observers, and misclassification of data are paraphrased in the following discussion.

Questionnaires are widely used in policy research and, of course, they may introduce certain kinds of errors. It is difficult to ask good questions. One must be careful because the answers may vary according to the type of approach: oral, written, or mail. Care also has to be taken because certain words have political connotations that could interfere with objectivity. Other threats to objectivity are evasions, lies, forgetfulness, and the inability to interpret one's own actions. The only way to judge the merits of data gathered through questionnaires is to read the questionnaire. If the questionnaire is not published with the results, the results should be rejected. One should also beware of surveys that break their results into many small groups. The number of subjects falling into some of these categories will be small and results will be inaccurate. "These 'outliers' are rare events (in terms of the sample) and probably belong to a different statistical distribution than the one encountered."[12] (See the discussion of *pitfalls of mistreating outliers* in the following section.)

Errors also occur in policy research because observers are generally not trained. This can sometimes lead to mass error. Thousands of observers may be employed for mass observations; there may result, for example, "a summation of errors by the successive processing of the data (observers at various stages not being able or trained to correct the errors of the lower stages). These errors frequently are cumulative. There is no guarantee that they will cancel."[12]

In population statistics, millions of people are sometimes omitted or counted twice. The error associated with mass observations also arises in the frequent cases where different agencies record the same activity and come up with different measurements.

Part of the reason that agencies have different approaches is that there is a lack of definition of classification of phenomena to be measured or recorded. Error is inevitable when there is uncertainty about whether a phenomenon is to be classified as an "X" or a "Y." Here again, the natural sciences are in a much better position. The theoretical characteristics of a wavelength are well understood, while the theoretical characteristics of, for example, health services and their relationship to quality of health care have yet to be established.

Employment categories are particularly hard to define. Therefore, when looking at unemployment figures, remember that at least two sources of error may be at work: doubtful classifications and untrained observers. Another problem is classifying large organizations, like General Motors. Is "auto industry" an appropriate classification for a corporation that produces automobiles as well as airplane engines, electrical appliances, and heating equipment?

THE DATA ANALYSIS PROCESS

The data that form the input to the data analysis process provide snapshots of the phenomena being studied. The information obtained from a datum will usually be of limited value. It is the general pattern that emerges from all the data that provides the first level of information required by decision makers.

A second level of information may emerge if the data contain extreme cases that appear to contradict the general pattern. Such extreme cases, called outliers, can arise for a variety of reasons. One reason may simply be that someone made a mistake or error in acquiring or processing the data. Another explanation may be that under the condition that gave rise to the outlier, a pattern different from the one originally identified describes the phenomena.

The pitfall of mistreating outliers can confront the data analyst in a number of different situations. If the outlier arises from an error, it can be discarded because it would not contradict the theory embodied by the model to be built. If it is included, it would distort the model and the corresponding theory. If the outlier is not due to an error, but reflects a true situation, then discarding it limits the model and may lead to an incorrect theory. The data themselves may not answer the question of what has given rise to the outlier. However, attention must be paid to answering this question before the data analyst falls into this pitfall.

In fitting models to data, we make a statement about the effect of different variables on a phenomenon. By explicitly including a variable in the model, we

are stating that the phenomenon behaves differently for at least some values of the variable. Variations in the data may suggest that a variable has an effect on the phenomenon. However, it is necessary to determine whether the differences observed between the different subgroups of the data defined by different values of the variables are (a) systematic differences due to the effect of the variable on the phenomena; or (b) random differences unrelated to the different values of the variable before a conclusion can be drawn. Therefore, it makes sense to include only those variables that describe patterns of "true" differences in the phenomenon.

Hypothesis testing provides the data analyst with information concerning which of the above two cases explain the observed differences between the subgroups of data. If the appropriate hypothesis test provides evidence to support the conclusion that the effects are systematic, then the differences are referred to as statistically significant differences. The process of hypothesis formulation and testing has a number of pitfalls associated with it. In general, the hypothesis to be tested is designed to answer the question: "Is the evidence strong enough to contradict a hypothesis of no difference?" or perhaps the more meaningful question: "Is the evidence strong enough to support a belief that the observed difference has the correct sign?"[13]

If a variable does not reflect statistically significant differences in the phenomena, then, in general, it will not be included in the model. The fact that a variable may differentiate statistically significant differences in the phenomena may not be reason enough to include the variable in the model. There are two reasons to avoid including every term that tests reveal as statistically significant in the model. The first reason is related to the *pitfall of unnecessary complication.* The second reason is related to the *pitfall of operational insignificance.*

The pitfall of unnecessary complication is described by Alonso[14] as "The more, the better." As Alonso points out, the more, the better is often not true in modeling, especially if greater detail entails additional calculations that compound or aggravate the errors in the data. Crissey observes that "a simple model may have better performance characteristics over the long run than a complicted model, even though it provides less information. As more detail is added, a model can become more opaque to the user, harder to understand, and more expensive to operate."[15] Tukey points out:

> In general, parsimony and simplicity will be achieved in the summary description either at the price of inadequacy of description or at the price of complexity in the model or in the analysis. Typically, those who insist on doing only primitive analysis must often be satisfied with complex—not parsimonious—summaries which often miss important points.[16]

The pitfall of operational insignificance is related to the notion of usefulness to the user. The data analyst falls into this pitfall if the assumption is made that every statistically significant difference must be operationally significant. A

variable may identify differences that are statistically significant but that in real terms cannot be acted upon or that in their interpretation provide no useful information to the user. Such differences are said to be operationally insignificant. Only the user can decide what is operationally significant and what is operationally insignificant. In some cases the analyst and user may disagree, but in such cases the burden of proof is on the analyst, not the user.

The process of the step-by-step fitting of a model to the data (as shown in Figure 1) ends when there are no longer any relationships to be found between the residuals and the independent variables and predicted dependent variables (values predicted by the current model). However, the analysis does not end at this point. To end here is to fall into the *pitfall of the premature stop.*

This pitfall occurs when the analyst forgets or neglects to analyze the distribution of the final residuals. If the distribution appears to be descriptive of a single random distribution, such as a normal distribution, the analyst can terminate the process of model building. If the distribution appears to be bimodal or multimodal, this may be a clue to the analyst that other effects exist and are yet to be accounted for. Continuing the analysis may reveal that other effects that exist were not accounted for in the initial definition of the problem. If they had been, then at least one relationship would still be indicated between the residuals and the independent and predicted dependent variables. Overlooking this clue (revealed through the distribution of the residuals), the analyst may have missed a critical factor that explains the phenomena being studied.

CONFUSING THE PAST WITH THE FUTURE

Data provide historical information about different aspects of a phenomenon. As such, they provide insights into what happened. The assumption is often made that the future will be similar to the past, and patterns are developed by extrapolating from historical data.[11] This can lead one into the *pitfall of false continuity*. This pitfall occurs when the data analyst explores historical patterns under the incorrect assumption that the future is simply a continuation of the past. In this case, deviations between predicted values and actual values provide clues that changes are occurring. But this does not help us if data are analyzed to develop models upon which future prediction and plans are based. A data analyst must always be sure to ask the question: "Is the future really going to act like the past?" or "Is the phenomenon changing slowly enough that predictions based on past data are useful in the near future?" The farther the prediction is projected into the future, the shakier will be the ground upon which the analyst walks.

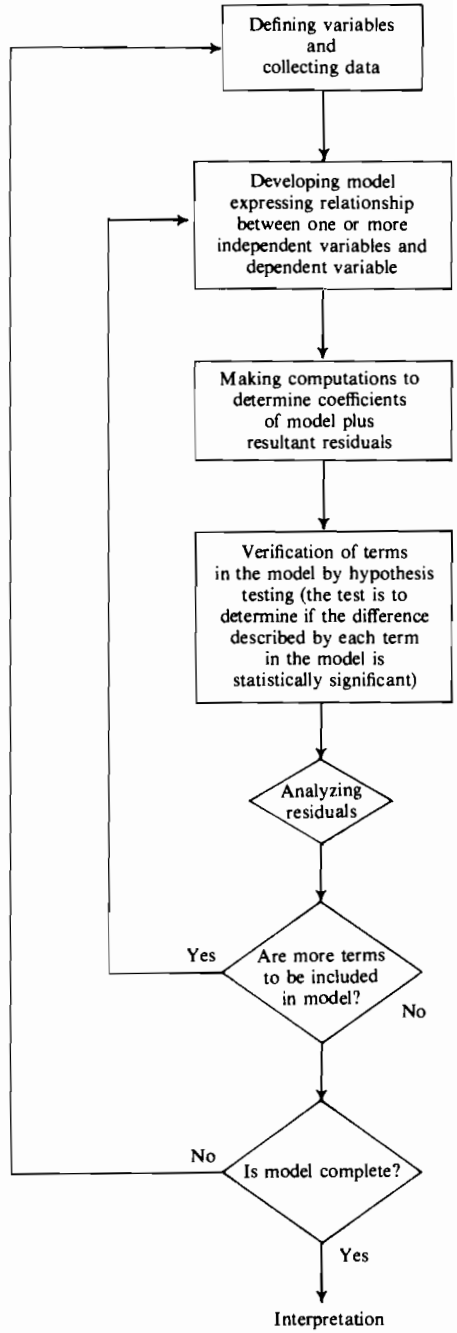

Figure 1 Flow diagram summarizing process of data analysis.

INTERPRETATION

The interpretation of data depends in part on the skill of the analyst. An important pitfall to watch out for in the process of constructing models from data is taking spurious relationships and treating them as direct, causal relationships. This pitfall is particularly tricky because it is always possible that models built from data describe relationships of "related variables" but not relationships that are causal.

> . . . [I]mports of gold and the annual number of marriages are positively correlated over years (because both reflect economic health), but suddenly increasing gold imports by government action could scarcely be expected to change the number of marriages.[17]

Zeisel presents other examples of this pitfall.[18]

After the question has been translated into quantitative terms, data analysis techniques are a bridge to the answers. Quantitative answers obtained from data analysis may require translation into qualitative terms, to ensure effective communication. The various translations that occur before and after the actual analysis offer numerous opportunities for falling into the *pitfall of misunderstanding*. When analysts describe the findings of a study of the daily demands of clients requesting services in terms of daily averages, a decision maker may reschedule personnel to meet these average daily demands without understanding the dynamics of actual variations in demand from day to day and from week to week. The pitfall of misunderstanding continually confronts analysts and decision makers when these two roles are not filled by the same individual. Modelers should always work closely with the model's user[19] to ensure that important factors are accounted for. The same reminder is required in analyzing data. Context and technique must be coordinated and integrated, if meaningful interpretations are to be made at the conclusion of the analytical process.

The separation between user and data analyst can produce effective work only if the communication between them is direct and clear. But communication between those who are technically trained and those who are not has always been a problem because of the specialized languages developed by different professions. Moreover, in extending communication beyond the inner circle occupied by professionals, one encounters the problem of differences in meaning for key words and phrases. The possibility of falling into the pitfall of misunderstanding increases as messages pass back and forth between two groups.

In the early 1970s there was strong disagreement in the United States over what policy to pursue to stimulate the economy. The executive branch of the federal government had decided to increase the money supply to achieve this end. When this policy did not produce the desired results, the Congress held hearings to study the matter. One of the analyses cited at the hearings as providing evidence that the original policy choice was justified was a statistical

analysis developed by two staff economists at the U.S. office of Budget and Management. However, while testifying at the congressional hearings, the economists stated

> that in their original paper they denied the ability or intent of their model to identify causal relations. They declared that the results can only show statistical correlations (not) cause and effect. Yet that is not how their results were received.[20]

DATA ANALYSIS: EVIDENCE, NOT CONCLUSIONS

The data analyst supplies the user with evidence from which conclusions can be drawn. The conclusion itself is not the product of analysis. Conclusions involve subjective judgments about the relative importance of different pieces of the evidence uncovered. They also involve subjective judgments about the importance of relationships uncovered. But, most important, drawing conclusions involves a willingness to take a leap of faith and make a statement about what is believed to be true even though uncertainty may surround the truth of the statement.

When all is said and done, data analysis is concerned with providing answers to questions. Every answer has many possible interpretations. The challenge of analysis is to narrow the choice down to what is most likely to occur under a given set of circumstances. But remember, most likely is not the same as absolutely certain, unless, of course, you walk across the hall afterwards and look. This is called *evaluation.*

FINAL NOTE

This discussion is by no means exhaustive. The list of pitfalls and dualities of interpretation is endless. The interested reader is directed to a number of papers for further readings.[21]

ACKNOWLEDGMENT

I wish to thank the other authors of this volume and Ms. Celesta Birnbaum, for their many helpful comments and suggestions. I would also like to thank Mr. Michael Hanson, a graduate student at the W. Averell Harriman College, for helpful suggestions.

NOTES AND REFERENCES

1. Greenberger, M., M. A. Crenson, B. L. Crissey. 1976. Models in the Policy Process: Public Decision Making in the Computer Era. New York: Russell Sage Foundation.
2. Tukey, J., and M. Wilk. 1970. Data analysis and statistics: techniques and approaches, pp. 370–389. *In* The Quantitative Analysis of Social Problems (E. Tufte, ed.). Reading, Massachusetts: Addison–Wesley.
3. Ehrenberg, A. S. C. 1977. Some rules of data presentation, pp. 305–310. *In* Statistical Reporter. Washington, D.C.: U.S. Government Printing Office.
4. Cuthbert, D., and F. Wood. 1971. Fitting Equations to Data. New York: Wiley.
5. Tukey, J. W. 1977. Exploratory Data Analysis. Reading, Massachusetts: Addison–Wesley.
   Tukey, J., and F. Mosteller, 1977. Data Analysis and Regression. Reading, Massachusetts: Addison–Wesley.
   Also see
   Altman, S. 1977. Analyzing data. *In* Productivity Handbook for State and Local Government. Washington, D.C.: National Academy of Public Administration.
6. Worswick, G. D. N. 1977. Is progress in economic science possible? Economic Journal 82: 73–86.
7. Hatry, H. P. Overview of Modern Program Analysis Characteristics and Techniques: Modern Program Analysis—Hero or Villain. Report No. 43–107–59. Washington, D.C.: Urban Institute.
8. Hanson, M. 1978. Numbers, Numbers Everywhere but Not a Fact To Have: Of the Uses and Limitations of Data. Prepared for the Top 40's Project. New York City: Urban Academy of Management.
9. Altman, S. 1976. The dilemma of data rich, information poor service organizations: analyzing operation data. Journal of Urban Analysis 3(1): 61–76.
10. Chaiken, J. M., M. W. Lawless, and K. A. Stevenson. 1974. The impact of police activity on subway crime. Journal of Urban Analysis 2(2): 173–206.
11. Morgenstern, O. 1963. On the Accuracy of Economic Observation. 2nd ed. Princeton, New Jersey: Princeton University Press.
12. Morgenstern[11] (p. 33).
13. Tukey, J. 1965. Future of processes of data analysis, pp. 671–729. *In* Proceedings of the 10th Conference on the Design of Experiments in Army Research, Development and Testing. Durham, North Carolina: U.S. Army Research Office.
14. Alonso, W. 1968. Predicting best with imperfect data. American Institute of Planners Journal 34(4): 248–255.
15. Crissey, B. L. 1974. A conceptualized framework for selecting quality and level of aggregation for dynamic computer simulation runs, p. 68. *In* 1974 Summer Simulation Conference Proceedings. Houston, Texas.
16. Tukey.[13]
17. Andersen, L. C. 1972. Statistics for public financial policy, pp. 321–325. *In* Tanur *et al.* (1972).[21]
18. Zeisel, H. 1958. Say It with Figures. 4th ed. New York: Harpers.
19. Walker, W. 1978. A review of models in the policy process. The Engineering Economist: 23(2): 92.
20. Greenberger *et al.*[1] (p. 17).
21. Other readings include

    Moore, G. H., and Julius Shiskin. Early warning signals for the economy.
    Brown, B. W., Jr. Statistics, scientific method, and smoking.
    Press, S. J. Police manpower versus crime.

Campbell, D. T. Measuring the effects of social innovations by means of time series.
Zeisel, H., and H. Kalven, Jr. Parking tickets and missing women.

The above articles and others can be found in

Tanur, J., F. Mosteller, W. H. Kruskal, R. F. Link, R. S. Pieters, and G. R. Rising, ed. 1972. Statistics: A Guide to the Unknown. San Francisco: Holden–Day.

Tufte, E. R. 1974. Data Analysis for Politics and Policy. Englewood Cliffs, New Jersey: Prentice–Hall.

Ross, H. L., D. T. Campbell, and G. V. Glass. 1979. Determining the social effects of a legal reform: the British "Breathalyser" crackdown of 1967. American Behavioral Scientist 15(1): 110–113.

Werner, W., and S. Altman. 1977. The Norfolk Health Clinic Case Study. Case Study and Teaching Note. W. Averell Harriman College for Urban and Policy Sciences, State University of New York at Stony Brook. Case 9–678–664. Soldiers Field, Boston: Intercollegiate Case Clearing House.

5 Pitfalls in the Analysis of Costs

Robert E. Bickner

IGNORING COSTS

We intend only a passing comment about the folly of ignoring costs altogether, for it is too gross a blunder to require astute exposition. Yet the pitfall is there, with its deceiving logic and enduring temptation. And the blunder is common enough.

The specious logic has two popular variants. First is the theme of the overriding objective, the transcendent goal that must be attained "regardless of costs." Second is the theme of noble concerns—concerns about lives, or human dignity, or aesthetics, or art—matters too precious to be sullied by the crass calculus of costs.

The cost analysts, however, unless they fall into some of the pitfalls mentioned here, are every bit as dedicated to ultimate goals and noble concerns as those critics who are in haste to ignore costs. The critics' undue haste reflects only an ignorance of the essential meaning of costs, for costs are nothing less than goals unfulfilled and concerns compromised. Costs are benefits lost. Costs lie in the same realm as benefits; costs have the same dimensions as benefits. Costs are in no wise less important.

The specious logic rationalizing the neglect of costs is accompanied by understandable temptation. Not only is the identification and appraisal of costs hard work, but the attention to costs often generates evidence contrary to one's pet schemes or personal interests.

The blunder of ignoring costs altogether is easily recognized, and it is a blunder more characteristic of decision makers than analysts. Less easily recognized, and more characteristic of analysts, is the pitfall of ignoring various components of costs: the pitfall of mistaking some conventional or convenient part of costs for the whole.

MISTAKING A PART OF COSTS FOR THE WHOLE

The painstaking care with which cost data are collected, the sophisticated formulae with which cost estimates are extrapolated, the expressions of caution that accompany aggregate cost projections, all distract our attention from the most critical characteristic of a cost analysis: its completeness. The challenge of cost analysis is not the development of impressively supported fine estimates of some very uncertain fraction of total costs. Such cost estimates are not simply useless; they misguide decisions, or else they reinforce the unfortunate but understandable skepticism with which many decision makers view the merits of cost analysis.[1]

Yet the temptation to produce such cost estimates is well-nigh irresistible. Like the proverbial drunk who searches for his missing wallet under the street lamp rather than in the dark, where he lost it, we search for costs where we have data. The pitfall is camouflaged by a specious reasonableness and a false modesty: the cost analysts may argue that they should restrict their efforts to within the bounds of their competence. But there is little true modesty in restricting one's attention to the easily measurable aspects of costs while presenting the results as an estimate of total cost.

Limited conceptualization of the meaning of costs adds further inducement, and further rationalization, to succumb to this common tendency to mistake a part of costs for the whole. To some people, for example, the meaning of costs is restricted to monetary expenditures. Once the probable monetary outlays associated with a proposed plan of action are identified and summed, then total cost has been estimated. The limitation of this view of cost is perhaps too obvious to warrant exposure here, but somewhat more comprehensive, yet still incomplete, concepts of cost are often accepted. Only slightly more sophisticated and adequate is the concept of cost that would include, in addition to outright expenditures, estimates of the monetary rent that might be imputed for the use of marketable durable resources: land, buildings, equipment.[2]

The experienced cost analyst will be aware that: (a) there will often be monetary expenditures relating to a decision or plan; (b) there will usually be other costs that might be readily measurable in monetary units; (c) there will often be still other costs that might be quantifiable in some way, though not in monetary units; and (d) there may yet remain costs that resist quantification in any trustworthy way. *All* of these costs need attention. Neither the allure of an inadequate concept of cost, nor the convenience of searching only where data are readily available and where units of measurement are at hand, nor the false modesty that rationalizes neglect of difficult challenges, should divert the analysts from their primary responsibility: comprehensive identification of costs. And cost means all of the sacrifices, every benefit lost, as a result of a decision.

A common, and especially enticing, variant of this pitfall is the neglect of

long-run costs. Both the analysts and their clients may be inclined toward this neglect: the analysts because long-run costs are difficult to identify and difficult to assess; the clients because decision horizons, especially where short-term political appointments or terms of office are concerned, are apt to be short.

CONFUSING DIFFERENT ASPECTS OF COSTS

CONFUSING RELEVANT AND IRRELEVANT COSTS

Having warned against the pitfall of ignoring various costs, it is now necessary to acknowledge the equally common pitfall of including, in cost estimates, irrelevant costs.

Costs are consequences of decisions. Relevant costs are those costs, only those costs, that result from the specific decision or choice being analyzed. Any expenditure, for example, or resource commitment or sacrifice, that will occur regardless of the decision being analyzed must not be reckoned as a cost of that decision.

The principle is simple enough, and every primary course in economics dwells on this principle. The distinctions between fixed costs and variable costs, between sunk costs and incremental costs, between nonrecurring and recurring costs, are all intended to fix attention on the careful separation of relevant and irrelevant costs—the distinctions in each case, or the definitions, being precisely governed by the scope of the decision under review. Whether a certain type of expenditure is considered a "variable cost" or a "fixed cost," for example, depends simply upon whether the expenditure is influenced by, or independent of, the decision being analyzed.

Yet the pitfall of including irrelevant costs in an analysis is so deceptively reasonable and inviting that only continuous care will avoid it. Here the meritorious concern for comprehensive coverage of costs can easily turn into an exaggeration of costs, and here the common habit of focusing attention on average costs can turn easily into a snare. In deciding whether or not to expand a program, for instance—in estimating the costs of the proposed expansion—fixed costs do not belong in the cost estimate. In deciding whether to complete a program or to continue a program, sunk costs do not belong in the cost estimate. And any average cost estimate is more than likely, unless we exercise great caution, to include cost elements that are simply irrelevant to the decision under examination.

This does not mean, incidentally, that resources on hand are cost-free. It means that the relevant cost estimate for resources must always be found in currently available alternative uses for them, not in past history.[3] And it means that only those sacrifices incurred as a result of the specific decision under consideration belong in an estimate of the cost of that decision.

CONFUSING THE QUESTION OF RELEVANCE WITH THE QUESTION OF CONCERN

If you decide to paint your house a brilliant purple, your decision may not only reduce the value of your house but also the value of your neighbor's house. Of course, you may decide that this doesn't concern you. Or perhaps it concerns you—but not very much. The question of concern is different from the question of relevance, and it is often a much harder question to resolve. Relevance is more of a factual matter, while concern is more a matter of volition.

It is very tempting to assume that cost analysts should restrict their attention to factual matters. Perhaps they should, but among the facts that they must confront are the concerns or lack of concern of their clients. The question of concern cannot be evaded.

Again, it may seem contrary to our earlier admonition to be comprehensive in identifying costs, but costs to other people will inevitably concern us less than costs to ourselves, and costs to some people may not concern us at all. And if cost analysts believe that they are going to avoid the arbitrary or moralistic judgments involved in this problem, they are mistaken. If they believe that they are going to concern themselves with costs to everyone, and to everyone equally, they are deceiving themselves and their clients.

Perhaps the tendency of one government agency or one community to ignore the cost impacts of its decisions on other governmental agencies or other communities is a regrettable deficiency that the cost analyst should attempt to transcend. But will the analyst's broader scope of concern have no bounds? Even if their philanthropy embraces the whole world, even if their cost calculus weighs, somehow, the interests of all living people equally, will their calculus give equal weight to future generations? All future generations?[4]

So perplexing is the problem of assigning relative weights to cost impacts on different people that cost analysts are apt to jump willingly into some pitfall as a way of hiding from the problem. They can choose from many pitfalls, some of which we have already noted, others that we will come upon shortly. They can find shelter, for example, in some very restricted concept of costs, such as monetary expenditures. But note that this concept is no more sufficient than a concept of benefits that is restricted to monetary receipts. Note also that treating all monetary expenditures or receipts as equivalent, unit for unit, regardless of who pays or receives the money, is itself a weighting system that is arbitrary and lacking in any profound justification whatever. It would be better if cost analysts, rather than hiding from the problem, confronted it directly, being careful, or at the very least conscious, of what they are doing.

CONFUSING LEVELS OF ANALYSIS

The analysis of costs may proceed on different levels. The analyst may attempt to identify the resources required to carry out a plan or decision. The analyst may also attempt to identify the most attractive alternative uses of the required

resources, or the most promising alternatives to the decision. He or she may even attempt to assess the benefits of these alternatives. None of these tasks is likely to be easy, and competent effort at any of these successive levels of analysis will illuminate the choice to be made. In the final analysis, however, the decision maker, if not the cost analyst, must attempt to assess the benefits of alternatives and weigh them against the benefits of the proposed decision.

Obviously, if different components of the cost analysis have been carried to different levels of analysis, summing up the analysis for the decision-maker requires special care. The analyst has abundant opportunity when piecing together the disparate elements of such analyses either to double count or to omit costs, or both.

The analyst also faces, not a pitfall, but a Scylla and Charybdis as he chooses among levels of analysis. On one hand, an analyst who restricts his efforts simply to identifying required resources may reasonably be faulted for evading the essential challenge of cost analysis: the evaluation of alternative benefits. Furthermore, since there will typically be a wide range of alternative resources usable in carrying out any plan or decision, the analyst must then either identify a host of options available to the decision maker or else identify the preferred, or the "least-cost," option. But this latter process itself inevitably involves an evaluation of alternative benefits. On the other hand, if the analyst boldly accepts the challenge of evaluating alternative benefits, he or she may easily be forced into making assumptions that fail to convince the decision maker fully, leaving him unsure about how much reliance to place in the analysis presented. There is no easy escape from this dilemma; only hard work, judgment, and communication with the decision maker can guide the analyst between these opposing hazards.

EXAGGERATING POSSIBILITIES OF REDUCING COSTS TO A SINGLE MONETARY DIMENSION

PRESUMING THAT A SCHILLING IS A SCHILLING IS A SCHILLING

Schillings are not all alike. Neither are marks or francs or liras. No pitfall is more tempting than the presumption that money, or monetary measures, simply because they might all be expressed in terms of some common monetary unit, can sensibly be added. Under certain conditions that procedure is reasonable, but just as often it is a dubious process, not unlike adding tomatoes, currants, and grapes after labeling them all berries. A schilling taken from your pocket is certainly not, to you, equivalent to a schilling taken from someone else's pocket, no matter how much they may look alike.

Costs, or benefits, to two different individuals cannot easily be compared, even when the costs, or benefits, can be measured in monetary units. This regrettable predicament is so easily forgotten that it wants emphasis, and the

distinction between appropriate and inappropriate additions or comparisons of monetary units may sometimes be subtle.[5]

Consider a government agency choosing between two prospective sites for a new, and unwelcome, power generating plant or trash disposal dump or vehicle parking lot. All other considerations are judged equal, but the agency, wishing to be fair, seeks to compare the costs that would be imposed on the residents neighboring one site with the costs that would be imposed if the alternative site were chosen. After a skillful survey is undertaken, it is estimated that 50,000 schillings annually would be required to compensate the residents at one site for the unwelcome intrusion, and 60,000 schillings annually would compensate the residents at the other site. Can we compare the 50,000 schillings with the 60,000, and presume the former is less?

The answer is yes or no, depending on whether or not the residents will actually be compensated. If compensation is intended, then we are comparing a prospective 50,000-schilling expenditure from the funds of the agency with a 60,000-schilling expenditure from the same funds. A direct comparison of the two amounts is reasonable. If, on the other hand, no compensation is actually intended, we are comparing costs imposed on one group of people with costs imposed on another. This is a very different matter indeed, and if, for example, the monetary income of one group is much higher than the monetary income of the other, the comparison of costs in terms of monetary units will be of dubious legitimacy.[6]

RELYING ON SIMPLISTIC SOLUTIONS TO THE PROBLEM OF COMPARING CURRENT SCHILLINGS WITH FUTURE SCHILLINGS

Everyone knows that a schilling on hand today is worth more than a schilling deliverable next year. This is another way in which one schilling can differ from a second, even though each of them is a schilling. Future schillings are discounted. The appropriate rate of discount is debated, but recommendations for governmental cost analyses typically suggest some uniform rate related to prime interest rates or government bond rates or estimates of the marginal efficiency of capital, or to something similar. There is a plausibility and a reasonableness in such suggestions that distract us from their frequent limitations.

An agency, with a limited annual budget, will have numerous options for postponing or advancing expenditures. Typically, but not always, postponements will increase undiscounted total costs while advancements will reduce them. If budgets are stringent relative to immediate needs, the agency may be incurring considerable costs to postpone expenditures—costs considerably greater than commonly proposed discount rates. Alternatively, if the agency has surplus current funds that it will lose if they are not spent, it may be advancing expenditures at little or no savings, or even at extra cost. In short,

the effective discount rate in specific situations can vary over a very wide range.

Failure to analyze the actual costs, in the specific situations, of postponing or advancing expenditures, relying instead on some conventional normative discount rate, is a pitfall camouflaged by admirable intentions. Economic efficiency would be served by having actual discount rates uniform, rather than widely variant, and the use of uniform discount rates in cost analysis is recommended in the hope of promoting this optimal harmony. But actual, effective rates can be brought into balance only by shifting budgets, not by legislating uniform rates for cost analysis. Seeking optimality is a virtue. Presuming optimality is a blunder.[7]

Crucial to the question of the appropriate discount rate to use in a particular situation is the question of who the decision maker is. The meaning of costs must always be found in alternatives, and decision makers at different hierarchical levels will have a different range of alternatives available. The effective discount rate must be found within the options available to the relevant decision maker.

Perhaps an example will clarify these points. If an agency is postponing needed repairs to the roof of a building, because of a temporary stringency of funds, it may find that the cost of repairs next year, along with the cost of damage due to the faulty roof, will exceed by 50 percent the cost of repairing the roof immediately. In this circumstance, it would be inappropriate for the agency to expend a comparable sum of money on some other maintenance project that could be postponed a year at a cost of only 25 percent, notwithstanding the possibility that some economist might estimate the social discount rate, or the opportunity cost of capital, to be only 10 percent. On the other hand, if we focus critical attention on some higher level of authority—some authority capable of shifting funds to this agency from some other agencies—then it is certainly possible that the project saving 25 percent should be funded (as well as, but not instead of, the project saving 50 percent).

The discount rate issue is often intertwined with two other matters, risk and uncertainty, and also with the issue of inflation, requiring their careful disentanglement by the analyst lest he become entangled himself. If the discount rate issue takes the extreme form of comparing costs, or benefits, accruing to future generations with those accruing to the current generation, then all of the intractable difficulties involved in interpersonal comparisons, mentioned in the preceding section, come into prominence. Discount rates can obscure this problem more easily than they can resolve it.[8]

OVERRELIANCE ON MONEY AS THE MEASURE OF COSTS

It may be true, as frequently reported, that money can buy anything—but it is also true that there is often legitimate dispute over the proper price. It is, of

course, part of the analyst's challenge to measure costs wherever possible, and, whenever possible, to reduce these measures to some common denominator. Money is, by all odds, the most frequently applicable measure and the most common of common denominators. It should not be disdained. But the measure, like the medium, can corrupt us.

This is not the place for a disquisition into the complexities of money and monetary measures, but the analyst must have a keen awareness of these complexities. Money has a familiar face and a deceptive simplicity that invites naive use and overreliance. The ingenuous analyst can easily be lured into faulty applications of monetary measures, or into neglect of alternative or supplementary measures where they might be advisable, or into neglect of cost elements not readily included in monetary measures.

For example, the use of market prices as estimates of costs, while often appropriate, is frequently inappropriate without careful review and adjustment. As a case in point, market prices, if competitive, will reflect marginal costs. However, if we are reviewing any plan or program large enough to have significant market impact, large enough to affect economies of production or to alter supplies and prices, then actual costs may be higher or lower than current market prices would suggest. Noncompetitive prices are even more likely to be misleading. Market prices will often be influenced by market imperfections, including government-introduced constraints. Depending upon the concerns of the decision maker in those circumstances, it is very likely that such market prices will either understate or exaggerate costs.

Exclusive reliance on monetary measures can be a fault as much as misapplication of monetary measures. Translation of other measures into monetary measures is a worthy ambition, but if the procedures employed to make the translation are insubstantial or highly disputable, then the usefulness of the entire cost analysis can be diminished. Other measures can usefully supplement, and sometimes displace, monetary measures of costs. For many smaller optimization problems, for instance, cost impacts in other terms such as failure rates, or kilograms, or hours, or liters, or units of output, or frequencies, or indices of one kind or another, may be more convenient and meaningful than monetary cost measures, provided that sight is not lost of relevant possibilities of substitution.[9] And for many major decisions or large-scale projects, the only meaningful procedure for comprehending and assessing costs may lie in a direct exploration of alternatives rather than in an effort to calculate monetary expenditures or losses. Certainly the marginal prices generated by the marketplace are not going to tell us the cost of a program that diverts half of a nation's scientific research capability into some limited area of effort, or the cost of a policy that redirects a large fraction of college students from the pursuit of medical education into other pursuits, and so on. Similar examples could be adduced without end.

Yet the most common pitfall created by our fascination with money is the

simple neglect of costs—costs that are not automatically, or at least not easily, grasped in our monetary calculations. Money is a most useful measure of costs, but if the focus of our attention shifts from the ultimate target, costs, to the means for measurement, money, the risk of omission of important costs becomes large.[10]

Before leaving this section, we must attempt to avoid a misinterpretation. Nothing said herein is intended to support the supposition that it is immoral to attach monetary assessments to sacred virtues or noble values—lives, for instance. We save money daily at the risk of life. Innumerable governmental decisions trade off monetary expenditures, on one hand, versus risk to life on the other. Refusing to look at the monetary evaluations implicit in our decisions does not exalt life: it only increases the likelihood of our making inconsistent trade-offs and, hence, making nonoptimal decisions.

FAILING TO COMPREHEND AND CONFRONT THE ESSENTIAL CHALLENGE OF COST ANALYSIS: THE IDENTIFICATION AND EVALUATION OF ALTERNATIVES

The purpose of analysis is to guide decisions. The purpose of analysis is to illuminate the choice among alternative courses of action. The purpose of analysis is to help identify and evaluate the differences that will result from choosing one course of action rather than another. Perhaps so obvious and fundamental an observation doesn't need such reiteration, but an analyst who has lost sight of his ultimate destination is in a very deep pitfall indeed.

It should not be presumed that analysts who have stumbled into this pitfall will experience either loneliness or vexation. They may, instead, find both company and comfort there, for the pitfall is a haven from challenge.

Costs, like benefits, must always be found in a comparison of alternatives: i.e., in a comparison of two or more different chains of events, or different states of the world, that would result from different choices. Failure to recognize that cost analysis is an analysis of alternatives, a comparison of options, leaves the analysis without its essential guidance and direction. Variant facets of this common failure, variant forms of this pitfall, include those briefly noted below.

GIVING INADEQUATE ATTENTION TO ALTERNATIVES

The alternatives relevant to a particular plan of action may not be explicit or obvious. This characteristic of alternatives in no way diminishes their paramount importance; it only invites their neglect. And if the cost analyst is unaware or unconvinced of their importance, their neglect is virtually assured.

Yet the alternatives are there, explicit or implicit, recognized or denied, attended or neglected, clarified or confused, and the quality of the cost analysis rests upon the meaningfulness of the alternatives presumed.

The implicit character of most alternatives is only one of the inducements to their neglect. The difficulty of discerning the relevant alternatives is another. The remarkable, but mesmerizing, performance of money as an indicator and evaluator of alternatives is yet another.

Money is so frequently and marvelously effective in identifying and assessing alternatives that, in cost analysis as in life, money often becomes the end rather than the means. Money is a useful measure of costs because, but only because, and only to the extent that, it indicates and evaluates the relevant alternatives. The importance of sophisticated understanding, rather than naive faith, in judging the capacity of money to perform this feat has already been stressed.

For those who insist that the essence of costs is to be found in money rather than in alternatives, we might add some incidental observations. Note, first, that the only reason you hesitate to spend money is because of its many alternative uses. Note, second, that even if cost analysis were simply the addition of a long list of monetary expenditures, the analyst would still need some key to determine which specific expenditures were, and which were not, appropriately attributable as costs of the plan being analyzed. The method, if rational, will be to determine which expenditures would change if the plan were not implemented. But that determination depends upon the alternative presumed. The fact is that we only know how to estimate costs of choosing one course of action rather than some alternative. We know how to evaluate decisions only by assessing the alternatives available.

PRESUMING THAT THERE IS ONLY ONE ANSWER TO THE QUESTION OF COSTS

If there is more than one pertinent alternative to a plan of action, then there is more than one pertinent answer to the question: What will it cost? If you and I have divergent perceptions of the pertinent alternatives, then we will have divergent perceptions of the costs of the plan. If there is some irreconcilable dispute over the pertinent alternative, then there is an irreconcilable dispute over the costs of the plan. And if we don't know the relevant alternatives, then we don't know the costs of the plan.

It may be that some nominal, neutral, inert alternative can be used as a common baseline against which to measure costs of several competing plans of action. This procedure will, of course, generate a unique estimate of the cost of a plan, but it should be recognized that (a) costs measured against a nominal alternative are only nominal costs; (b) the nominal alternative is as likely as the other alternatives to need careful delineation; and (c) nominal alternatives such as "do nothing," or "leave matters to the marketplace," are rarely as simple as they seem. In any event, the specification of a nominal alternative will not tell

us the real costs of a plan, and the specification will inevitably be arbitrary and variable.[11]

Other considerations should further discourage the presumption that there is only one valid answer to the question of costs. Since costs must be found in alternatives, and since the range of available alternatives depends upon the authority, power, perception, influence, and persuasiveness of the decision maker, it also follows that there will be different answers concerning the cost of a plan depending upon the locus of the decision. Higher levels of authority, for instance, may have different alternative uses for resources utilized in a plan or different means for implementing the plan. The meaning of the costs of the plan, and meaningful estimates of these costs, will differ accordingly.

Yet another reason for expecting multiple cost estimates is, of course, the uncertainty of the future. This uncertainty leads not only to different cost estimates in a simple sense but also to related issues of risk acceptance or avoidance, and issues related to the assessment of strategies for enhancing the flexibility of plans.

So far we have simply noted that there may well be more than one relevant meaning to the cost of any decision. We must now also note that any singular meaning can have more than one meaningful measure. We are referring now to the problem of concern, discussed earlier. Different clients of cost analyses will inevitably be concerned with different costs. Even if there is only one relevant alternative to a decision, and even if that alternative is clearly and precisely defined, the question of how much weight or concern to attach to different kinds of costs, or to costs falling on different people, remains. Different, but valid, cost estimates will result.[12]

BELIEVING THAT COSTS ARE OBJECTIVE AND SIMPLE

Costs, like benefits, are inevitably matters of conjecture. We do not mean simply that costs must be anticipated or projected. Even in retrospect, even after a decision has been made and implemented and after history has run its course, costs are a matter of conjecture. An estimate of the consequence of any decision is inevitably a comparison between two contrasting courses of events: one course that follows upon the decision if it is, or was, made, and the contrasting course that will follow, or would have followed, upon an alternative decision. At least one, if not both, of these courses of events will be conjectural. Costs, like history, require skillful interpretation.

The popular illusion that cost estimation is objective and simple, or that it is relatively simple compared with the estimation of benefits, is easily maintained. The cost analyst can, if he wishes, confine his efforts to those assignments in which the meaning and the measurement of costs are relatively easy. Or he can, if he wishes, define costs as those considerations relating to a decision that are easily identified and measured. These procedures are common enough to

have obtained the status of orthodoxy. But to those who are willing to confront, rather than to diminish or evade, the challenge of cost analysis, costs are not simple, nor is cost analysis routine. Analysts who understand the essential nature of costs will also appreciate the complexity of the analytic challenge. They will conduct their analyses with both imagination and caution—and interpret their results with great care.

NOTES AND REFERENCES

1. The point is perhaps included in one of Enthoven's ten maxims for policy analysis: "Number 5. It is better to be roughly right than exactly wrong."
   Enthoven, A. C. 1975. Benefit–Cost and Policy Analysis. Chicago: Aldine.
2. There is no intent here to deny that cash flows, or financial liquidity, may frequently be a special problem demanding special attention. But this frequent concern with current monetary expenditure does not justify a grossly inadequate conceptualization of costs. Other items besides cash flows will also frequently demand special attention, items such as manpower availability, time, space constraints, energy consumption, water supplies, political capital, and good will. None of these items, however critical they may be, constitute the meaning of cost.
3. In an otherwise excellent book, Seiler states: "The concept of inheritance leads directly to the concept of incremental cost, which means that only the requirements beyond those that can be inherited are charged to the new system. . . ." Actually, the correct concept is that the relevant cost of inherited items lies in their alternative uses, not in their historical acquisition costs. See
   Seiler, K. 1969. Introduction to Systems Cost-Effectiveness. London: Wiley, p. 6.
4. If the ambitious analysts believe they know how to give appropriate weight to the interests of the unborn, they might tackle an even more difficult challenge: In appraising family planning programs, what weight should be given to the interests of those who might, as a consequence of the program, never be born?
5. A very interesting discussion of the problems and considerations involved in weighting the interests of different individuals is found in Chapter 8 of
   Frost, M. J. 1975. How to Use Cost Benefit Analysis in Project Appraisal. Guildford, Surrey: Biddles.
6. A concise summary of the "compensation principle" and of its relevance in comparing alternatives can be found in
   Cohn, E. 1972. Public Expenditure Analysis. Lexington, Massachusetts: Heath.
7. A very helpful treatment of discount rate issues can be found in
   Beesley, M. E. 1973. Urban Transport: Studies in Economic Policy. London: Butterworth.
   On page 107 Beesley notes the difference between basing discount rates on estimates of what *will* happen versus what *ought to* happen, and on page 108 he affirms that the dilemma, if it is to be resolved, "must include a view of the *consequences* of requiring a particular rate to be adopted in a real decision context."
8. Churchman, pondering the difficult problem of balancing interests of future generations with our own, expresses skepticism about the appropriateness of some use of discount rates, and he alludes to his "feeling-deficient" friends who say: "The hell with the values of a generation yet unborn, or at least 10 percent the hell per annum." Churchman, C. W. 1974. On the facility, felicity, and morality of measuring social change. *In* Accounting for Social Goals (J. L. Livingstone and S. C. Gunn, ed.). New York: Harper and Row.

9. In assessing alternative ways to improve reliability of a spacecraft by adding redundant equipment, for example, the most meaningful simple indicator of comparative costs may be the weights of alternative items of equipment, and hence their comparative effects on payload, rather than their monetary costs. Of course, any simple indicator must be used with caution.
10. To emphasize this risk, Anthony Downs lists 22 categories of losses (costs) imposed by urban highway and renewal projects other than construction costs represented by monetary expenditures.
    Downs, A. 1970. Uncompensated nonconstruction costs which urban highways and urban renewal impose upon residential households. *In* The Analysis of Public Output (J. Margolis, ed.). New York: Columbia University Press.
11. The appropriate alternative is identified by Squire and van der Tak of the World Bank: "In many cases the situation without the project is not simply a continuation of the status quo, but rather the situation that is expected to exist if the project is not undertaken. . . ."
    Squire, L. and H. G. van der Tak. 1975. Economic Analysis of Projects. Baltimore: John Hopkins University Press, p. 19.
12. To choose an example almost at random, Borus and Tash, in analyzing costs of manpower programs, develop separate estimates for costs to (a) participants, (b) employers, (c) government, and (d) society as a whole.
    Borus, M. E. and W. R. Tash. 1970. Measuring the Impact of Manpower Programs. Wayne, Michigan: The University of Michigan—Wayne State University.

6 Pitfalls in Optimization Methods

E. M. L. Beale

This chapter is devoted to pitfalls in the use of optimization methods. An applied systems analyst spends more time using known optimization methods than in developing new ones, and a discussion of pitfalls in algorithm development would require more mathematics than is appropriate to this volume. Throughout the chapter, it is assumed that the optimization methods discussed will be implemented on a computer. Given any optimization model, one should certainly think about whether it can be solved analytically. If it can, though, the model probably does not capture the essence of the problem being studied. For example, analysis may show that the optimum solution to the problem as posed is to set all the decision variables to zero or, alternatively, to set them all to their upper bounds. It is a pitfall (into which I sometimes fall) to assume that all real problems must be more difficult than this. But if a problem is subjected to systems analysis, it is nearly always because there are partially conflicting objectives that cannot be resolved so easily. In the preliminary stages of a study, rough calculations made without a computer can help to fix the range of data and parameter values of interest in the main study, and some such calculations can be regarded as applications of optimization methods. But the real benefits from optimization come when its systematic methodology is applied to problems that involve the interaction of so many factors that intuitive methods are ineffective. And these benefits require automatic data handling and computation.

THE PURPOSE OF OPTIMIZATION

The immediate purpose of optimization is to find the optimum solution to some explicit model of a real situation. Thus, a discussion of the practical value of

optimization must be concerned as much with the development of such optimization models as with optimization techniques or methods.

Optimization models may be classified as either established or new. An established model is run from time to time with updated input data as part of some operational decision-making routine. The purpose is then to suggest a specific course of action to management, and the suggestion will usually be accepted. A new model may also be used in this way, but it is more often used to gain greater understanding of the situation. The model may be run under a variety of assumptions that lead to different conclusions, and the model itself will not suggest which set of assumptions is most appropriate.

The use of established optimization models is relatively free from pitfalls. It should not be too hard to find out whether the optimization process leads to decisions that are about as good as, or better than, the decisions that would otherwise be taken, and then to assess whether the optimization process is cost-effective. This assessment should take account of the savings realized from not having to carry out alternative procedures, and of the relative speeds of the different decision-making procedures. The use of formal optimization may also improve management's ability to react to new circumstances, since these may require only a small change in the optimization model, or perhaps just a change in the data, while other methods of decision making may lead to decisions that are no longer appropriate.

An optimization model can be used to explore the appropriate reaction to a large number of hypothetical situations, and the understanding provided by this exercise may well justify the initial investment in the development of the model. But this is not a reason for using the model on a day-to-day basis.

Systems analysts are more often concerned with new situations, and the use of optimization may then be more controversial. Koopman (1956) drew attention to the pitfalls of "linearitis" and "maximitis," and yet linear programming is now widely accepted as a very successful tool in applied systems analysis. Why is this? It is partly because linear programming provides a true optimal solution to a wide class of problems that could previously be solved only approximately. But much more important, it is because the concepts of mathematical programming provide a convenient framework for thinking about the types of data that are relevant to the problem at hand, and a focus for collecting these data. The mathematical programming model is then used by the project manager to coordinate the outputs from other studies. The model will produce an optimal policy according to some explicitly stated criterion, given some first approximation to the data. This policy will often be unacceptable. But the calculations will show where the data are unrealistic, which additional features need to be considered, and perhaps where additional data need to be collected. Because optimization tends to produce extreme solutions, and because it requires a precise specification of the policy options, it is a much more powerful tool for model development than, say, simulation.

There is therefore nothing wrong with making exact calculations on inexact data. On the other hand, very elaborate calculations on inexact data may be unwise for the following reasons.

One must try to avoid a situation where everyone working on a project is so busy with the mechanics of obtaining numerical solutions that no one has the time or energy to think about whether these solutions are reasonable, or whether some assumptions or some numerical data need to be changed. These mechanics may involve the development of a computer program to carry out the optimization process itself. More often, they involve the development of computer programs to read the data in a convenient format, to express them in the formats required to carry out the optimization using some standard software such as a mathematical programming system, and to print out both the data and a report on the solution in convenient formats. These programs are usually called matrix generators and report writers. In addition to this programming effort, the sheer process of feeding a large number of alternative sets of data into a complex computer package and assessing the solutions can take a remarkable amount of time and energy, particularly if the matrix generator and report writer have been designed or implemented casually. This serious pitfall can often be avoided only if the whole project team works very hard while the project deadline is still some time in the future. And this is perhaps contrary to human nature.

A related pitfall is to underestimate the number of times that an optimization model needs to be run, with alternative sets of data and probably with some changes in the model structure. Perhaps the most common cause of failure of optimization models is that analysts, and their clients, are so frightened of over-simplification that they produce a model that is too expensive to be used for more than a few runs. They then fail to obtain any real understanding of the problem.

One can try to avoid expensive runs by truncating the iterative process used to solve the problem and accepting an intermediate result as an "approximate" solution. This is rarely sensible. The amount of computer time saved is usually small unless one is willing to accept a possibly very poor approximate solution. And the pattern of the solution may be obscured in an approximate solution even if the objective function is not severely degraded. This can greatly reduce the value of the work. In integer programming applications one must often accept a "good" solution that is not guaranteed to be optimum because it would take too long to complete the tree search. This is unfortunate, but at least the solution is conditionally optimum, given possibly arbitrary assignments of values to some of the integer variables.

OPTIMIZATION TECHNIQUES

The cost of developing and running a model may depend crucially on whether it can be solved using some standard optimization technique, and more specifically on whether the analyst has access to effective computer software implementing the technique and to people who know how to use it.

Optimization techniques can be classified into five groups:

- Linear programming
- Integer programming
- Nonlinear unconstrained optimization
- Nonlinear constrained optimization
- Dynamic programming

LINEAR PROGRAMMING

Linear programming minimizes an objective function of continuous variables subject to equality and inequality constraints. In principle, the objective function and constraints must be linear. In practice, a nonlinear objective function and nonlinear inequality constraints can be handled by piecewise linearization provided that both of two conditions hold:

- The objective and constraint functions are defined globally as explicit functions of the decision variables.
- The objective function and the constraints are convex, or, in other words, the model does not represent any economies of scale.

Linear programming on large computers is a relatively reliable technique if the number of constraints is less than, say, 5,000 and the number of decision variables is less than, say, 50,000. But remember that many constraints and decision variables may be needed to represent "obvious" aspects of the model in an explicitly linear form. Software is available for solving much larger linear programming problems with particular special structures, such as network flow problems, which include the classical transportation problem.

INTEGER PROGRAMMING

Integer programming is a direct generalization of linear programming in which one can include variables required to take integer values, and particularly zero–one variables, representing the omission or inclusion of some activity. Special ordered sets can best be considered as part of integer programming. These allow nonconvex nonlinear constraints, and therefore economies of scale, again with the proviso that all functions are defined explicitly.

Integer programming is a less reliable technique than linear programming, in that some types of model cannot be solved in a reasonable amount of computer time if they contain more than, say, 100 integer variables. However, many models containing several hundred integer variables are solved routinely. And if good results can be obtained on a realistic set of test data, then they can usually be obtained with new sets of data for the same class of problem, provided that the dimensions of the problem are not significantly increased.

NONLINEAR UNCONSTRAINED PROGRAMMING

Nonlinear unconstrained optimization techniques find local minima of functions defined only by subroutines that give the function value, and perhaps the vector of first derivatives, and again perhaps the matrix of second derivatives, for any set of trial values of the decision variables. Reasonably reliable methods for twice-differentiable functions (whether or not the derivatives are computed explicitly) exist for up to about 100 variables, and many real problems are solved with a few thousand variables. Methods for nondifferentiable convex functions have been developed and have been found to work well on some problems but not on others.

NONLINEAR CONSTRAINED PROGRAMMING

Nonlinear constrained optimization, or nonlinear programming, is a simple extension of unconstrained optimization if all the constraints are simple lower and upper bounds on individual decision variables. The problem becomes somewhat more difficult with general linear constraints, and somewhat more difficult again with general nonlinear constraints. But even in this case, software for handling hundreds of constraints and variables has been developed, and some much larger models have been solved accurately enough for practical purposes.

DYNAMIC PROGRAMMING

Dynamic programming is more a modeling technique than an optimization technique: apart from one extension discussed in the section on dynamic programming below, it solves problems expressed in terms of finding the shortest path through a network. Once the problem is in this form, the task of finding a global optimum solution is quite easy if the numbers of nodes and arcs in the network are not too gigantic. The classical dynamic programming formulation is one in which the nodes are grouped into "stages" in such a way that all arcs lead from a node in one stage to a node in the next stage. The nodes at any stage then represent the possible states of the system at this stage, and the lengths of the arcs represent the cost of moving from one state in one

stage to the appropriate state in the next stage. The shortest-route formulation implies that the states are defined so that this cost does not depend on how the previous state was reached.

Dynamic programming is sometimes described in terms that make it seem much more general than other branches of mathematical programming. But for computational purposes this is misleading. Instead, dynamic programming should be regarded as applicable to a narrow range of important problems, some of which cannot easily be solved any other way. The technique works well if the number of states at each stage is less than about 10,000. Unfortunately, this is not a large number if several independent variables are needed to define the state: for example it allows for up to ten possible values of each of four variables. Sometimes the difficulties can be overcome by using approximate methods.

Special software is not usually needed for dynamic programming.

THE RELATIONSHIP BETWEEN THE FORMULATION AND THE SOLUTION METHOD

Having outlined the reasons for optimizing, and the available techniques, we now turn to the formulation of optimization models. It is a pitfall to confuse the activities of finding out what the real problem is and those of deciding how to solve it. This is in a sense a generalization of Koopman's linearitis. But, having decided what the problem is, one should choose a mathematical structure to represent it that makes computing as simple as possible while capturing the essence of the problem. This generally means representing the partially conflicting objectives.

To illustrate this point, consider two different problems concerning the production and storage of a number of products. The analyst may be asked to help make an annual plan when demands for the products are seasonal. The problem here is to compromise between a policy that produces to meet the demand as it arises, which may minimize storage cost at the expense of high production costs, and a policy of steady production rates of all products, which may minimize production costs at the expense of high storage costs. A multi-time-period linear programming model using deterministic estimates of all demands, capacities, and costs may well be adequate for this problem, although the actual stock levels used will include an allowance for uncertainties in the supply and demand for materials and for operational flexibility. But if the analyst is asked to help choose the best compromise between holding expensive stocks and the risk of failing to meet unexpectedly high demand, then some explicit representation of the stochastic nature of demand is essential.

Sometimes the simplest model with the required logical structure needs data that seem so abstract to the client that he cannot supply them. A more dis-

aggregated model may then be necessary. But we must resist unnecessary disaggregation. Inexperienced clients sometimes feel that features of their operation that are important to them must be included in any model "to make it more realistic." But realism in a mathematical model is not a primary requirement, and the test must be whether the features are really important to the decision that the model is being used to study. The dangers of being misled by an oversimplified model may be the most obvious ones at an early stage of the project. But, if realistic data for the simple model can be collected, then the dangers of having a model that is too cumbersome to be used effectively, discussed above, are probably greater. A model built by a committee is particularly prone to excessive detail. The model will need to incorporate information from many sources, so this pitfall can be avoided only if there is someone in charge of the project who understands the danger and has the authority to do what is needed.

Even in the initial discussions before detailed model-building starts, the analyst probably has some solution technique in mind to help him to structure his thoughts. Nevertheless, he must not shut his eyes to relevant information that cannot be processed by this technique, so he should be in no hurry to make a final choice.

Even the general concepts of mathematical programming, namely, decision variables and explicitly defined constraint and objective functions, can blind the analyst to stochastic aspects of the problem. This may be inevitable. There are no general modeling approaches that allow effective optimization when the stochastic information evolves as successive decisions are made. Multistage stochastic programming can be applied only to very simple models.

Although dynamic programming allows for some random variables, these must be statistically independent of all previous random variables, given the values of the current state variables. Representing uncertainty in such a naive way may be even more misleading than omitting it from the formal part of the analysis. Informal guidance can often be obtained from sensitivity analyses of the effects of changes in the data.

Even if a mathematical programming approach is accepted, there is no reason to think specifically of linear programming if there are relevant data that can be expressed more conveniently in nonlinear or integer terms. Once the structure and approximate values of the data have been established, the analyst should think seriously about whether a purely linear programming model, or even a network flow model, might be adequate. If so, the tasks of developing and running the model are simplified, and the model becomes less dependent on specialized computer software and therefore more portable.

Two more specific pitfalls in optimization modeling are worth noting. One is that the ratio of two variable quantities is rarely, if ever, a good way to express an objective function. For example, one may be interested in the rate of return on some portfolio of investments, but to choose the portfolio to maximize the

ratio of net revenue to capital employed can lead to the omission of useful investments because this omission reduces the denominator by a more significant amount than the numerator. It is nearly always better to ask the client to specify the value of capital, or whatever is measured by the denominator, possibly as a function of the amount of capital used. One can then maximize the difference between the value of the net revenue, or whatever is measured by the numerator of the original ratio, and the cost of the resource measured by the denominator.

The other pitfall is to forget that there may be many solutions with the same, or almost the same, value of the objective function. There is usually no reason to prefer one such solution to another, but, if there is a reason, it should be incorporated in the model. So, for example, if a linear programming model is used monthly to allocate the production of materials at different sources to different destinations, it may be important to include the information that, other things being equal, it is more convenient to supply a destination from the same source as last month. Otherwise, the model is quite likely to recommend changes for trivial reasons.

This is a convenient place to point out two contrasting pitfalls in the computer implementation of models. It seems inevitable that while a model for a practical problem is being developed, and while the initial data are being collected, the detailed requirements will change. One serious pitfall is for the analyst to allow changes in the specification during the initial computer implementation of the model. Even apparently trivial changes can have unexpected ramifications, delaying the production of working programs and increasing the risk that the developed programs will contain errors. But an almost equally serious pitfall is for the analyst or his client to become so impressed with the difficulty and danger of program changes that he continues to use an unsatisfactory implementation of the model. Once the initial development is complete, changes may be very worthwhile. The input can often be simplified. It may be found that some quantities that were originally conceived as individual items of data are in fact computed from other more basic data. Provided these basic data are not much more voluminous, they should be read instead, and the computations consolidated into the model. Experience may suggest the desirability of introducing more data checks, or, on the other hand, of making some data formats more flexible, even if this means removing some checks.

The output can often be improved in the light of experience. And changes in the mathematical structure of an optimization model are often easy as long as they do not require a change of solution technique. Indeed such changes often cause far less trouble than in a simulation model. This is probably because optimization requires that the model structure be set out in a more disciplined way than does simulation.

DEFINITION OF THE MODEL

It seems appropriate to say something about the task of formulating an optimization model, although this cannot be reduced to simple rules. Beale (1968) suggests that the constraints can be classified as availabilities, capacities, qualities, demands, and material balances and that the activities, or decision variables, can be classified as buying, making, and moving. This classification is still valid and helpful, although there is a good case for treating demand constraints merely as special types of material balance constraints and for treating "making" as another type of activity whose characteristic feature is that it has a negative coefficient in one or more capacity constraints.

Confusion over signs is an obvious pitfall in any mathematical work. So it is wise to have a rigorous convention about signs in material balance constraints. The usual convention is to use a positive coefficient for activities that bring material in and a negative coefficient for activities that take material away. Thus the general form of the constraint is

Stock on hand from the previous time period (or initial stock)	+	Stock bought, transferred in, or made	−	Stock sold, transferred out or used	−	Stock on hand at the end of the time period	=	0

This is very natural. The only aspect that is at all counterintuitive is the signs attached to the storage activities, but these follow inevitably from the rest of the equation. It is worth noting, however, that there is a superficial contradiction between this sign convention and the convention that says that a capacity constraint takes the general form:

Capacity used is less than or equal to capacity available.

Thus the use of capacity is represented by a positive coefficient while the use of a material is represented by a negative coefficient. An availability constraint can be regarded as a material balance constraint except that it is written with the sign conventions of a capacity constraint to remove the negative coefficients.

Signs in the objective function can also cause trouble. Unless instructed otherwise, mathematical programming systems maximize the dummy variable x_0 if the objective function row is written

$$x_0 + \sum_j a_{0j} x_j = b_0$$

So a positive coefficient a_{0_j} in the objecuve function row represents a cost, and a negative coefficient a revenue. But a positive constant term b_0 represents a

fixed revenue. This operation of maximizing x_0 is, of course, equivalent to minimizing the expression

$$\sum_j a_{0j} x_j - b_0$$

It is recommended that storage activities indexed by time period t should refer to the end of this time period and not the start. Initial stocks then have a subscript of zero, which is natural, and one can normally avoid any activities indexed by a time period after the last one.

One rather technical point about linear programming formulations so often causes trouble that it seems worth mentioning. Typical production activities give only a single output, and these are naturally represented by single decision variables, even though these variables may have negative coefficients in several material balance rows if they consume several raw materials in fixed proportions. If the activity produces one or more by-products together with the main product, however, many people feel the need to represent the by-products by additional decision variables. This feeling must be suppressed. By-product production is represented by positive coefficients in the appropriate material balance row in the same way that raw material use is represented by negative coefficients. Of course, if there are alternative modes of production giving different products, or even the same products in different proportions, this is a very different matter, and one must introduce separate decision variables representing the different modes.

One pitfall with optimization models, and particularly with large-scale mathematical programming models, is that the detailed formulation and implementation can easily become unintelligible to anyone except their authors, who may overlook something vital, particularly if the model needs amendment. A first step away from this pitfall is the adoption of a systematic way of documenting models. The approach recommended by Beale *et al.* (1974) works well. This is based on the fact that an intelligible detailed formulation of a large-scale model depends on the compactness of an algebraic formulation using subscripts and summation signs. By including summation over one or more subscripts, one can write a single line of algebra to define an equation connecting many variables. Furthermore, if other subscripts not covered by summation signs are included, the same single line of algebra can define several similar equations.

It is therefore natural to base the computer implementation of a mathematical programming model on an algebraic formulation. But this formulation itself can easily become unintelligible if it is not set out systematically. The following order of presentation is recommended.

1. *Subscripts.* Use lowercase letters, with the corresponding capital letter representing the maximum possible value of the subscript.

2. *Sets*. These may be needed to give precise definitions of ranges of summation, or conditions under which a constraint, variable, or constant is defined. Use capital letters.
3. *Constants*. Specify the physical dimensions. Use capital letters.
4. *Variables*. Specify the physical dimensions. Use lowercase letters.
5. *Constraints*.

Avoid using any letter to mean more than one thing. If there are too few different letters to use mnemonic single letters for all quantities, then use capital letters as literal subscripts to define composite letters. Such literal subscripts never take numerical values, while lowercase subscripts always take numerical values. So, for example, C_{Mfp} may denote the manufacturing cost for product p at factory f (in £/tonne), and C_{Ddfp} the distribution cost to destination d (in £/tonne).

This may seem a perverse approach, since with a very simple model one can go straight to the constraints and then explain them by defining the constants and variables. The subscripts are then defined implicitly in the course of the other definitions. Some of the reasons for the recommendations are therefore spelled out.

A definition of all the subscripts immediately indicates the overall scope of the model. If the symbols are then used consistently, the subscripted constants and variables can be understood more easily and defined more compactly since the subscripts need not be reinterpreted.

Although sets may not be needed in simple models and will generally be developed at a late stage in the formulation, they often provide the only compact way of giving a precise definition of domains of summation, or conditions for the existence of other entities. It seems natural to define them immediately after the subscripts to which they refer.

The constants, variables, and constraints are normally developed in parallel. The constants are listed first because they are part of the conditions of the problem and logically precede the decision variables, which represent the approach to its solution.

The structure of the constraints is clarified if the notation distinguishes between the constants, or assumed quantities, and the decision variables to be determined by the model. This may not be clear from the context. For example, the model may take the amounts of each product to be made either as data or as quantities to be determined. Confusion is avoided by always using lowercase letters for decision variables and capital letters for constants. This notation conflicts with that used in more theoretical work, where one is concerned with comparatively few types of subscripts and variables, and where one can reserve letters near the beginning of the alphabet for constants. But this approach is not convenient for applied work. On the other hand, in applied work there is no merit in using capital letters to denote matrices, since constants and variables

often have three or more subscripts and in any case their physical significance is clarified if the subscripts are written explicitly.

In applied work, quantities are sometimes given names consisting of two or more letters, as they are in computer programming languages. But it seems best to maintain the mathematical convention that multiplication is implied when two letters are written side by side and not as subscripts. These ideas can be reconciled by writing the name with all letters after the first as literal subscripts. These letters are written in capitals to distinguish them from subscripts that take numerical values, which are always written as lowercase letters.

These literal subscripts can be appended to constants, variables, and sets. They can, if necessary, also be appended to subscripts. Subscripts may also take numerical subscripts if several quantities of the same type are needed in the same constraint. So the amount of material transported from location i_1 to location i_2 may be written as $y_{i_1 i_2}$, and if the demand at location i is D_i, we may have a constraint

$$\sum_{i_1} y_{i_1 i} - \sum_{i_2} y_{i i_2} = D_i$$

Similarly, if c_t denotes the capacity introduced in time period t and x_t the capacity used in time period t, we may write the capacity constraint as

$$x_t - \sum_{t_1=1}^{t} c_{t_1} \leqslant C_0$$

where C_0 denotes the initial capacity. But this form of constraint, relating the current situation back to the initial situation, may represent a formulation pitfall, since the constraints often contain significantly fewer nonzero coefficients if one time period is related to the previous time period. The constraint can be written as

$$x_t - \sum_{t_1=1}^{t} c_{t_1} + e_t = C_0$$

where e_t is an "explicit slack," and if the constraint for time period $t-1$ is subtracted from it, we have the constraints

$$x_1 - c_1 + e_1 = C_0$$

$$x_t - x_{t-1} - c_t + e_t - e_{t-1} = 0 \quad (t \geqslant 1)$$

This constraint has fewer nonzero coefficients than the original one when $t > 3$.

Given the algebraic formulation of the model set out in this way, the assuming that the values of the constants can be read into the computer, it is a straightforward task to write a matrix generator computer program to write the data in the format required by a mathematical programming system such as MPSX. Indeed, the task is so straightforward that it can be automated. Scicon, for example, has written a matrix generator generator program, which produces a matrix generator program, given the algebraic formulation and instructions for computing the coefficients in the form of FORTRAN arithmetic statements or function-type subroutines.

Whether the matrix generator program is produced automatically or manually, Beale *et al.* (1974) recommend that one should try to keep the documentation of the algebraic formulation up to date when the model is changed. They were wary of proposing rigid standards, since they knew the pressures under which practical program developers work. But subsequent experience has shown that it is quite easy and very desirable to accept this requirement unconditionally: whenever the algebraic structure of the model is changed, however small or however urgent the change, the master copy of the documentation of the algebraic formulation must be updated before the computer program is altered.

LINEAR PROGRAMMING

For many good reasons, linear programming is the most widely used optimization technique in systems analysis. Reliable software for solving quite large linear programming models exists for many computers, and there are remarkably few pitfalls in its application.

The advantages of linear programming can be summarized under three headings:

- Ease of modeling and data collection
- The fact that one can easily find a global optimal solution
- The fact that sensitivity analysis is easy, and in particular that an immediate by-product of the solution is a set of shadow prices, or Lagrange multipliers, representing the effects on the objective function of small changes in the constraints

It is sometimes suggested that the reason people use linear models is simply that they cannot solve nonlinear models. This would not be a bad reason except that it ignores recent progress in numerical methods for nonlinear programming. The real reason, however, has much more to do with the feasibility of data collection and a judgment of what is relevant. It may be inexact to use a linear material balance constraint as spelled out in the preceding section. This ignores the possibility that breakages, or theft by employees, may depend in a nonlinear way on the level of activity in the store. But even though it may sometimes be desirable to allow for these "losses," it is rarely sensible to consider them in such detail. Capacity constraints on multipurpose machinery are another example of a linear constraint simplifying reality by allowing for the time lost changing over from one product to another in an inexact way. But we try to avoid representing these changeovers in detail, not only because it would make the optimization much more difficult but because it would make data collection much more difficult.

This situation can be summarized by saying that any fool can think of a reason why a linear model may not represent the real world exactly. But this is rarely useful unless he can offer some quantitative correction to the linear model. The only real exception occurs when the nonlinearity is the essence of the problem even though it is hard to quantify. In a model for farm planning, one decision to be considered may be the amount of fertilizer to be applied to the farmer's grass. A linear model is likely to conclude that if fertilizer is useful at all, one should apply as much as possible. It seems necessary to allow for the nonlinearity of the yield as a function of fertilizer level, even though the data on this nonlinearity may not be very reliable. Fortunately, such effects can be accommodated within a linear programming model by using piecewise linearization.

It is important to be aware that quite large linear programming problems can be generated (using a matrix generator program) and solved easily. From a logical viewpoint, it is also important that all linear programming problems are convex and that any local optimum is therefore also a global optimum. Practical linear programming models usually have between about 500 and 1,200 constraints and a rather larger number of variables. It is interesting that these model sizes are much the same as those used 15 years ago, even though computer capabilities for solving linear programming problems have increased vastly. These increased computer capabilities owe at least as much to improvements in mathematical methods for handling sparse matrices as they do to improvements in computer hardware. So, although it is conventional to measure the size of a linear programming model by the number of rows (essentially the number of constraints), because the solution time increases as the cube of the number of rows (other things being equal), the number of nonzero coefficients in the matrix is much more significant. This means that one may be very disappointed by the computer's performance on a model with many more than six nonzero coefficients in an average column. The problem will require more computer storage than a typical problem with the same number of rows and columns. The time per iteration of the simplex method will be increased, the number of iterations may be greatly increased, and there may even be problems of numerical accuracy. Sometimes these difficulties can be overcome by introducing extra variables defining intermediate quantities that occur in many constraints. Adding to the numbers of rows and columns in this way is well worthwhile if it drastically reduces the number of nonzero coefficients.

It is a pitfall to think that no one could ever reasonably want to generate a linear programming model with several thousand constraints. Such models arise all too easily in studying operations that extend over several locations and several time periods. So the art of model building in linear programming consists of reducing the "natural" problem size to something manageable by aggregation or by using other simplifications in the formulation.

Careless scaling of the numbers can cause difficulties with any optimization method, but it seems best to consider this problem with each method individually. Modern linear programming codes are fairly robust against bad scaling. In particular, they have some automatic rescaling option, and they also provide some variant of the dynamic scaling algorithm introduced by Harris (1973). The codes still retain a number of fixed tolerances, so it is still desirable that the magnitudes of the nonzero coefficients in the original matrix should lie between 0.01 and 100 as far as possible. This can usually be achieved by scaling the units in which constraints and activities are measured by suitable powers of 10^3.

INTEGER PROGRAMMING

Integer programming is a useful approach to optimization problems involving economies of scale. Many large integer programming problems are now solved routinely. Thus the task of developing an integer programming model for a new application should be approached optimistically. This optimism must be cautious, however, since the amount of work required to solve the problem in principle increases exponentially with the number of integer variables, and many models in which the solution of the linear programming relaxation of the problem does not give good guidance toward the corresponding integer solution have proved useless.

Difficulties with integer programming models can sometimes be overcome by reformulation. Williams (1974) provides some illuminating examples. Integer programming is not as robust a technique as linear programming, and although we can use it to find the best combination of subsystems, we must try to avoid using it to find the best options within subsystems when these can easily be found by other means. As an extreme example, in a model used to study the economic viability of a system of offshore gas-gathering pipelines, it is natural to introduce integer variables representing alternative possible starting times for a shore terminal, and integer variables representing alternative possible starting times for the different links in the pipeline system. If there is only one possible link to some shore terminal in the system being considered, then it stands to reason that this link and the terminal should start operating together. But an integer programming model that treats these as separate decisions can waste a great deal of time before it recognizes the fact.

A more general pitfall in integer programming formulations concerns constraints of the form

$$\sum_j x_j - M\delta \leqslant 0$$

where δ is a zero–one variable. This is a very powerful modeling device for imposing the constraint that each x_j must be zero unless $\delta = 1$. Problems with

many such constraints tend to be very hard to solve, however, particularly if the numbers M are larger than they need to be. Williams notes that these difficulties can often be alleviated by disaggregating the constraint. Specifically, if the upper bound M_j on an individual x_j is lower than the upper bound on their sum, then it is probably better to write

$$x_j - M_j\delta \leqslant 0 - j$$

even though this may greatly increase the number of constraints.

Once the formulation is settled, the user can sometimes still have a significant effect on the progress of the algorithm by varying the solution strategy. In particular, if the standard strategy does not work too well, he should consider assigning a priority ordering to the integer variables and special ordered sets.

Integer programming methods tend to work less well on purely combinatorial optimization problems, so it is important to remember the basic arithmetic of combinatorial optimization. A problem with n independent yes-or-no decisions has 2_n possible solutions. It can therefore be solved by enumerating them. If $n = 10$, this means enumerating just over 10^3 possibilities, which may well be feasible. If $n = 20$ there are just over 10^6 possibilities, and enumeration is probably infeasible. The idea of branch-and-bound methods is to avoid the need to examine all possibilities. But if this approach saves 99.9 percent of the work required for a complete enumeration, there are still over 1,000 problems to be solved if $n = 20$, and the amount of work doubles when n is increased by one. Fortunately, there are many applications where branch-and-bound is much more effective than this, and problems with hundreds of integer variables can be and are solved.

The task of finding the optimal permutation of a set of elements increases even faster with the number of elements. Since 6! is only 720, one can often handle up to six elements without considering anything more sophisticated than complete enumeration. Even 10! is only just over 3×10^6, and this might be a manageable number. But every single extra element then increases the amount of work by a factor of more than 10, so the approach cannot then be considered practical using even the fastest computer. This is the starkest example of the pitfall of thinking that the computer can do arithmetic at effectively infinite speed.

UNCONSTRAINED OPTIMIZATION

Unconstrained optimization is more rarely used in systems analysis than the techniques discussed so far. The most important applications are in fitting models to data by choosing parameter values that minimize some function of the residuals, which are defined as the deviations of the observed values from

the fitted values. Special methods can be used for such problems, based on local linear approximations to the individual residuals. The function is often taken as the sum of the squares of the residuals, and the problem is then usually solved using the Gauss–Newton method as modified by Levenberg (1944) and Marquardt (1963). These methods usually work well, but if they fail to converge in a number of iterations of the same order as the number of unknown parameters, then the problem can often be solved more effectively as a general unconstrained optimization model.

The main pitfall in this type of unconstrained optimization is to rely on the computational technique to cope with an unnecessarily awkward formulation. While there is no reason to abandon a model in favor of another model that makes less physical sense just because it is easier to compute, there is every reason to express the model in terms of different parameters if the effects on the fitted values of changing one parameter could be largely counteracted by compensating changes in other parameters. This pitfall is discussed constructively by Ross (1970).

NONLINEAR PROGRAMMING

There are so many types of nonlinear programming problem, and correspondingly so many methods for nonlinear programming, that it is hard to identify general pitfalls.

One pitfall is to use a primitive general method, particularly one that does not require derivatives, and to rely on the power of the computer to overcome any lack of sophistication in the algorithm. In practice, round-off errors can easily cause the algorithm to stop some way from any local optimum.

Penalty function methods are convenient for small problems if one has a good unconstrained optimization code, but they are intrinsically weaker than other methods. And one must beware of the inherent ill-conditioning of the problems as the penalty parameter tends to zero. Various ideas have been proposed for avoiding this difficulty, but it is doubtful if they will ever be competitive with the best methods that use explicit linear approximations to the constraints. These methods are mostly based on the reduced gradient method introduced by Wolfe (1963) and generalized to nonlinear constraints by Abadie and Carpentier (1969).

DYNAMIC PROGRAMMING

The major pitfall of dynamic programming is that, even if the problem can be accommodated within the computer, the amount of computing required may

be excessive. Careful thought must thus be given to whether some of the calculations are being repeated unnecessarily. In other words, there may be common elements in the calculation of the effective lengths of different arcs in the network.

As noted in the section on optimization techniques above, it is a pitfall to think that dynamic programming deals only with shortest-route problems. A minor extension of the standard algorithm solves optimization problems on a tree in a sense that can perhaps be best explained by considering a specific example.

Suppose that one has the design for a network of pipes in which material enters the network at a number of sources and proceeds by a unique route to a single sink. A number of nodes can be identified in this network. There must be a node at each junction point, and we may introduce additional nodes. Without loss of generality, we may assume that only two pipes join at a junction, since multiple junctions can be represented as neighboring simple junctions. The problem is to choose some property of the network that depends on the "state" at each node. For example, we may need to choose the diameters of each pipeline link, and the state at the node may represent its depth. The cost and feasibility of any link may then depend on the diameter chosen and on the depths at the upstream and downstream ends of the link. Such problems can be solved by dynamic programming, by defining a value function $V_n\ (x_n)$ for each node n, depending on the state x_n at this node, as the minimum achievable cost for all links upstream of node n, given that the state at this node is x_n.

CONCLUSION

In spite of the pitfalls noted in this chapter, numerical optimization techniques are becoming increasingly useful to applied systems analysts. This is partly the result of the development of understanding of the strengths and weaknesses of the techniques and of appropriate modeling strategies. It is the result of the development of reliable and convenient computer hardware and software, so that models can be developed and implemented quickly. If the first numerical results are unexpected or unsatisfactory, there is then plenty of time to correct the data, or even restructure the model if necessary.

One final pitfall is to dismiss numerical optimization as an old-fashioned and unchanging aspect of systems analysis. To use it effectively, the analyst may well need to take advantage of current developments in mathematical modeling techniques, in numerical algorithms, and in computer software design.

REFERENCES

Abadie, J., and J. Carpentier. 1969. Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints, pp. 34–39. *In* Optimization (R. Fletcher, ed.). London: Academic.

Beale, E. M. L. 1968. Mathematical Programming in Practice. London: Pitman.

Beale, E. M. L., G. C. Beare, and P. Bryan-Tatham. 1974. The DOAE reinforcement and redeployment study: a case study in mathematical programming, pp. 417–442. *In* Mathematical Programming in Theory and Practice (P. L. Hammer and G. Zoutendijk, ed.). Amsterdam: North-Holland.

Harris, P. M. J. 1973. Pivot selection methods of the Devex LP Code. Mathematical Programming 5: 1–28.

Koopman, B. O. 1956. Fallacies in operations research. Operations Research 4: 422–426.

Levenberg, K. 1944. A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Mathematics 2: 164–168.

Marquardt, D. W. 1963. An algorithm for least squares estimation of non-linear parameters. SIAM Journal of Applied Mathematics 11: 431–441.

Ross, G. J. S. 1970. The efficient use of function minimization in non-linear maximum-likelihood estimation. Applied Statistics 19: 205–221.

Williams, H. P. 1974. Experiments in the formulation of integer programming problems. Mathematical Programming Study 2: 180–197.

Wolfe, P. 1963. Methods of nonlinear programming, pp. 67–86. *In* Recent Advances in Mathematical Programming (R. L. Graves and P. Wolfe, ed.). New York: McGraw-Hill.

7 The User's Perspective

Laurence E. Lynn, Jr.

Whether based in government or in private organizations, systems analysts are recognized as playing an important role in the making of public policy. Because most policy choices are complex, public officials often need help in sorting out the issues and alternatives and in marshaling evidence concerning their likely consequences.[1] Systems analysts provide that help by bringing the knowledge, methods, and ideas of the scientific community to bear on the tasks of designing and choosing public policies.

Yet decision makers are often dissatisfied with the usefulness of systems analysis. In part, their criticisms are aimed at incompetently done studies or at analysts who lack skill and experience. However much of their dissatisfaction is more fundamental. It is directed at studies and advice that, though professionally competent, they view as misguided and irrelevant. It is the latter type of dissatisfaction with which this chapter is concerned.

Some see dissatisfaction of this sort as rooted in the sharply different ways that policy makers and policy analysts think. Laurence H. Tribe, for example, argues that

> in part because of their particular form of commitment to a posture of detached and impersonal objectivity, the policy sciences tend to partition and warp reality in certain patterned ways, generating a fairly understandable, and sometimes quite unfortunate, system of blind spots and distortions.[2]

The reality to which he refers, and in which policy makers are enmeshed, is characterized by the inseparability of facts and values, of private values and the public interest, of political process and policy outcomes, and of decision making and moral judgment. Analysts, Tribe is saying, fail to come to grips with that reality.

This is an important issue. To prepare the ground for identifying pitfalls from the perspective of a user of analysis, the next section of this chapter

explores different models and images of decision making in the public sector. With this as background, the discussion moves to an identification of pitfalls in systems analysis from the perspective of a government official facing a complex policy decision. Conclusions follow.

DECISION MAKING IN THE PUBLIC SECTOR

RATIONAL ACTORS AND CLASSICAL ANALYSIS

Many ideas about the application of systems analysis have been derived from a particular paradigm of the decision-making process and of the role of analysis in shaping decisions. Christened the "rational actor model" by Allison[3] and "the problem solving image" by Rein and White,[4] this idealization characterizes the decision maker as a goal-oriented seeker after the knowledge that will enable him or her to make "good policy." Through a rational process of identifying specific objectives and evaluating alternative ways of achieving them, the decision maker chooses policies and allocates resources toward their achievement. Rational decision makers, furthermore, are participants in a policy process that is purposeful.

> Implicit in [the] basic characterization of problems and solutions [by traditional policy analysts] is the image of someone or some collective mechanism that is actively involved in steering, making the policy decisions that guide the ship of state. The image involves both a steersman and a steering wheel well connected to the rudder.[5]

The task of the systems analyst in supporting the rational decision maker is to do good analysis in the classic sense. E. S. Quade characterizes such analysis as follows:

> The idea of an analysis to provide advice is . . ., in concept, . . . simple and rather obvious. One strives to look at the entire problem, as a whole, in context, and to compare alternative choices in the light of their possible outcomes. Three sorts of inquiry are required, any of which can modify the others as the work proceeds. There is a need, first of all, for a systematic investigation of the decisionmaker's objectives and of the relevant criteria for deciding among the alternatives that promise to achieve these objectives. Next, the alternatives need to be identified, examined for feasibility, and then compared in terms of their effectiveness and cost, taking time and risk into account. Finally, an attempt must be made to design better alternatives and select other goals if those previously examined are found wanting.[6]

More precisely, the elements of analysis, "present in every analysis of choice,"[7] are the objective (or objectives), the alternatives, the costs, a model (or models), and a criterion.

This paradigm of decision making is popular with systems analysts partly because it is compatible with the methods of rational inquiry that they use and partly because of its normative implications. Belief in the problem-solving

image rests, according to Rein and White,[8] on "the conviction that policy expresses a theory of action that includes: (1) a definition of the problem; (2) a set of possible courses of action; and (3) a goal or goals one seeks to achieve."[9] Thus, in this view, decision makers and systems analysts are engaged in identical cognitive activities. Moreover, the argument goes, policy making *should* be conducted this way if the public interest is to be served. The making of public policy depends to a large and increasing extent on the clarification of objectives and on the efficient use of resources, the achievement of which is beyond the reach of judgment unaided by systematic analysis of ends and means.[10]

In the classical formulation of systems analysis, pitfalls are defined as sources of errors in logic or reasoning that stem from weaknesses in the theoretical foundations of systems analysis and from lack of training and experience on the part of the analyst.[11] Quade further categorizes pitfalls as internal—those inherent in the analysis—and external—those concerned with getting analysis used.[12]

Internal pitfalls are those that decision makers and their advisers typically have in mind when they say, "This is a poor study." Pitfalls of this type include inadequate treatment of uncertainty—the inability of the analyst to answer questions of the form, "How would your results be affected if . . .?"; failure to treat all considerations relevant to the decision at hand; failure to formulate the problem in a way that is responsive to the decision maker's concerns; bias or parochialism exhibited at any stage of the analysis; and excessive attention to technique or to detail to the neglect of the policy issue or problem.

External pitfalls include failure to get the message across to the client. Such failure may stem from arrogance on the part of the analyst and from the analyst's having to bear a message the client does not want to hear; failure to anticipate organizational resistance to change; and "substituting a model for the decision maker," i.e., discounting or neglecting factors outside the analyst's model that may have an important bearing on the decision.

The origin of the idea of a dichotomy between internal and external pitfalls is perhaps to be found in the distinction drawn in the early systems analysis literature between analysis and judgment. This distinction has often been interpreted as justifying a hierarchical separation of analytic considerations and policy-making considerations. Systems analysis should be done in accordance with the highest standards of scientific excellence and the results submitted to, or injected into, the processes of political decision making. That is, analysts make their judgments, then policy makers make theirs, sacrificing "objective" values in the interest of achieving political ends or "second best" solutions.

This interpretation often leads to the view that non-use of competent analysis can be attributed to flaws, often termed "irrationalities," in the policy-making process or to mistakes by decision makers. Politics are seen as obstacles to rational decision making and to the uncompromised use of good analysis.

Although some policy decisions in some systems of government are the province of a single decision maker, in which case "rational actor" or "problem-solving" models might be appropriate, most are not. That the rational actor model is not an adequate characterization of most public sector decision making is widely recognized. A number of alternative conceptualizations of decision making have been developed that represent the nature of the policy-making process with greater realism. Each has important implications for the practice of systems analysis and for the identification of pitfalls.

Some conceptualizations are concerned with explaining policy outcomes in terms of interactions among social institutions, groups, and individual actors sharing the power to influence decisions. One type of model, associated with Charles E. Lindblom, identifies the interdependence of ends and means, the prevalence of value conflicts, the dispersion of power, and the processes of incremental adjustment and compromise as essential characteristics of policy-making systems.[13] Rein and White describe a variant of this model according to which "policy formation involves (1) problem setting as a step toward issue setting; (2) mobilization of the fine structure of governmental action; (3) achievement of settlements in the face of dilemmas and trade-offs among values."[14]

Other conceptual frameworks are concerned with the actions of decision makers within organizational settings. According to Alexander L. George, for example, a "policy making system is comprised of *individuals* who often come together in *small groups* within the framework of an *organization* that is characterized by hierarchy, division of labor, and specialization."[15] He goes on to note that impediments to the use of analysis and other information are to be found in

> (1) the dynamics of individual behavior; (2) the ways in which small, policy-making groups are structured, their internal processes managed, and their communication with outsiders regulated; (3) the patterns of behaviour that emerge in complex organizations.[16]

Allison has developed two alternatives to the rational actor model, each of which incorporates important elements of reality. According to his "organizational process model," governmental behavior can be understood less as a product of deliberate choices and more as the outputs of large organizations functioning according to standard or routinized patterns of behavior. In contrast, his "bureaucratic politics model" explains government behavior in terms of political bargaining along regularized circuits among players positioned hierarchically within the government.[17]

Combining the ideas of complex policy-making systems and complex organizations leads to a view of policy making considerably different from the problem-solving image. The objectives and values of participants are not given

and immutable, nor is the establishment of goals the first stage in policy making. Often values are created and objectives are selected in the course of choosing among alternatives. Conflict and compromise are the methods of decision making, and rational calculation—to the extent that it occurs—is a diffuse, difficult, and largely unobservable activity.[18]

For the most part, the implications of these alternative models of decision making for the practice of systems analysis—and for the identification of pitfalls—have not been systematically spelled out. Some observers have offered general advice. Tribe, for example, advocates that analysis move

> in the general direction of a subtler, more holistic, and more complex style of problem-solving, undoubtedly involving several iterations between problem formulation and problem-solving and relying at each stage on the careful articulation of a wide range of interrelated values and constraints through the development of several distinct "perspectives" on a given problem, and couched in an idiom true to its internal structure rather than translated into some "common denominator."[19]

Lindblom argues that "analytical methods cannot be restricted to tidy scholarly procedures."[20] The analyst must recognize, and incorporate into his or her work, strategies and dodges reflecting the actual behavior of decision makers: "satisficing" instead of maximizing; proceeding incrementally; focusing on specific bottlenecks and the elimination of specific disadvantages. Especially useful in this context are the views of those who have at different times been both producers and users of analysis. This largely anecdotal body of lore is a rich source of insights into the practice of analysis based on insiders' perspectives on the decision-making environment.

LESSONS FROM EXPERIENCE

Alain Enthoven, for example, formerly the Assistant Secretary for Systems Analysis of the U.S. Department of Defense and a distinguished practitioner of systems analysis, has distilled from his experience in government ten practical principles for systems analysis.[21] For example: "A good analysis will search out and highlight the key questions of values, the uncertainties and the intangibles, and not bury them."[22] A good analyst will "keep it simple"[23] and recognize that "It is better to be roughly right than exactly wrong."[24] A good analyst will "consider the basis on which decisions are actually being made, and ask whether [he or she] can improve on it,"[25] and will "think critically and realistically about the prospects for implementation."[26]

Following service as the Director of the U.S. Bureau of the Budget, economist Charles Schultze echoed the general theme of Enthoven's principles:

> "Systematic" in the term systematic analysis has more than decorative meaning. It implies that problems should be analyzed in the context of the relevant environment. Successful analysis of the federal government's domestic programs does not isolate facts or relationships from the system in which they exist. It does not ignore the feed-

backs between the results and initial conditions or major spillovers of costs or benefits. It avoids solutions that have a high risk of failure because they run directly across the grain of the system on which they are imposed.[27]

A number of officials and systems analysts who worked in the U.S. Department of Health, Education, and Welfare (HEW) during the 1960s and 1970s have furnished advice to analysts. Economist Alice M. Rivlin, formerly Assistant Secretary for Planning and Evaluation at HEW and the first Director of the U.S. Congressional Budget Office, has written:

Analysts would be wasting time and effort if they gave high priority to making dollar estimates of the benefits of social action programs, for politicians and decision makers are unlikely to pay attention to them. They and their constituents have strong, intuitive ideas about the relative importance of health, education and social well-being that are not likely to be shaken by benefit–cost estimates.[28]

Rufus E. Miles, Jr., for 12 years HEW's Director of Administration, says:

The . . . role of the assistant secretary for Planning and Evaluation requires sophistication in understanding the competing political forces that act on and through the Congress, as well as in economic analysis. Enormous amounts of energy may be wasted if it is devoted to enterprises that have no realistic chance of acceptance by the Congress.[29]

After assessing a group of studies done in the early days of the use of systems analysis by HEW, reporter Elizabeth Drew called attention to recurrent problems:

Some of the studies are more conscientious than are others about pointing out assumptions which are only guesses. . . . Others at times reach levels of near unintelligibility. . . . Some belabor the obvious, are super-repetitive (we shall discuss . . . we are discussing . . . we have just discussed), are littered with references to arcane studies, leave the reader to find the page which explains the chart, and serve up these vague euphemisms at which bureaucrats specialize. . . . Finally, and most disconcertingly, some of the figures in the charts on benefit–cost ratios have been afflicted with typographical errors.[30]

Constantine Menges, a former HEW analyst, writes in a similar vein of his experiences with evaluation studies:

Reports are often esoteric, highly detailed, and not easily digested by high or lower level practitioners. . . . Decision-makers expecting clear and broad answers have been disappointed when they encountered poor products or poor reports that were too narrow in scope.[31]

The common theme running through these observations and admonitions is that systems analysts must base their work on a deeper understanding of how policy is made and on clearer insights into how government works. Doing useful analysis is not only a matter of technical proficiency but also of good judgment and good instincts in identifying issues and designing policies. But how can such advice—which sounds like a vague counsel of perfection—be

made operational to the practicing systems analyst? What are the specific pitfalls they should avoid?

A useful place to begin answering these questions is with Giandomenico Majone's view that "[u]ncritical acceptance of the metaphysical distinction between ideas and action is probably the most serious pitfall of applied analysis."[32] He suggests that analysts who implicitly accept distinctions between logic and action, or between analysis and politics, are apt to place "excessive reliance ... on normative decision models and the pure logic of choice, on analysis rather than synthesis and design,"[33] thus ignoring or mishandling important aspects of the reality with which decision makers must contend. In this context, Majone defines a pitfall as "a *conceptual error* into which, because of its specious plausibility, people frequently and easily fall."[34] He argues that

> The real significance of the category "pitfall" for systems analysis is best appreciated in relation to the craft aspects of the field. ... Good analytic work cannot be produced mechanically.... "Style" plays as big a role in determining the value and acceptability of the analytic product as it does in the results of the craftsman's work.[35]

From the perspective of decision makers, then, Majone's view is that systems analysts can be viewed as craftsmen employing their skills to create useful analysis. In addition to mechanical or technical mistakes, decision makers are apt to be disturbed, and alienated, by failures of craftsmanship—the exhibition of an irrelevant style.

In fundamental respects, this formulation goes to the heart of the matter. Decision maker dissatisfaction with analysis is often based on the feeling that the analysts have failed to relate the practice of their craft to the decision maker's needs and situation, to what Schultze termed "the relevant environment" and what Enthoven termed "the basis on which decisions are actually being made." In this context, Majone's definition of pitfall might usefully be extended to include *errors of perception* into which, because of their analytic convenience, systems analysts frequently and easily fall.

In what follows, I shall organize the identification of pitfalls in terms of Majone's four components of the analyst *qua* craftsman's task: (1) problem setting, data, and information; (2) tools and methods; (3) evidence and argument; and (4) conclusions, communication, and implementation. In illustrating pitfalls, I shall try wherever possible to give examples that conform to the following criteria: There is an identifiable analytic product—a particular study or group of related studies; an explicit purpose of the analysis was to be useful to decision makers; the analysis was competently done except for the pitfall(s); there is evidence of reaction by decision makers or their representatives; and these reactions are to the study's success in producing relevant information and do not constitute acceptance or rejection on essentially arbitrary grounds. Where all criteria are met except that decision maker reactions are

unknown, I will supply plausible grounds for criticism from a user's perspective.

DEFINING THE PROBLEM

This initial phase of the analytic task involves "translation of a policy situation into a policy problem," the development of the questions to be addressed in the analysis, and identification of the data and other information to be sought and used.[36]

Without doubt, this phase is the greatest source of pitfalls for the analyst. During this phase, training is least relevant, and instinctive feel for the decision-making process most important. "Because of the unstructured character of a problem situation," Majone points out, "imagination, judgement, and analogical and associative thinking play a bigger role than rigor and technical skills."[37] Faced with the inherent difficulty of structuring a complex problem, analysts are tempted to trim the questions to the available data, to make both questions and data collection conform to methods they know or prefer, or to frame the problem in accordance with preconceptions about the issue. Among the most common types of pitfalls under this heading are addressing the wrong question, failing to consider the institutional and political context, and failing to identify an adequate range of alternatives. A typical illustration is the following.

- The Secretary of the U.S. Department of the Interior faced a decision on whether to continue construction on the Bonneville Unit of the Central Utah Project (CUP), a large, multipurpose irrigation, municipal, and industrial water development project. The Bureau of Reclamation, the agency responsible for the CUP, needed the Secretary's authorization to award contracts for construction of Currant Creek Dam, the next part of the Bonneville Unit to be built. The Secretary asked for an analysis of the controversial project before proceeding.

His analysts studied the issues and alternatives and presented their findings. Completion of the Bonneville Unit would, they argued, create problems that could easily be avoided. Much less expensive and less environmentally damaging alternatives were available to meet the most important water needs in central Utah. Moreover, implementing these alternatives would avoid troublesome questions concerning the water rights of the Ute Indian tribe. Therefore, all construction should be stopped and the CUP reformulated.

But, the Secretary responded, you are describing the consequences of completing the entire Bonneville Unit. Suppose we built only Currant Creek Dam. Would it be useful by itself, i.e., without constructing any other tunnels, aqueducts, or pumping plants? If so, and if building it would not pose serious cost, environmental, or Indian water rights problems, wouldn't I be better off to

go ahead with it while taking more time to study the bigger issues? Enamored of the task of evaluating the entire Bonneville Unit, the analysts had not considered a narrower possibility that was attractive to the Secretary. Under political pressure to go ahead with construction and facing other issues that were more important to the Department, he had wanted to know whether he could simply duck the bigger issues for the time being. In the end, the answer was yes.

The point is not that the analysts erred in evaluating the Bonneville Unit as a whole. Their mistake was in getting carried away with it and neglecting to consider less exciting but more practical alternatives. Their neglect cost them an important measure of credibility.

• An analyst completed a careful study of the role of foreign medical school graduates in the U.S. health manpower picture. While discussing his recommendation that the U.S. Department of Labor terminate the "critical skill" designation that permitted large numbers of foreign medical graduates to enter the U.S., he was asked how approval of this recommendation would affect overall U.S. immigration policy, and how foreign governments facing a drain of other types of skilled manpower to the U.S. would react to such a precedent. Preoccupied with studying U.S. health manpower issues, the analyst had failed to consider questions about which the responsible official would inevitably be concerned. Indeed, the dilemma neglected by the analyst was that health manpower officials with the greatest concern for the problem were located in another agency and had little power to do anything about it, whereas the officials with the necessary power to act saw the problem in a very different light.

• Even when addressing the right questions, systems analysts may fail to appreciate the institutional and political context in which the decision will be reached. For example, analysts of welfare reform in the United States have seen their efforts go for naught in the Congress time and time again because the logic of their analysis has failed to fit the perspectives of elected officials. Of President Richard M. Nixon's proposal to reform U.S. welfare programs, Vincent and Van Burke wrote:

> Its unorthodox origin, the drawing board of economists, plagued the Family Assistance Plan [the name of Nixon's proposal] for it denoted a national plan aimed at directly reducing poverty, not a plan to attract other constituencies for their own self interests. And when it reached Congress, the politicans and special interest groups fell upon it.[38]

William A. Morrill, following service as Assistant Secretary for Planning and Evaluation at HEW, during which he supervised an analysis of welfare reform, summarized his impressions of the Family Assistance Plan (FAP) experience.

. . . I think [my] staff felt strongly that the FAP designers really hadn't understood what it was they were up to well enough, not because they were stupid but because the state of the art was such that they took some important compromises far too early in the game. And in some ways they destroyed the chance for the plan very early on. The original FAP design hadn't really dealt much with the work problem. It sort of came on them subsequently, and then they began to say and do things that they hadn't thought through very well and they got themselves into all kinds of problems.

Another dimension which they hadn't thought through seriously enough was the administrative one: where they were assigning responsibilities for what. For instance, as soon as they decided they had to do something federal about the work test and its administration, that brought with it all the problems of saying, "Well, what are you going to do about the training, and what are you going to do about the day care." And then that got pulled up into the federal level; yet, if we knew anything, we knew that federally administered manpower programs were hardly a thriving success. And the system began to get clogged up with bad roles and missions that inevitably dumped a whole bundle of cost on the federal structure that wasn't there before. And the FAP designers just lost control of the problem at that point.[39]

• A similar but more complex example is that of a study of the San Francisco, California, housing market conducted as part of the city's Community Renewal Program. The overall objective of the program was the development of a comprehensive, integrated program of public and private actions aimed at improving the living environment of the city, with special concentration on the housing market. The task of the analysts, then, was to develop a framework that adequately replicated the operation of the private housing market.

Among the problems facing the analysts was the choice of a geographical boundary for the study. They chose the legal–jurisdictional territory of the City of San Francisco, since their charter applied to the city only. However, they realized they were on the horns of a dilemma. The relevant housing market was larger than the city. Plans developed for the city alone would be greatly weakened by what they necessarily left out of the picture. On the other hand, plans developed for the natural social unit (the region) would be irrelevant, because there was no decision-making entity with the necessary power to act or to induce the cities or the region to act. The model builders were apparently in a no-win situation.

In evaluating this experience, Kenneth L. Kraemer points out that:

In practical terms, . . . the types of analysts selected to work on a problem should often include not only model builders and experts in the issue involved, but also organizational and behavioral scientists who understand the social and institutional context in which decision is arrived at and action eventually carried out.[40]

In a broader vein, Kraemer notes that:

Generally, analysts (model builders, planners, and other professionals, etc.) do not face problems as other decision makers must face them. They tend to ignore two things. First, they ignore many decision options which other decision makers would regard as important. Second, in considering this limited set of options, they fail to consider many pros and cons that these other decision makers cannot ignore.[41]

During the next phase of analysis, judgments must be made concerning the tools and methods to be used in data production, manipulation, or interpretation. Pitfalls are typically associated with using tools whose limitations are not understood, using tools for their own sake, and using techniques and models of unknown reliability. Such inappropriate uses are most likely to occur when tools are employed because of their supposedly intrinsic merits rather than because they fit the problem. Pitfalls in the choice of tools and methods are often encountered in social experimentation, econometric analysis, benefit–cost analysis, and simulations.

- Social experiments have been particularly vulnerable to problems that vitiate their use.[42] In their report on the experiments in educational performance contracting, for example, Edward M. Gramlich and Patricia P. Koshel conclude that

> . . . the experimental strategy is no panacea. . . . The experiment encountered difficulties at almost every stage—there were problems in determining exactly what was to be tested; in design; in selecting firms, school districts, and schools; in writing and negotiating contracts; in developing the institutional programs; in enlisting the cooperation of local personnel; and in measuring the success of the contractors. Any one of the problems, if even slightly aggravated, could have completely frustrated the enterprise. Thus the first important lesson from the experience is that social scientists should be modest about their ability to perform experiments, at least in the area of education. Experiments can be done, but they are much more difficult than the early rhetoric implied.[43]

With regard to the planned variation experiments in education, James Coleman has noted:

> The experimental paradigm adopted at the federal level appears to be wrong, or at least incomplete, for it assumes the existence of a single decision maker who requires some kind of information feedback. In schools, above all, there are a number of decision makers.

He goes on to point out that:

> These decision makers are not all concerned solely about cognitive achievement. A number of outcomes are often more important to local school systems than grade-equivalent scores in reading achievement, and a number of outcomes are often more important to parents and children.[44]

In assessing the same experiments, Carol VanDeusen Lukas noted that successful experimental treatments might not be replicated in other locations, the presumed goal of the experiment, because the models are not specified in behavioral terms.

> Even if all models were complete in the sense of being firmly based on clear and integrated theories on how to teach children, some of the sponsors [i.e., the designers of

the models] still would not be able to provide detailed operational description of day-to-day classroom activity because their philosophy runs counter to close prescription. Instead they set out general principles and encourage teachers to carry them out in the manner that best fits their style and the needs of their students. . . . [W]here lack of specification is inherent in the nature of the model, variation in implementation will continue to occur.[45]

Alice M. Rivlin and P. Michael Timpane summarize the experience with planned variation experiments in education by noting that

The model developers held diverse views about the aims of early education. Hence the objective of testing models individually would have implied the development of many sets of outcome measures, each appropriate to the goals of a particular model or group of models. A comparison of models would have required that some common measures be applied, at moments in time when the full program effects would be captured. The most useful strategy might have been to do both, developing measures appropriate to specific groups of models, but finding out how each model stacked up on the various criteria. It would then have been feasible not only to rate models on their own objectives, but to say how they rated in terms of the objectives of other models. It might then have been possible to give potential users and other interested parties the information that one model rated higher on a particular set of outcomes and low on another set while the reverse was true of another model, and let them apply their own values in weighing the various measures.[46]

- A definitive analysis of pitfalls in choosing statistical techniques was generated by a symposium on the "Integration of Sociology and Other Social Sciences through Policy Analysis."[47] The principal paper was written by sociologist James S. Coleman, and discussion centered on the statistical techniques used in Coleman's classic study, *Equality of Educational Opportunity*. The study examined the question, "What factors in schools as they are currently organized in the United States contribute most to a child's educational experience?" As Coleman points out:

It was the answer to this analytical question . . . which could provide the most extensive avenues for policy, for this would indicate what kinds of policy changes in schools might be most effective in increasing educational opportunity.

In his paper, Coleman outlines three "general styles of strategies" in the use of regression analysis. These strategies are

- Use of regression analysis as a technique for estimating parameters of a well-specified model, with a known or hypothesized structure, but with unknown values or parameters
- Use of regression analysis as a technique for uncovering the causal structure in a set of variables when something is known about, or some prior assumption can be made about, the causal priority in a system of variables
- The use of regression analysis as a method for obtaining as much information as possible about the structure of relations among variables when

nothing is known or assumed about the independent variables other than that any relation between them and the dependent variable is not produced by an effect from the dependent variable to the independent variables

"It is clear," Coleman points out, "that belief in the use of one or another measure has been dictated by the general strategy in the use of regression analysis by the different disciplines." Coleman further points out that the fact that his report was shaped by sociologists had a major effect on the report's policy impact. The sociologists chose a two-level regression technique with individual student performance as the dependent variable and with the background characteristics of individual students as one set of independent variables and other school level variables as another set of independent variables. Coleman's results, which identified the importance of school level variables to education outcomes and gave support to efforts toward racial integration, attracted the greatest policy interest. Investigation of the effects of such variables, he says, is peculiar to sociology, since it stems from a concern for the influence of other persons or influences in the environment. Economists would ignore such influences, because they cannot easily be incorporated into a production function. Educational psychologists and statisticians would tend not to look for them because educators did not view them as educational inputs and, besides, they are beyond their control.

Though an academic historian, Andrew Hacker provided what could well be a decision maker's response to Coleman's paper and his report:

While the report purports to provide illumination for policy makers on how they may more effectively deploy their resources, the very manner in which that research was conducted induces the suspicion that it was also made to serve as an exercise in sociological scholarship and a testing ground for certain disciplinary techniques. Indeed, the criticisms of that report ... dwell less on public questions of effective education and more on the private dialectics of social science.

• Problems in the use of benefit–cost analysis have been widely discussed. They have been well summarized by the economist Robert H. Haveman, who was drawing in part on his experience as a member of the staff of the Joint Economic Committee of the U.S. Congress:

... while policy analysis focuses on the costs and benefits of policy measures to society as a whole, few individual policymakers find the net social benefit criterion relevant to their more limited objectives. If the legislator takes the costs of a decision and spreads them widely over citizens or in some other way camouflages them, while he concentrates the benefits on a vested group whose support he seeks, the fact that aggregate costs are greatly in excess of benefits is of little interest to him. In effect, policy analysis answers questions that few legislative policymakers are interested in either asking or having asked.[48]

More generally, Burton A. Weisbrod has argued that the typical aversion of economists to questions of distribution in program analysis is one important

reason why decision makers so often seem to ignore the analysis. He traces this aversion to the logical distinction economists like to draw between efficiency and equity. Choices concerning the latter are essentially political, in this view, and can be implemented with appropriate tax-transfer policies. Weisbrod holds that this orthodoxy assumes that tax-transfer programs are costless to effect, which of course they are not.[49]

Another example of problems arising in the economic analysis of benefits and cost is contained in a 1963 exchange between economist Jerome Rothenberg and Frederick O'R. Hayes, then an official at the U.S. Office of Economic Opportunity. With respect to Rothenberg's study of the benefits and costs of urban renewal programs, Hayes says:

> . . . public investment can seldom be justified solely on the basis of market consideration. Some part of the support must be found in other benefits. In renewal, a major part of support rests on an assumption of net social benefits, some measureable and others not, some clearly identifiable and others representing the satisfaction of a kind of itch on our collective superegos. . . . The most careful effort to analyze the costs and benefit of renewal is, therefore, unlikely to produce a definite answer.[50]

Hayes goes on to criticize the Rothenberg study for more specific reasons: concentration on too narrow a range of renewal projects, failure to consider benefits from income redistribution associated with projects, the assumption that resources are fully employed, neglect of the effects of projects on community functioning, and failure to produce data on benefits and costs meaningful to central city decision makers.

- An example of a pitfall in the choice of simulation modeling techniques is provided by the effort to model the San Francisco housing market mentioned earlier. Members of the model-building team soon came into conflict over

> The different levels of detail desired by the different participants to support their normal pattern of work and interest. The city planners, for example, maintained insistence on a level of detail which bordered on impracticality, in terms of input data requirements and output analysis. . . . The operations research personnel wished to retain simplicity in the initial model, even at the cost of possible loss of validity and utility of results. Experience with the use of simulation in other fields had indicated a consistent pattern. Initial models were found to be over-detailed, and successive versions were almost invariably made more aggregated. However, this process would presumably result in a model which could not answer all questions of interest to the planners. . . . The compromise finally reached probably satisfied neither party.[51]

Following their extensive review of the use of models in policy making, Martin Greenberger and coworkers concluded, first, that the political setting in which a model is presented may be decisive in how the model is received by policy makers, and, second, that the organizational framework in which a model is developed can also be a crucial determinant in its usefulness.

> We find that there is often a considerable gulf between the policymaker and the policy modeler, a reflection, we believe, of the different constraints under which the respective

parties work and the dissimilar value systems to which they respond. We find that modelers, especially those who are at a distance from the actual workings of the policy process, tend to overestimate the real relevance and usefulness of models for practical decision making.[52]

EVIDENCE AND ARGUMENT

Arguments linking data and information with conclusions will typically be a complex blend of factual statements and subjective evaluations. ". . . [A] wrong assessment of the strength and fit of evidence before it is included in the argument can lead to pitfalls in the drawing of conclusions."[53] Pitfalls include overly formalized styles of presentation, the use of data in contexts other than ones for which they were obtained, and setting inappropriate standards for acceptance of information or opinion. "No amount of technical skill can compensate for a lack of sophistication in the structuring of arguments or for carelessness in drawing the necessary distinctions between data, information, supporting evidence, and conclusions."[54] A variety of examples illustrate pitfalls of this type.

- Wallace K. Swan reports on efforts by the Minnesota Department of Public Welfare to undertake a series of studies leading to the preparation of a comprehensive master plan for shifting from reliance on larger multipurpose state hospitals to smaller community-based facilities for the mentally ill, the mentally retarded, and those dependent on drugs. Based on an analysis of this experience, Swan calls attention to a common pitfall:

> Care must be taken in a research analysis to indicate the value judgments which underlie . . . analyses. The assumption that community-based care is better than hospital-based care is closer to a value judgment than an established fact.[55]

In part because value judgments were presented as facts, the results of the study effort encountered a swarm of difficulties with the state legislature.

- An important pitfall was encountered during analyses of whether catalytic converters on automobiles produced sulfuric acid emissions that would create a hazard to human health and whether a delay by the U.S. Environmental Protection Agency (EPA) in implementing the statutory standards for HC, CO, and NO_x specified in the Clean Air Act would be justified.[56]

Human exposure to sulfuric acid concentrations in the air cannot be measured directly. One group of EPA analysts used a CO dispersion/physical-activity model to approximate peak hourly exposure of pedestrians near major arterial roads for both normal and adverse meteorological conditions. Based on calculations using this model, the analysts concluded that the "health risks [of using catalytic converters] exceed the benefits after four model years are equipped with catalysts." Based in part on such conclusions, the EPA

Administrator made a controversial decision to reduce the pressure on the automobile companies to install catalytic converters.

Subsequent analyses showed that the conclusion rested on a series of "worst case" assumptions, the improbability of which called into question the earlier conclusions. Had the uncertainties in the data, the assumptions, and the calculations used in the model been made explicit and appropriate sensitivity analyses conducted, a very different picture of the problem would have emerged. Later developments in estimating sulfuric acid exposures were to show that the problem had, in fact, been greatly exaggerated.

• Additional examples are found in the relationships between analysts and the U.S. Congress on issues of higher education. Thomas R. Wolanin points out that the criteria used in Congress to determine what is relevant information and the criteria applied by those who supply information are different.

> [L]egislative policy makers and policy analysts in the executive branch and outside of government reach decisions on desirable courses of action in basically different ways. Analysts quite properly begin with questions and finally produce the answers. The legislators, as politicians, are concerned above all with the answers, the results, the impact, who will get what—"the bottom line." To them, analysis is often perceived as an analytic trap. Implicit in most questions are values and preferences. These are often unintended and unknown even to the analyst. This is particularly true as the questions and analytic techniques become more abstract and are cast in terms of statistical procedures, models and simulations. A legislator who buys into the apparently innocent and objective questions runs the risk of being locked into conclusions that are politically unacceptable. For example, the 1973 report of the National Commission on the Financing of Postsecondary Education has had little apparent impact on the 1974–76 review of the basic higher education legislation. An important reason for this is that the Commission was charged with developing "alternative models for the long range solution to the problems of financing postsecondary education" and it carried out its task in large part by producing "an analytic framework."[57]

• Thomas R. Dye compares the way the now-defunct U.S. Office of Economic Opportunity hurried to report "preliminary findings" of the New Jersey income maintenance experiment, which supported the agency's views that guaranteed levels under a negative income tax did not produce massive reductions in work effort, with the same agency's reluctance to release preliminary results of the Westinghouse Learning Corporation/Ohio University evaluation of the Head Start early child education program, which did not support the agency's views. Dye asks:

> Are government sponsored research projects predisposed to produce results supportive of popular reform proposals? Are social scientists, whose personal political values are generally liberal and reformist, inclined to produce findings in support of liberal reform measures?[58]

• A detailed assessment of the New York City RAND Corporation's experiences in providing analytical advice to the city's agencies illustrates the

pitfalls of using models as evidence. A RAND analyst had created a sophisticated queueing model to study the problem of empty beds in New York City's hospitals. Analysts in the City's Health Services Administration were notably cool to the development and proposed use of this model. In their eyes, the model "did not add much to what could be gained by simply ordering hospital officials to close down empty beds 'to the point where occupancy rates reached 90 percent.'" "I don't have much faith in computer simulation," explained the official who was director of program analysis at the time.

> I think that you can use a lot less sophisticated quantitative techniques a lot more successfully. Frankly, I don't think it's worth the effort—the time, and especially the expense—to do a computer simulation when you can do it a lot less sophisticatedly. ... There was a problem of taking a study that sophisticated and turning it into a policy. .... I felt, for our purposes at least, that things were happening so quickly ... that we just couldn't rely on that kind of sophisticated method.[59]

CONCLUSIONS, COMMUNICATION, IMPLEMENTATION

The way conclusions are drawn and the persuasiveness with which they are communicated to decision makers can determine the fate of an entire systems analysis study. Meltsner puts the proposition cogently:

> Analysis is a social process, and this fact is nowhere more apparent as when we try to communicate the results of our work. .... Communication and analysis are interdependent activities. Not to recognize this interdependence and its social implications are primary errors.[60]

The test of an analysis is whether it is convincing "even to a hostile and disbelieving but intelligent audience."[61] A related pitfall is failure to consider the problems of implementation adequately, which Majone attributes to the excessive reliance of many analysts on normative decision models and the pure logic of choice, on analysis rather than on synthesis and design.[62]

- An example is furnished by the higher education situation discussed above.

> From their respective institutions, legislators, executive branch officials and postsecondary educators often fail to communicate effectively with each other because each speaks the language of their specialized environment. Congressmen work and communicate in legislative language and explain their efforts in everyday language ... the executive branch officials speak in the convoluted style born of bureaucratic caution and complexity.... The educators tend to speak in the eloquent rhetoric of the academy. They leave the legislators with a warm glow of high ideals, but frequently without the words of legislative art that could transform noble sentiments into legislative realities. Alternatively, academics and bureaucrats speak in the tongues of education and social science research, lacing their remarks with "multiple regressions," "cognitive constructs," "learning modules," and "synergy at the interfaces." ... Unfortunately, there are too few effective translators when they come together, so legislators fail to use the information they receive because they often cannot understand it.[63]

• In discussing a set of papers analyzing results of the New Jersey income maintenance experiment, Henry Aaron said:

> . . . the various reports on the experiment thus far available may be scientifically useful . . . but they are of limited short-run political value. [They] are not fully comprehensible to any but well trained economists. Even for this latter group, the various reports suffer from uneven style, employ diverse models that cannot be compared, and in general reflect a kind of scholarly laissez faire that makes integration and evaluation of the findings quite difficult. . . .[64]

• A brilliant example of failure to consider implementation problems is provided by Davis and Salasin.[65] Research at the National Institute of Mental Health showed that the rehospitalization of mental patients following discharge could be reduced if a social worker from a patient's county of residence visited the hospital, established a working relationship with the patient, and participated in planning for discharge. They proposed to a state Commissioner of Public Welfare that he implement the practice statewide.

Instead of leaping for joy at their proposal, Davis and Salasin relate, he said he wanted to ask a few questions:

> "How will people in our hospital social service departments feel about county workers coming in and taking over a major portion of what they have seen as their roles? And how will the counties feel about extending the duties of their employees beyond the responsibilities which they normally carry?"
>
> "In your project, you used experienced psychiatric social workers and public health nurses with masters degrees. How do I know that our county welfare workers will be able to match the skills of your project workers? How will we pay for the training programs necessary to prepare the county workers to carry out the same aftercare services? Where do I find funds to pay for their travel expenses to the state hospitals? Who will carry out the work that they will be unable to accomplish while they are spending the required time at the hospital helping the patients prepare for discharges?"
>
> "Your findings sound almost too simple and pat. How do you know the results you obtained did not stem from the skills of the workers on the project? And how do you know that their small caseload of only about six patients at any time wasn't the determining factor?"
>
> "In your project, your workers had their offices in the city very close to the hospital and the locations of the patients after discharge were also fairly close. There aren't many parts of the state where things are that convenient. Will the plan still be feasible?"
>
> "The counties aren't going to volunteer to use their scarce resources for added service unless legislative and budget adjustments are made. The state legislature just met. How will I bridge things throughout the rest of the biennium?"
>
> "Readmission rates are already respectably low. Who is so critically concerned about the problem that the increased expenditures would be warranted? Of course, I'd like to see readmission made unnecessary even for one patient, but it would help if the legislature, the governor, or at least some groups were concerned enough to back this policy."
>
> "The need for social workers in our hospital would be considerably less. Some may lose their jobs. How will I handle their unhappiness? And the county workers are going to be raising Cain because of the hardships they'll face, even if compensation is

arranged; they will have to stay away from their families during trips to the hospital, for instance."

"Though your results are statistically significant, would the improvement in the readmission rate be sufficient even to be noticed? Will anyone feel better for having gone along with this policy if it should be adopted?"

• A vivid example of planners and analysts stumbling into an implementation pitfall is provided by Lehman's study of the Canadian experience with welfare reform. Lehman reports as follows:

The strategy charted by federal officials in the Social Security Review was a deductive one. T. R. Robinson, an Assistant Deputy Minister who mapped federal strategy in the Review, described the method in this way:

"The idea was to rethink the basic concepts, examine the evolving nature of social security issues, and work from basic principles—i.e., 'from the top down'—in pursuit of viable reform proposals. Constraints resulting from potential costs, administrative exigencies, existing programme patterns, and finally any difficulties with respect to financial arrangements or jurisdictional issues, were to emerge as the issues (and choices) became more and more specific. They were not to encumber the first stages of the review. The evolution was to be from the 'desirable' to the 'practical,' with as little departure from the former as possible" (Robinson, 1975, p. 20).

Of course, the Review did not proceed smoothly along these lines. The issues of cost-sharing and jurisdiction derailed attempts to reach a consensus. The deductive approach had ignored crucial questions of practicality that desperately needed resolution. Because no effort had been made to resolve these difficult questions in the early stages of the Social Security Review, no room for compromise existed when they finally reached the agenda. Thus just when conflicts over jurisdiction and financing had become most intense in the April, 1975 meeting of welfare ministers, the federal communique emerging from this meeting dreamily declared: "A consensus has been reached on the broad features of a new guaranteed income system," and officials have been "instructed to work out the details of an operational design for an income support and supplementation (guaranteed income) system of the kind Ministers [have] agreed on" (quoted in Johnson, 1975, p. 42).

In view of the disappointing results of the Social Security Review, some federal officials closely involved with it later themselves regretted not having taken an incremental or inductive approach that tried to accommodate these practical problems piecemeal. The disappointing outcome is testimony enough to the deficiencies of the deductive approach.[66]

IS ANALYSIS EVER DONE RIGHT?

The preceding catalogue might make the prospect of pitfall-free analysis seem depressingly remote.

Unfortunately, successes are even harder to document than pitfalls. This is not to say that analysis and analysts have not been influential in policy making. Enthoven and Smith have written, "Good systems analysts were willing to work for [U.S. Secretary of Defense Robert S.] McNamara and [McNamara's successor, Clark] Clifford because they knew that when they did good work it

would be acted upon and would influence decisions."[67] Many of the growing number of current and former members of policy analysis offices in government and of policy-related research and analysis institutions outside of government have testified in various forums that their work was used. Many former policy makers have lauded their analytic staffs. What is hard to come by are examples showing how attention to the four aspects of craftsmanship discussed above contributes directly to the acceptance and use of an analytic product.

Illustrations of analytic craftsmanship that come close to filling the need are described in two teaching cases developed at the John F. Kennedy School of Government at Harvard University.

The first case, *The 1972 Education Finance Initiative*,[68] describes how analysts in the U.S. Department of Health, Education, and Welfare's Planning and Evaluation Office participated in departmental and presidential decision making concerning an initiative, proposed by the President's staff, to reform the financing of elementary and secondary education. The issue had become prominent because of state court decisions declaring the property tax to be an inherently unfair basis for financing education. At the direction of HEW Secretary Elliot L. Richardson, and in association with officials from the U.S. Office of Education and the White House Domestic Council, analysts over the course of a year assessed the nature of the school finance problem, possible goals of a federal government initiative, and the costs and likely consequences of alternative ways of meeting different goals. Drawing on their work, Richardson recommended to the President a program to replace the property tax as a source of school funds by federal revenues, with the primary educational goal of equalizing interstate differences in spending per pupil. (At the time, Alabama was spending an average of $438 per pupil; New York's per pupil spending averaged $1,237.)

The second case, *Caspar Weinberger and Welfare Reform*,[69] describes how a different group of HEW analysts assisted HEW Secretary Caspar W. Weinberger in reviewing U.S. welfare programs and deciding what policies to recommend to the President. The analysts conducted a comprehensive review of the possible objectives of an income maintenance system, the costs and consequences of existing welfare programs, and alternatives for reform. Lengthy discussions of the analysis with Weinberger led to his recommending to the President in November 1973 that he propose to Congress a comprehensive overhaul of U.S. welfare programs and the enactment of a negative income tax.

Both cases illustrate the importance of craftsmanship to the success of analysis.

1. The program and the issues were defined both with analytic rigor and with awareness of the political and institutional context. Moreover, the analysts became involved in a sustained dialogue with the decision makers concerning the nature of the problem, the issues, and the interrelationship between policy

goals and the means to achieve them. In the end, fundamental agreement had been reached on how to think about the subject, and the issues and alternatives had been sharply drawn.

The lengthy discussions of the problem appeared to clarify and focus both Secretaries' thinking.

- School finance analysts, for example, posed questions such as "What is the nature of [the property tax relief] goal? Is it to reduce property tax burdens or to provide net tax relief? Or is it to make the property tax or the general tax structure more equitable and progressive? For whom is the relief primarily intended: homeowners, renters, agriculture, commercial property? The answers to these questions can dramatically affect the form of our program."[68] Richardson came to emphasize the goal of interstate equalization of per pupil spending for the new initiative after detailed analysis of this sort. Only the federal government could achieve this goal, he concluded, whereas other goals—intrastate equalization, general fiscal relief for the states, encouraging educational reform—were less appropriate or could be pursued in other, perhaps more efficient ways.
- Weinberger began with an apparent preference for tougher welfare management and skepticism concerning welfare reform schemes. The analysts surveyed possible objectives of welfare programs and the effectiveness of the existing system in achieving them. In recommending comprehensive welfare reform, following extensive review of the analysis, Weinberger concluded: "The interrelationships [among] the domestic assistance programs have become as important as the content of each one. To ignore these interrelationships is to pile one program on top of another with undesirable, unintended and costly results."[69]

2. A rather complex array of data had to be assembled and analyzed. (In both instances, computer simulations were needed to analyze the consequences of different policy designs.) By maintaining their focus on the problem and the policy issues, the analysts were able to create a coherent, appropriate, and credible body of evidence without being trapped or mesmerized by technique.

- The school finance analysis, for example, encompassed both straightfoward examinations of school spending trends (comparisons of education spending increases with changes in other local expenditures and with changes in pupil–teacher ratios) and computer models simulating the effects of different federal spending formulas on patterns of per pupil spending by individual school districts. Conclusions were drawn in terms of the issues to be resolved:

> [Power equalizing grants, i.e., federal grants that are larger for less wealthy states, provide] a strong interstate equalizing thrust and a strong upward push on educational spending. As such [they] ... conflict particularly with an objective of net tax relief, since they impel the States toward higher levels of effort.[68]

- William A. Morrill, who supervised Weinberger's welfare reform analysis, describes the process of structuring the analysis:

We just worked it through mutually as to what the problem was and what we could bring to Cap [Weinberger] in order to make it a comprehensive problem. We began to divide it into a group of sensible bites—the front end of which was largely descriptive: what's there? What does the system now look like in programmatic content? And then go on to say: "Okay, what can we say out of all of the research that's been done? What are the characteristics of this population? How do they behave? What do people allege and how much research do we have to underpin any particular view of an issue of either a purely analytic or empirical nature?"

In other words, the analysis was carried out with the informational needs of a definite audience kept clearly in mind.

3. The analysts were generally careful to distinguish between facts, evidence derived from various analyses, and conjectures and other subjective assessments, though these elements were often intertwined in their communications.

- In assessing the nature of the financial squeeze on schools, school finance analysts noted, for example, that

In general, the meaning one should imput to the "taxpayer's revolt" is not clear. We may simply be facing a change in consumer preference (based on widespread doubts of educational effectiveness) or a scapegoating (where educational finance issues are the only opportunity for voters to express directly a more general taxpayer's discontent).

At another point, the analysts said that they "continue to have doubts about the merits of such [an initiative]. . . . We have little confidence that increased expenditures in the range we envision will produce additional achievement for the average child." "You may anticipate," they said, "that [the Office of Education] will have a somewhat stronger view as to the fiscal predicament of our schools and will be more inclined [than we are] to promote additional net expenditures for education. . . ."

- Welfare reform analysts needed to address the question of whether various welfare reform alternatives would discourage work. Evidence from experiments conducted in New Jersey suggested that there would be little work withdrawal, but many doubted the validity of these results. Faced with this situation, the analysts said:

What we argued was that while the experimental results were not conclusive, they were consistent with [other] results. In other words, a totally different methodology yielded similar results. Viewed in isolation the New Jersey results were not that influential, but viewed in the context of a whole range of evidence, they're one more factor that seems to indicate that large-scale withdrawal from the labor force or work reduction is not going to result.

4. The analysts employed skill and discernment in carrying out what Meltsner has termed their social role in the bureaucracy. It would be too much to claim that they had convinced even hostile audiences—those who did not share the analyst's value orientations remained unconvinced—but respect for the communication process and sensitivity to organizational needs were evident. "During the past two weeks [Office of Education (OE)] and [Office of the

Secretary (OS)] staff have worked extensively with Domestic Council staff to prepare an education finance initiative for the consideration of ... the President," the education finance analysts reported to Richardson at one point.

It is not certain what, if any, major decisions you should be asked to make. What appeared to be major conflicts and problems a few days ago are, slowly but surely, submitting to analysis. There are, at this point, no clearcut differences among OE, OS and Domestic Council staff, that require your intervention.

Welfare reform analysts were especially sensitive to Weinberger's instincts, which tended to be more conservative than theirs.

Our conscious effort was not to push a viewpoint, but to present as broadly and objectively as possible the background information that the Secretary needed. ... There were certain kinds of questions that he was going to ask: obviously you tried to anticipate those questions so that you were doing good staff work for him and giving him what he wanted.

In the end, because of the analysts' efforts, disagreements centered on questions of values and goals much more than on the nature of the issues, alternatives, and evidence.

In both cases, considerable, if not decisive, credit must be given to the intelligence, integrity, patience, and administrative skills of the users of the analysis. They achieved a high degree of rapport with their analysts but did not coopt them. The craftsman benefits greatly from a discerning and forbearing patron; indeed, the discerning craftsman will seek out just such a patron.

CONCLUSIONS

The central conclusion of this survey of pitfalls is that systems analysis often does not matter in decision making because systems analysts do not know what matters to decision makers. The usefulness of analysis will increase to the extent that analysts develop better understanding of decision-making contexts and conduct their work in the light of that understanding.

But in following this advice analysts encounter yet another pitfall. Gaining an understanding of the decision maker's problem may shade into learning to think exactly like the decision maker, into adopting uncritically the decision maker's point of view and sharing his or her biases and blind spots. This would be a mistake on two grounds. First, it would be fundamentally irresponsible. The systems analyst, no less than any other craftsman, must maintain integrity. One can argue, without reintroducing the myths of scientific detachment and objectivity, that integrity for the analyst consists of what Rein has termed "temporary detachment" from one's own values and from the values of one's superiors or patrons while examining the policy problem critically.[70] Second,

decision makers are not interested in analysis solely because it can solve problems. Carol Weiss concludes on the basis of her study of the uses of research by government decision makers:

> Evidence suggests that government officials use research less to arrive at solutions than to orient themselves to problems. They use research to help them in thinking about the issues and defining the problematics of the situation, to gain new ideas and new perspectives. They use research to help *formulate* problems and set the agenda for future policy actions.[71]

Officials, she is saying, use research and analysis for enlightenment as well as for getting solutions to specific problems.

It must also be recognized that decision makers frequently make mistakes in their use of systems analysis. The responsibility for avoiding pitfalls in the use of systems analysis does not rest solely with the analysts. Decision makers share that responsibility. They should monitor the performance of their advisory systems to eliminate what George[72] refers to as "malfunctions." They should avoid judging studies solely on the basis of the conclusions that are reached, second- or third-hand opinions that may be unreliable, the credentials of the analysts, or the elegance of the models that are employed.

The most important questions systems analysts face are: Toward what purpose am I working? Whom am I trying to influence and in what directions? What types of analytic products are needed? What approaches to the performance of these analytic tasks are most likely to be successful? These are matters more of judgment than of technique, and it is by the perceptive exercise of these types of judgments that the pitfalls in producing useful analysis can be avoided.

NOTES AND REFERENCES

1. I shall follow Steinbruner in defining complexity to mean the presence of two or more competing values, the existence of significant uncertainty, and the dispersal of power over outcomes among a number of actors and organizations.
   Steinbruner, J. D. 1974. The Cybernetic Theory of Decision: New Dimensions of Political Analysis. Princeton: Princeton University Press, p. 16.
2. Tribe, L. H. 1973. Policy science: analysis or ideology? Philosophy and Public Affairs 2(1): 66–110. Reprinted *In* Benefit–Cost and Policy Analysis, 1972. Chicago: Aldine, p. 43.
3. Allison, G. T. 1971. Essence of Decision. Boston: Little, Brown, p. 10.
   See also
   Lynn, L. E., Jr. 1978. The question of relevance. *In* Knowledge and Policy: The Uncertain Connection (L. E. Lynn, Jr., ed.). Washington, D.C.: National Academy of Sciences.
4. Rein, M., and S. H. White. 1977. Policy research: belief and doubt. Policy Analysis 3(2): 249.
5. Nelson, R. R. 1974. Intellectualizing about the moon–ghetto metaphor: a study of the current malaise of rational analysis of social problems. Policy Sciences 5: 381.
6. Quade, E. S. 1968. Introduction. *In* Systems Analysis and Policy Planning (E. S. Quade and W. I. Boucher, ed.). New York: American Elsevier, p. 11.

7. Quade, E. S., and W. I. Boucher. 1968. Systems Analysis and Policy Planning. New York: American Elsevier, p. 12.
8. Rein and White[4] (pp. 261–262) use the term "policy research" in ways that suggest that their definition subsumes what I call systems analysis as well as policy and social science research.
9. Rein and White[4] (p. 249).
10. Quade, E. S. 1975. Analysis for Public Decisions. New York: American Elsevier, p. 4.
11. Quade and Boucher[7] (p. 348); Quade[10] (p. 298).
12. Quade[10] (p. 300).
13. For a discussion of Lindblom's work in the context of analyzing the role of policy analysis in decision making, see
    Schultze, C. L. 1968. The Politics and Economics of Public Spending. Washington, D.C.: The Brookings Institution, Chapters 3–4.
14. Rein and White[4] (p. 262).
15. George, A. L. 1975. Introduction and overview. *In* The Use of Information. Appendix D to the Report of the Commission of the Government for the Conduct of Foreign Policy. Washington, D.C.: U.S. Government Printing Office.
16. George[15] (p. 10).
17. Allison[3] (pp. 67, 144).
18. In Steinbruner's view, the best-developed themes of collective behavior are based nonetheless on an analytic paradigm within which individual actors are assumed to act in accordance with rational calculations of self-interest. In contrast, he argues, actors behaving in accordance with a cognitive or cybernetic paradigm "will not display the same degree of deliberate accommodation, will act more independently, and will by-pass bargains which under analytic assumptions would appear to be obvious." See Steinbruner[1] (p. 147). I will not explore the implications of this view in this chapter.
19. Tribe[2] (p. 44).
20. Lindblom, C. E. 1968. The Policy-Making Process. Englewood Cliffs, New Jersey: Prentice-Hall, p. 21.
21. Enthoven, A. C. 1975. Ten practical principles for policy and program analysis, pp. 456–465. *In* Benefit–Cost and Policy Analysis, 1974. Chicago: Aldine.
22. Enthoven[21] (p. 45).
23. Enthoven[21] (p. 459).
24. Idem.
25. Enthoven[21] (p. 462).
26. Enthoven[21] (p. 464).
27. Schultze[13] (p. 103).
28. Rivlin, A. M. 1971. Systematic Thinking for Social Action. Washington, D.C.: The Brookings Institution, p. 59.
29. Miles, R. E., Jr. 1974. The Department of Health, Education and Welfare. New York: Praeger, p. 72.
30. Drew, E. 1967. HEW Grapples with PPBS. The Public Interest. No. 8 (Summer 1967), pp. 12–13.
31. Menges, C. 1974. Toward more effective social policy: lessons from contemporary evaluations of federal programs [processed, August 1974]. p. 5.
32. Majone, G. 1977. Pitfalls of analysis and the analysis of pitfalls. Journal of Urban Analysis 4: 251.
33. Majone[32] (p. 250).
34. Majone[32] (p. 235).
35. Majone[32] (p. 238).
36. Majone[32] (p. 239).

37. Majone[32] (p. 239).
38. Burke, V. J., and V. Burke. 1974. Nixon's Good Deed—Welfare Reform. New York: Columbia University Press.
39. John F. Kennedy School of Government. 1977. Caspar Weinberger and Welfare Reform. Case C95–77–199. Cambridge, Massachusetts: Harvard University, p. 20.
40. Kraemer, K. L. 1973. A Systems Approach to Decision Making, Policy Analysis in Local Government. Washington, D.C.: International City Management Association, pp. 71–72.
41. Kraemer[40] (p. 70).
42. Experimental results have been influential with decision makers despite serious pitfalls in their derivation. A match-up between a decision maker with a naive faith in scientific methods or a tendency to judge analysis by its findings, and analysts with similar failings can be especially dangerous to the public interest.
43. Gramlich, E. M., and P. P. Koshel. 1975. Educational Performance Contracting. Washington, D.C.: The Brookings Institution, p. 75.
44. Coleman, J. 1975. Comment by James Coleman, p. 173. *In* Planned Variation in Education (A. M. Rivlin and P. M. Timpane, ed.). Washington, D.C.: The Brookings Institution.
45. Lukas, C. VanD. 1975. Problems in implementing Head Start planned variation models, pp. 118–119. *In* Planned Variation in Education (A. M. Rivlin and P. M. Timpane, ed.). Washington, D.C.: The Brookings Institution.
46. Rivlin, A. M., and P. M. Timpane. Planned variation in education: as assessment, p. 13. *In* Planned Variation in Education. (A. M. Rivlin and P. M. Timpane, ed.) Washington, D.C.: The Brookings Institution.
47. Coleman's paper is entitled "Integration of Sociology and the Other Social Sciences Through Policy Analysis." Quotations from it and the discussion it generated are in
Charlesworth, J. C. (ed.). 1972. Integration of the Social Sciences Through Policy Analysis. Philadelphia: The American Academy of Policial and Social Science, pp. 162–195.
48. Haveman, R. H. 1976. Policy analysis and the Congress: an economist's view. Policy Analysis 2(2): 247.
49. Weisbrod's views are summarized by
Chase, S. B., Jr. 1968. Introduction and summary, pp. 20–21. *In* Problems in Public Expenditure Analysis (S. B. Chase, ed.). Washington, D.C.: The Brookings Institution.
50. Dorfman, R. (ed.). 1965. Measuring the Benefits of Government Investments. Washington, D.C.: The Brookings Institution, pp. 352 ff.
51. Kraemer[40] (p. 75).
52. Greenberger, M., M. A. Crenson, and B. L. Crissey. 1976. Models in the Policy Process. New York: Russell Sage Foundation, pp. xiv–xv and 20.
53. Majone[32] (p. 245).
54. Majone[32] (p. 248).
55. Swan, W. K. 1977. Research and the political process: experience with a plan for deinstitutionalization. Policy Studies 5(3): 258.
56. This account is based on information in the teaching case
John F. Kennedy School of Government. 1976. Automobile Emissions Control: The Sulfates Problem. Case C515–76–077, –078, and –079. Cambridge, Massachusetts: Harvard University.
57. Wolanin, T. R. 1976. Congress, information and policy making for postsecondary education: "Don't trouble me with the facts." Policy Studies 4(4): 390.
58. Dye, T. R. 1976. Policy Analysis. University, Alabama: University of Alabama Press, p. 13. Carol Weiss has observed that "... Evaluation researchers tend to be liberal, reformist, humanitarian, and advocates of the underdog ..." which, to the extent that it is true, cannot help but influence the structuring of arguments and conclusions. See
Weiss, C. H. 1973. Where politics and evaluation research meet. Evaluation 1(3): 44.

59. Greenberger *et al.* (p. 305).
60. Meltsner, A. J. 1976. Policy Analysists in the Bureaucracy. Berkeley: University of California Press.
61. Kahn, H., and I. Mann. Techniques of Systems Analysis. RM–1829. Santa Monica, California: The Rand Corporation. Cited by Majone.[32]
62. Idem.
63. Wolanin[57] (pp. 390–391).
64. Aaron, H. J. 1975. Cautionary notes on the experiment. *In* Work Incentives and Income Guarantees: The New Jersey Tax Incentive Experiment (J. A. Pechman and P. M. Timpane, ed.). Washington, D.C.: The Brookings Institution.
65. Davis, H. R., and S. Salasin. 1978. Strengthening the contribution of social R & D to policy making. *In* Knowledge and Policy: The Uncertain Connection (L. E. Lynn, Jr., ed.). Washington, D.C.: National Academy of Sciences.
66. Lehman, C. K. 1977. Welfare Reform and the Working Poor: The Setting of Poverty Policy in Canada and the United States. a Thesis presented to the Department of Government, Harvard University, pp. 192–193.
67. Enthoven, A. C., and K. Wayne Smith. 1971. How Much Is Enough? New York: Harper & Row, p. 323.
68. John F. Kennedy School of Government. 1977. The 1972 Education Finance Initiative. Case C95–77–192. Cambridge, Massachusetts: Harvard University.
69. John F. Kennedy School of Government.[39]
70. Rein, M. 1976. Social Science and Public Policy. Harmondsworth, Middlesex, England: Penguin, p. 169.
71. Weiss, C. H. 1976. Research for policy's sake. Speech given at the Symposium on Applied Sociology: Patterns and Problems, April 22, 1976. Cleveland, Ohio: Case Western Reserve University, p. 4.
72. George.[15]

8 Don't Slight Communication: Some Problems of Analytical Practice

Arnold J. Meltsner

Analysis is a social process. Normative accounts implying that analysis is some objective exercise uninfluenced by human feelings, values, and beliefs could not be more mistaken. Analysis is produced by people operating in social contexts. Analysts work together as colleagues to produce analysis. They consult one another and review one another's work. They interact with many people besides analysts to secure data and to understand different perceptions of a problem. They report to a client whose values and predispositions strongly affect the conduct and outcome of analysis. With that client's support, and sometimes without it, they disseminate their results to many people. That analysis is a social process is at no time more apparent than when we try to communicate the results of our work.

By communication, I mean the act of gaining acceptance and asserting a claim on another person's beliefs. Now this definition is inadequate, and I am sure it will not satisfy those who are used to the jargon of sources, messages, channels, and receivers. Its ambiguity allows me to stress a number of points, however. First, communication involves human interaction; in its simplest form communication between an analyst and his immediate client is the relationship of two people. Neither participant can completely control this relationship or determine its effectiveness. Second, my subject here is one side of this relationship—namely, the analyst and his means of gaining acceptance for himself and his work. Third, I interpret the results of analytical work to be "assertions" about reality or desirable changes in reality. These "assertions" may or may not be grounded in theory and evidence. In either case, the analyst makes a claim on his client or the person with whom he is trying to communicate. "A man who makes an assertion puts forward a claim," Toulmin informs us, "a claim on our attention and to our belief. Unlike one who speaks frivolously, . . . a man who asserts something intends his statement to be taken seriously . . ." (Toulmin, 1964, p. 19).

In making my own assertions about communication, I have felt the lack of empirical literature dealing with the communication of *analytical* information. Ideally, I would have liked to discuss a range of behaviors associated with communication. This range could be determined by the kind of analyst (e.g., policy analyst, systems analyst, planner), the analyst's role in a specific organizational context (e.g., a research technician working as a contractor for a large governmental agency), the type of client (e.g., a business executive with little patience), and the choice of different means of communication (e.g., memoranda, briefings, formal reports, movies). Instead, I have fallen back on my own experience in working with governmental agencies and have sought to identify common errors in communication that cut across the wide range of possible analytical behavior.

Generally, communication has two highly interrelated parts: social and substantive. One way to understand the distinction between them is to think of the skills of the analyst. Writing and speaking are social, while modeling and analyzing data are substantive; formulating and defining a problem require both substantive and social skills. The social part, the context in which communication takes place, encompasses such factors as the persuasiveness and credibility of the analyst, the feelings of trust between analyst and client, and the extent to which values and beliefs are shared by the analyst and his various audiences. The substantive part has to do with the logic and content of the study itself, with the validity and reliability of the knowledge it contains, and with its policy and organizational implications. When an adequate study is believed and accepted or when an inadequate study is not believed and is rejected, the substantive and social parts of communication usually are compatible—and the acceptance or rejection can be validated for the analyst by his peers. When a problem of communication exists, the two parts usually are incompatible. This may mean, for example, that the analysis contains an adequate policy direction for the given context, but fails to convince or gain the acceptance of the client. Or it may mean—a problem analysts sometimes ignore—that the analysis contains misleading information, which the client accepts.

In the situation of easy rejection (see the table below) the analysis fits the imperatives of the organizational context. It follows the canons of current analytical practice; it is adequate in the sense that most analysts would have done it in the same way in the same environment. Yet the client and other organizational decision makers are not willing to accept either the study's assessment of a problem or its recommendations. In the situation of easy acceptance the study does not follow the standard practice; it is inadequate in the sense that other analysts would have done it in a different way in the same environment. If they were asked to judge the study, they might call the analysis "sloppy" or "incomplete." Neither the analyst nor the client may be aware that the study's information is misleading and inappropriate for the organization.

The client and other organizational decision makers accept the implications of the study and act on them. For the client, the analysis as presented does fit the imperatives of the organizational context. Thus communication problems can be depicted simply, as in the following table:

		Social	
		Credible	Noncredible
Substantive	Adequate	No communication problem	Easy rejection
	Inadequate	Easy acceptance	No communication problem

Accepting inadequate information is a serious problem for the client. Of course, many clients are too intelligent and perspicacious to be easily deceived, but some can be. Other than developing the skills to evaluate analytical products, assessing analysts on the basis of their past records, and resorting to multiple and independent sources of information, a client can do little to protect himself against a social context that encourages his tendency to accept what he is being told. The analyst also has something to lose in this situation. In his desire to promote his study, he may be trading off short-term acceptance for long-term ridicule by his peers and abuse by his client if the recommendations of the study do not have the intended effect. In short, the social part of communication may have an empty victory over the substantive part.

A more familiar problem—and my chief concern here—is the domination of the social by the substantive: high-quality studies are rejected because the social part of communication is slighted. A number of years ago Kahn and Mann (1957, p. 45) identified the pitfall of "hermitism": "The problems of communication and persuasion are often ignored though they are central to getting recommendations translated into policy." According to them, the analyst must have the ability "to sell" his study and must spend a great deal of time doing so. Usually, some member of an analytical team writes well, talks well, or has the personal attributes to secure the trust of a number of clients and decision makers. The ability is there, but what is often lacking is an appreciation of the extensive time and effort that must be devoted to communication if it is to be persuasive.

Besides underestimating the resource requirements for communicating—money, as well as time and effort—analysts frequently do not understand when they should spend these resources. All too often, only when a lengthy project nears completion does the analytical staff belatedly turn its attention to

documenting results and disseminating recommendations. Considerable resources are expended on publications, illustrations, charts, slides, and other paraphernalia of professional systems and policy analysis. Yet much of this effort may be wasted because such communication is an afterthought. In order to be effective, communication must be an integral part of the analytical activity itself. As a relationship, communication is a two-way process: the participants must have the opportunity to listen to each other. Therefore, to enhance the receptivity of an analysis and also to obtain information and other assistance, the analyst should involve the client and other likely members of the study's audience in the project from the beginning. Determining who will be members of the study's audience and sizing up their perceptions of the problem are communication strategies central to the initial steps of analysis.

Unfortunately, far too many analysts perceive communication as something outside analysis. Even so experienced a practitioner as Edward Quade makes this mistake when he identifies some failures of communicating with a client as "external pitfalls" or "factors or actions that handicap the analysis but do not originate within the analysis itself" (Quade, 1975, p. 312). Some handicaps do result from the analysis itself; for example, a tight and logically structured analysis is easier to communicate than one that is loose and fuzzy. And a model that represents reality with simplicity but without doing violence to that reality will be easier for a client to understand and accept. For some purposes it is useful to make inside–outside distinctions, but in the case of communication, to do so is dysfunctional to our understanding of the analytical process. Communication and analysis are interdependent activities. Not to recognize this interdependence and its social implications is a primary error. Difficulties in communication are inherent in situations where there is conflict over values, objectives, and beliefs. Although a certain amount of rejection of quality work is to be expected, the difficulties in communication should become less severe as the participants in analysis begin to appreciate the social context of their endeavors.

Having pointed out the importance of understanding communication as part of the social process of analysis, I will now discuss certain errors involved in the conception and choice of *audience* (recipients), *argument* (content and form), and *means* (techniques and media) of communicating analysis.

AUDIENCE

Although the choice of whom to communicate with is a central decision, it is one that analysts too often leave to the client. They may not be given a choice, but in order to improve the effectiveness of communication, analysts should not let this decision be made by default. Once they commit themselves to getting involved in the making of this decision, analysts will see that what is needed is a

strategy for selecting and sequencing audiences. Such a strategy is not a simple matter to devise. It encompasses an appreciation of the organizational context of the analysis and of the fact that neither the client nor the analyst is the sole audience. The audience for analysis is seldom a single person.

IGNORING THE ORGANIZATIONAL CONTEXT

Analysis is conducted in an organizational context. All too often analysts sit "outside" the organization and have little exposure to the organizational imperatives that will influence their communications. To be outside means that the analysts are located so as to be *distant* from critical information and cues. Contractors and consultants often have this problem, but so do employees. For example, analysts who work, as employees, for the head of a large complex organization may be quite naive about the rest of the organization. They may not understand that the rest of the organization not only is a source of information but also has to be included, in some way, in the calculations about the audience for analytical communication. To convince the client is hardly sufficient when acceptance and implementation rest on these other members of the organization. What are some of the organizational factors that influence the conduct of analysis and its communication?

Time

One important factor is time. Through its routines and rules the organization sets the schedule for the analysis. Such rules not only place constraints on performance but also establish when it will be best to communicate. For example, an analysis that is intended to influence a decision obviously should be presented before that decision is made and frozen. If the analysis is predicated on reallocating resources, it may be necessary to be aware of budgetary routines in order to catch the appropriate part of the cycle of financial management. An analyst can be highly proficient in the art of communication, but if he has little sense of time—that is, if he does not know when it is opportune to communicate—his skills will not be worth much.

Organizational Structures

Another important factor is organizational structure and its capacity to absorb information. Over the years, an organization will develop a certain structure for the processing of work. Let us consider a fairly typical structure, a hierarchy composed of many levels, specialized offices, and centralized decision making. As part of its processing of work, such a structure also processes large quantities of data. In so doing it absorbs and filters that data. Hence an analytical report that is inserted into the organization will be modified as it goes

through numerous levels before reaching its final destination. Modifications can take place as each level either rewrites the report or encloses transmittal memoranda that select certain aspects of the report for treatment. This digesting of data occurs simply to avoid placing an informational overload on central decision makers, but it also serves as a protective device to prevent damaging information from being communicated to one's superiors. In any event, organizational structure introduces a measure of distortion into the communication of analysis.

Competition of Information

Analysts generally value their information more highly than do members of the organization. Not perceiving that the organization is already flooded with information, analysts are seldom sensitive to the costs of absorption. Of course, a single report will not overload the informational capacity of an organization. Assuming that the communication of analysis takes place in an informational vacuum, however, can lead to miscalculation. One can easily forget, for example, that competing information is available from the communication lines of the organizational structure, and that such information may be in conflict with the direction and findings of an analysis. The analyst must take into account in his analysis and in his communication of it that his audience will have received information that has been processed by the organizational structure. Although the analyst may circumvent the organizational structure to get his data, he will still have to contend with the sometimes distorted views of others.

Circumventing a number of superiors to reach a sympathetic ear or reaching directly down several levels to a friend for data are common tactics for avoiding the effects of organizational structure. While such tactics can be effective, they also have high personal costs in terms of the analyst being marked as disloyal or being denied access in the future. Analysts who have some perseverance might be better off if they cope directly with the organizational structure in their conception of the audience and in the design of the means of communication.

Identity

Organizations also have a self-image that asserts, "This is what we are, this is what we do, and this is how we do it." In communicating to an audience, the analyst has to tailor his message so that it is consistent with this organizational identity. Suppose an organization's identity involves a conservative streak and a respect for traditional ways of doing things. Analysts, usually biased toward significant change, will have to cloak their recommendations for divergence from that tradition. What is new will have to appear old or only slightly different from present practices. Effective analysis is the product of a number of

people and must be put into a framework consistent with the organizational identity.

Environment

Organizations also exist in a larger environment. They have permeable boundaries and are subject to the external influences of articulated interests. By acting as if these influences do not exist, as if the organization is a "closed" system, the analyst may make the mistake of concentrating his efforts at communication solely on sponsors and other members of the organization. Suppose an analyst examining alternative early childhood development programs focuses the report only in terms of the organization's perceived and stated interest in cognitive achievement, while downplaying or omitting his analysis of the other interests of parent and teacher groups. Such a report may not be well received by those groups when they review it; and, what is more important, the report may not be well received by the members of the organization, because they also perceive the problem through the lens of these environmental influences.

Functions

Finally, the analyst should be sensitive to the variety of functions that a particular communication fulfills for the organization or a component of the organization. A common error is to assume that the communication of analysis serves only a single function—transmitting policy and programmatic information. Sometimes communication is used to establish the jurisdiction of the organization, to show that the organization has control over, and is working in, an area. Sometimes communication serves to maintain the analytical unit and to secure resources for its activities. At other times it is used to develop political support and to neutralize the opposition of the organization. And then there are times when the communication of analysis is strictly a symbolic act indicating concern about a problem that the organization actually intends to do nothing about.

Conclusions

In short, in choosing and calibrating their audience, analysts must be sensitive to the organizational context of communication—to such factors as timing, structure, competition of information, identity, environment, and functions.

ASSUMING THE CLIENT IS THE AUDIENCE

Analysts would do much better at communicating if they used the notion of audience rather than of client. *Audience* is a neutral term; it does not prejudge

the situation of decision making. The use of audience allows the analyst to regard potential recipients without bias and frees the analyst to attempt to influence the selection of the audience's members. Assuming that the client is the only member of the audience, however, can lead to serious errors.

One such error occurs when the analyst assumes his client to be *the* policy maker. Generally we think of clients as policy makers—persons who instigated analysis and are expected to use its results. They are our links to policy making. Yet, all too frequently, the person who sets the process of analysis in motion—the immediate client—is merely one link in a complicated chain of advisors and policy makers. Indeed, that person may not be a policy maker at all, but a manager who has arranged and monitored an analysis for someone else. In complex organizations and political systems, the immediate client is, at best, only one of numerous policy makers who comprise the analyst's audience, and the analyst may be mistaken in focusing his communication on that single client.

Related to this often erroneous idea of a single policy maker is the analyst's belief that his client will be the sole recipient of his communication. No completely foolproof way exists for preventing the dissemination of analytical information. Whenever the analyst commits his information to paper, the information takes on an independent existence. Any reader of a study can become a member of the analyst's audience if he chooses to use the information. If the analyst recognizes that the material he is communicating is sensitive and can be misused, however, he can take some small steps to restrict its utility to others. For example, a study can be written so that it applies only to a specific context from which it would be difficult to extract the findings.

Shaping analytical communication to fit the audience is a highly desirable practice, but one partially dependent on appropriate selection and understanding of that audience. Thus, as we have seen, treating the immediate client as the sole member of the audience may lead to errors. In the first case, the analyst unduly restricts his communication so that it is not likely to appeal to other policy makers. In the second case, the analyst does not restrict his communication enough, and other policy makers misuse it.

Concentrating on the immediate client as sole recipient implies that the policy process is more or less constant. The analyst assumes that the person who contracted with him will be there to receive the product and that the decision will wait for his research. Yet all too often, when the analyst is ready, the immediate client has been replaced by an uninterested monitor, the key elected and administrative officials have left office, and the current crop of actors are on to another problem. Because policy and decision processes are fluid and dynamic, analysts cannot afford to tie themselves to a rigid conception of the audience.

In developing a strategy for the selection of the audience, the analyst should not become entangled in legal niceties. All too often, I have seen analysts

restrict themselves and their communication because the contract called for so many copies to be delivered to a particular person. A contractual or formal arrangement may in fact exist between client and analyst; there is no reason to assume that such arrangements have to be breached, but there is every reason to augment them. No one is suggesting that the analyst ignore the wishes of the client, whether they be contractual or otherwise; it would be foolish to do so. Indeed, it is in order to protect the interests of the client that the analyst has to examine the potential audience for his work. Whatever the sophistication and intelligence of the client, most of us would agree that in the practice of analysis it is inappropriate to accept only the client's definition of the problem without examining other definitions. It is also important that the analyst have room to maneuver in selecting the audience and in communicating the results of analysis.

Thinking strategically about the selection of the audience leads us to consider the "sequencing" of the audience. Sending the results of a study to the wrong person can cause the premature death of that study. The selection of the *initial* audience is crucial in this respect. Reaching a supportive staff person, for example, can do more to influence a decision maker than a more direct approach by the analyst. In a large agency with a well-established analytical office, the analyst can sometimes create a climate of receptivity for his work as he goes along. This process of building consensus can start quite modestly, with the analyst convincing first his fellow analysts, then the head of the analytical office. After having built up a cadre of supporters inside the analytical office, he can proceed to communicate with superiors and subordinates outside the office and eventually with peers and policy makers outside the agency.

In sum, when an analyst assumes that his immediate client is his audience, he gives up the opportunity to influence the choice of audience. From the viewpoint of communication, who is reached and when they are reached is as critical as what is said. Making up a distribution list for a report should not be treated as a trivial mechanical exercise. If analysis is to affect decisions, the analysis must reach the decision makers and those friends and enemies who can influence them. This is not to say that an analyst can always anticipate all those who will be in the audience, but that thinking strategically about the composition of the audience is essential for effective communication. There are many clients to be reached; some may be peripheral, some remote, some in the future, yet all are part of a potential audience (Meltsner, 1976, pp. 200–210).

ASSUMING THAT THE CLIENT IS AN ANALYST

One of the common errors in communication occurs when the analyst assumes that his audience or client is like himself, sharing his beliefs and values, his

definition of the problem, his language and education. In the particular client who requested the study he will expect a sympathetic and empathetic audience. Such expectations may be unrealistic. Running for office is not the same experience as running a computer. Policy makers and decision makers, while perhaps competent judges of analysis, have seldom been analysts. They may come from a different class and social background than the analyst. They may not even remember that they instigated a study. Even in situations where the client is a former analyst, the analyst should not assume that the client is *still* an analyst. True, the client and the analyst may share a common language, but it is unlikely that they will completely share policy preferences and definitions of the problem. Differences in perceptions and values are bound to arise because of the client's position in the decision-making system. Sitting where he does, the client develops concerns different from the analyst's—concerns about politics, administrative feasibility, and personal advancement.

The analyst should assess rather than assume the degree of similarity between himself and his client. Two general situations can bracket this assessment, as Rogers and Shoemaker (1971, p. 39) point out:

> *Heterophily* is the degree to which pairs of individuals who interact are different in certain attributes, such as beliefs, values, education, social status, and the like. The opposite of heterophily is *homophily*, the degree to which pairs of individuals who interact are similar in certain attributes. Generally, most human communication takes place between individuals who are homophilous, a situation that leads to more effective communication. Therefore, the extent of ... heterophily ... leads to special problems in securing effective communication.

If Rogers and Shoemaker are correct in their inference that "most human communication takes place between individuals who are homophilous," the analyst's assessment should indicate the necessary specific elements for effective communication. If these elements are lacking, it is the analyst's responsibility to adjust the situation. Obviously, the client does much to determine the effectiveness of communication; our focus here is on the analyst and what he should do—namely, increase the degree of similarity between himself and his audience. Analysts who have had experience in dealing with numerous clients usually do this without much thought: they listen sympathetically to the client, respect his knowledge of the situation and problem, and are able to convert the technical language of a report into the language of the client. They do not assume that the client is naive and easily manipulated, but they try skillfully to couch their arguments in terms that increase the appearance of shared values. They are so knowledgeable about their client that they read his mind. This ability to anticipate the reactions of a client is difficult to acquire and may depend on considerable prior exposure to the client. As the Chinese philosopher Han Fei Tzu (Watson, 1967, p. 73) reminds us:

> On the whole, the difficult thing about persuading others is not that one lacks the knowledge needed to state his case nor the audacity to exercise his abilities to the full.

On the whole, the difficult thing about persuasion is to know the mind of the person one is trying to persuade and to be able to fit one's words to it.

How to increase homophily between analyst and client is not at all clear. We seem to know what direction to take but not what steps will be effective in a variety of situations. The exact prescription rests on the skill of the analyst and the idiosyncratic aspects of the situation, not on empirically based generalizations. If the analyst stops attempting to make the client over in his own image and instead puts himself in the position of the client, communication will at least be enhanced.

ARGUMENT

Once the analyst knows his audience, he can determine how to communicate. He will have to develop arguments that combine the substantive aspects of his study with its social context. An argument is the theme that holds the various parts of the study together and leads them to a particular conclusion. The conclusion may urge the acceptance of a specific policy alternative or of a novel way of perceiving the problem, or it may suggest promising avenues of research. When the argument is well executed, we usually say that the conclusions of the study followed from the analysis: the argument has been so constructed that the study appears to be correct and its conclusions compelling. In short, an argument should be persuasive.

Now when I say compelling and persuasive, I do not mean that an argument as used in the practice of analysis is or should only be a strong appeal to the emotions of the audience. Commonplace techniques such as the use of analogy and comparison can also be persuasive. The use of *a fortiori* analysis, where we place our preferred alternative at a disadvantage to see if it is as good as we thought, is not only an analytical technique but also an effective argument in communicating with an audience.

Many mistakes in communication derive from the unwillingness of the analyst to recognize that his study's argument is not neutral, that it leads to some conclusion, and that it is meant to be persuasive. Other mistakes related to argumentation result from the difficulty of constructing an argument that tells a client something he does not want to hear or from the analyst's insistence on dragging the audience through an obscure methodological or mathematical line of reasoning.

ASSUMING AN ARGUMENT IS NEUTRAL

The argument of an analysis should be balanced, but this does not mean that it should not be persuasive. By definition, the communication of analysis is an act of persuasion. It is not the simple transfer of information from analyst to

recipient—as if information could be aseptically transferred by itself and acceptance depended solely on the quality of information. Is it not false humility to assert—as many analysts do—that we merely provide information to help the decision maker or to improve the quality of decision making? Even where the directional or steering content of the information is minimal, analysts are engaged in an act of persuasion. Our analysis has led us to accept an altered view of reality and that is the view we present to the recipient. We say: Here is our new reality; here is cause and effect; here is *A* and if you do *A*, *B* will follow. We do not have to ask the recipient to choose *A*, only to accept that if *A* is chosen *B* will follow. Essentially, we are asking the recipient not necessarily to do something, but to accept our view of reality and to change his framework of understanding. When we sincerely believe that our view is correct, and not clouded with uncertainty, and especially when we believe that our analysis indicates the correct path, what then? The most detached, unemotional, objective analyst cannot avoid seeking some confirmation from others. No matter how neutral he tries to be in his communication, he is engaged in persuading the recipient to believe him, and his problem is to do it well, convincingly, and evenhandedly.

Some analysts may object to the assertion that communication is an act of persuasion, because to them persuasion implies a manipulative appeal to emotions, and they see themselves as appealing solely to reason. Appeals to reason can also be manipulative, but as long as analysts recognize that they are appealing to something it will be sufficient for my purposes. Consider the scientific analyst who expects that giving his report to the decision maker is the end of his responsibility. Compare that analyst with one who advocates a particular policy and expects that with considerable activity and skillful communications the decision maker will adopt the policy. These analysts have differing views of what is being communicated, and they have different styles or employ different means of communication; but, oddly enough, they have similar expectations. They both expect the decision maker to embrace the communication and to behave differently because of it. In a sense, both analysts are advocates, although one of them is unaware of it.

TELLING THE CLIENT WHAT HE WANTS TO HEAR

It is a sad commentary on the state of analysis, but all too often a client will listen if you tell him what he wants to hear and will not listen if you tell him what he needs to know. Why should this be so? It has to do with a person's tendency to select and distort what he perceives on the basis of his predispositions. Analysts, of course, are subject to the same tendencies, but norms of professional practice and review usually eliminate or modify many of the possible distortions. Perhaps clients do not like suprises.

Research on mass communications, in the laboratory and the field, have

come up with several generalizations that seem applicable to the communication between analyst and client. Consider these generalizations from Berelson and Steiner (1964, pp. 529–541):

> People tend to see and hear communications that are favorable or congenial to their predispositions; they are more likely to see and hear congenial communications than neutral or hostile ones.
>
> People tend to misperceive and misinterpret persuasive communications in accordance with their own predispositions, by evading the message or by distorting it in a favorable direction.
>
> People respond to persuasive communications in line with their predispositions, and they change or resist change accordingly. Communications will be most effective—that is, will secure the response most in line with the intention of the communicator—when they are in accord with audience predispositions; when they tell people what they (most) want to be told.

Analysts clearly have a great deal to contend with if the behavior of clients is similar to that of most people. Of course, sometimes the client's predispositions are congruent with the analyst's, and in such cases the problems of communication are fewer. But what about a client whose predispositions are in conflict with the recommendations of the analyst? No amount of prior consultation with the client about the definition of the problem can prevent such a situation from arising. What then can be done to ensure a fair hearing of an analysis and increase the chances of its acceptance? What alternatives are open to the analyst?

First, the analyst can decide to make no attempt to understand why the client's predispositions are in conflict with his recommendations. He can ignore the conflict and present his facts and recommendations with the expectation that the study will speak for itself. This alternative operates on the assumption that human beings are reasonable and that knowledge can prevail. Second, the analyst can recognize the conflict and present his study with the same facts and recommendations, but do so using a personal style and form of communication that smooths out the abrasive aspects of the situation. This alternative operates on the assumption that human beings enjoy flattery and are prone to listen to those they trust. "The important thing in persuasion," Han Fei Tzu tells us, "is to learn how to play up the aspects that the person you are talking to is proud of, and play down the aspects he is ashamed of" (Watson, 1967, p. 75). Third, the analyst can recognize the conflict and modify his study so that the argument, with its selected facts and proposed recommendations, is more congruent, though not completely so, with the client's predispositions. This alternative operates on the assumption that it is preferable to compromise oneself partially than to be completely compromised or not listened to at all. Fourth, the analyst can recognize the conflict and change his study so that it is completely consistent with the client's predispositions. This alternative operates on

the assumption that future influence depends on present submission by the analyst and acceptance by the client.

As in many policy problems, not one of these alternatives is attractive. Each raises serious issues of both ethics and practicality regarding appropriate roles and standards for analysts. Yet surely we recognize that communication, by definition, does involve manipulation, anticipation of the client, and planned reaction by the analyst. The real question of professional ethics is not whether to manipulate the client but how much manipulation is acceptable. The first and the last alternative in all likelihood result in the client not hearing or accepting what he needs to know—a dubious ethical choice at best. The other alternatives, in which the analyst slightly changes his behavior or his study, offer the best chance for effective ethics and communication. Oddly enough, when viewed in this context, the second alternative, modifying the style of the analyst and the form of his presentation, seems most appropriate because it leaves the content and message of the study intact. The analyst does not have to settle for a less than preferable solution. Form is changed but not substance. Rather than telling the client what he wants to hear, the analyst tells him what he needs to know by making it appear that what is being said *is* what he wants to hear.

To do this involves tactics that orators and advisors have known and used throughout the centuries: making sure the time is right, stating the argument in terms familiar to the audience, presenting a departure from present policy as merely a slight extension of it, coupling the recommendation with the possibility of personal advancement for the client, and pointing out that the recommendation is really the client's and that he will be identified with it and get the credit for it. Add to this choice of clothes, haircut, and appropriate regional speech pattern, and the nuances of persuasion are endless. When John Stuart Mill was a young boy, he enjoyed reading the orations of Demosthenes. His father pointed out (Mill, 1924, pp. 14–15)

> the skill and art of the orator—how everything important to his purpose was said at the exact moment when he had brought the minds of his audience into the state most fitted to receive it; how he made steal into their minds, gradually and by insinuation, thoughts which, if expressed in a more direct manner would have aroused their opposition.

No doubt many analysts will be uncomfortable with the notion that they may have to manipulate a client or audience. They will argue that it is impossible to change form without changing substance, and they will consider attempts to do so expedient and unethical. They may well be right, but one thing is certain: if they do not use the manipulative arts of persuasion, their communications will be effective only in situations in which the values, attitudes, and beliefs of the participants are reflected in the analysis. With the exception of somewhat trivial and technical problems where such congruence by definition exists, such situations are not common.

TELLING THE CLIENT WHAT YOU DID

It is an axiom of the analytical world that clients are busy, impatient, and have short attention spans. They want to know only as much as they will need to explain their decisions to others. With the exception of crises—when the demand for information is real and urgent—a client often reacts to the communications of the analyst with feelings of boredom and indifference. In a face-to-face situation, he may do so because he does not want to appear committed. Reading a report in the privacy of his office, he may easily be distracted by more pressing items on his agenda; even though he continues to read, his mind and heart are likely to be elsewhere. Keeping the client's attention is a struggle, and many an analyst loses it because he insists on telling the client what he did rather than what he learned.

It is true that a client will not completely understand the analyst's conclusions and recommendations if he does not know how they were reached, but is total understanding what we are after? I would suggest that often what we are after is acceptance. Now most of us would agree that acceptance with understanding is preferable. The two ends are not always compatible or attainable, however, and thus we usually settle for acceptance with an illusion of understanding.

On the other hand, many analysts feel that acceptance itself is closely linked to a detailed documentation of methods. They reason that a presentation of methodology is one way of establishing the expertise of the analyst and the value of his analysis. The rationale here is legitimation and not necessarily enhanced understanding. The analyst believes that he will not be trusted or his work approved unless he presents his expertise in detail. How can one be viewed as an expert unless he shows the paraphernalia of his models, statistics, and the like? If the point of such demonstrations is to maintain one's reputation as an expert, then surely there must be other ways to do so. Association with a prestigious university or research firm, necessary credentials in terms of education and experience, and recommendation by well-known former clients also contribute to the maintenance of reputation. Indeed, if one has to demonstrate repeatedly that he is an expert, he is likely not to be—at least by reputation.

How an analysis is conducted can also affect the ease of communication. Consider the analyst who is taken with a new and sophisticated analytical method. He can hardly wait to try it out. When he does, he may find that he has trouble explaining it and why he used it. Generally, the analyst should choose a simple method that fits the problem and the quality of the data. The more obscure, specialized, and complex the method, the more difficult it will be to explain it and the more the analyst will be forced to provide detailed documentation so that the study can be validated by others.

The client is more likely to listen if the material is directly related to his concerns. If he wants to know how a program is operating, tell him about what

contributes to the success and failure of the program. If he has to make a decision, concentrate on the policy choices and their consequences. If he wants to enhance his knowledge of underlying causes and relationships, show him the products of research, not the methods. Of course, if he wants to hear about methods, to enhance his knowledge or perhaps to calibrate the analysis, then it is appropriate to inform him. But even in such situations, the sequencing and the amount of attention devoted to methods should be carefully controlled. Generally, methodological details are better left for footnotes and appendixes or for special meetings of staff and technicians.

MEANS

The forms of communication include not only many media but the numerous ways of using them. All too often, the analyst has a limited view of the available choices. He settles for a particular method and makes do. Having learned that a communication technique worked in one situation, he somewhat slavishly continues to use it. The analyst seldom realizes that he needs to develop multiple means of communicating with diverse audiences, rather than simply following mechanical rules.

The analyst is often not given the time, money, or other resources to break out of prevailing communication practices. Audiences and clients want to swallow some "instant analysis" pill, yet they usually are not willing to pay for it. Both clients and analysts are to busy to make the effort that would improve the effectiveness of communication. What can be done about this neglect? I have no satisfactory answer, other than that as analysts we must start with ourselves. We should not so overemphasize communication that our efforts are counterproductive, but we certainly can do more to seek out new ways of reaching out audiences. At the same time, we can educate our clients about the importance of communication.

ASSUMING ONCE IS ENOUGH

There is a trade-off between the resources used for the analysis and those used in communicating the results of the analysis. I have been arguing that analysts do not spend enough of their resources on communication and that they certainly do not make creative use of the resources they have allocated for communication. With a modicum of arrogance, analysts assume that the world of policy is anxiously waiting for their work. They put their results in one final report, do a few briefings, and assume that that will be enough. But how can one report be sufficient when the analyst is dealing with multiple audiences?

Usually, the analyst is aware of at least two audiences, his client and his peers. In fact, the client represents a number of decision makers, and the peers

represent numerous technically minded staffs, advisors, and fellow analysts. The former will participate in the formulation of a policy or the making of a decision. The latter, friendly or not, will review the study for its technical competence and elegance. The informational needs and standards of the two audiences are in conflict and tension, and yet, armed with a single report, an analyst attempts to address them. The logic of the situation would argue, however, that he should have multiple reports to communicate with multiple audiences. But this I do not mean a reshuffling of the material in a single report, but the actual execution of a number of separate communication products—different reports for different readers and different briefings for different audiences.

To many analysts, multiple communication seems unnecessary work and a waste of time. Besides, they argue, the immediate client has not provided sufficient resources, in time or money, for them to do more. From the perspective of most participants, a single product seems sufficient. In the face of this erroneous consensus, it is up to the analyst to assert the need for additional resources for communication. By thinking strategically about the different audiences, analysts can reach effective compromises with their clients and make effective use of available resources. Audiences can be grouped and matched with particular communication media. Nor must analysts stay wedded to traditional reporting formats and briefings. Indeed one analytical team, for example, made a documentary film to promote and disseminate its work (Brewer, 1977, pp. 28–32).

It may appear that such actions will cause redundancy and a clogging of communication channels with excess verbiage and paper. But while the analyst may find himself covering the same material, he will not do so in the same way. The language and the forms of communication will be different, and the results will not be entirely repetitious. Redundancy in communication, in any event, is a practical way of dealing with the audience's tendency to selective perception. When the analyst describes his work in terms—language and form—that will appeal to a specific audience, his analysis is not as likely to be screened out by conflicting predispositions.

Actually, what I am suggesting is only an extension of present analytical practice. Do we not create both reports and briefings? In reports, for example, do we not include a summary for the busy reader? Our reports also have a fairly stylized format: in the introduction tell the reader where he is going, take him there, and then conclude by telling him where he has been. In their efforts at communication, analysts too often behave like the characters in Lewis Carroll's poem "The Hunting of the Snark":

> Just the place for a Snark! I have said it twice:
> That alone should encourage the crew.
> Just the place for a Snark! I have said it thrice:
> What I tell you three times is true.

I do not want to be accused of comparing the pursuit of analysis to a Snark hunt, and I certainly do not believe that repetition creates truth. Yet redundancy, both within a communication product and through a number of such products, can facilitate acceptance of the analyst's message.

FOLLOWING MECHANICAL RULES

Giving advice about communication is a hazardous business. Few people besides students of rhetoric and social psychology have studied what works in communication, and to my knowledge no major study or body of literature evaluates the communication of analysis. In addition, communicating the results of analysis is very much grounded in the specific situation. Few rules of writing and speaking are universally applicable. As we have seen, many idiosyncratic factors seem to influence the effectiveness of communication: the analysis, the structuring of the argument, the client and the audience, the organizational and social context, and the choice of media and technique. While it is easy to set forth some rules that work in some situations, it is easier to set forth exceptions. That is why no analyst should follow in a mechanical fashion a rule of effective communication from me or anyone else. A few examples will illustrate the problem.

I teach my students that using short declarative sentences in their writing allows easier communication with most audiences and clients. Simplicity usually pays. One of my former students is now experiencing a certain amount of frustration because she insists on following my rule while working for a client who does *not* prefer simple, straightforward sentences. In reviewing the work of other analysts, I usually urge that the analyst use ordinary language and get rid of the jargon and technical terminology. Such esoteric language can easily mislead nonexperts. Using ordinary language is a good rule when the policy makers in an audience do not share a common technical vocabulary, and, as Perelman and Olbrechts-Tyteca (1969, p. 153) put it, "ordinary language can help to promote agreement on the ideas." When the audience does share a technical vocabulary, however, using ordinary language is not a good rule because the technical vocabulary is a useful shortcut. I also believe in avoiding weasel words, such as *probably, tends to, perhaps, likely*. We use such words to qualify our inferences and to make up for some inadequacy in the study. If these words are used with skill, the reader will never know what the analysis is suggesting. Yet when the inference stands alone without the benefit of such trappings, it may appear not only naked but flat-footed to some. A client who, because of education or culture, is accustomed to convoluted language would never appreciate such a practice.

Because analysis is fairly abstract, I have found it helpful to supplement the definition of the problem with a vivid example or story. Making things as concrete as possible seems effective, particularly in an oral presentation. But a con-

crete example can also deflect an audience's attention or cause disagreement. An analysis that stays at an abstract level may be more difficult to understand, but it also promotes agreement.

Sometimes it is difficult to know operationally whether one is following a specific rule. Consider the rule that a study report should stand on its own, that the reader should be able to go from a statement of the problem to the recommendations without having to look at anything else. It would be silly to expect total self-sufficiency of a report; yet because communications are constrained—in terms of pages or time, for example—the rule hints at a degree of self-sufficiency. How much should be included? What should be excluded? Should topics generally known and agreed upon be omitted? How is the analyst to identify such topics?

In short, follow a rule if you can, but only if the rule is congruent with a specific communication situation. This last rule supersedes any other mechanical rule you may confront.

OVEREMPHASIZING COMMUNICATION

An essay on communication all too easily overemphasizes the importance of communication at the expense of substance. The same sort of thing happens in the practice of analysis. Some experienced analysts are likely to appear too glib when communicating their results. Some analysts try to substitute communication for substance; words like *overselling* and *snowing* describe such a situation. As one perceptive reviewer of this essay recently told me:

> Many analysts, perhaps because of their own needs for esteem or prestige, perhaps from pressures by their organizations, tend to pack or puff up analytic reports with unnecessary material which—it seems—is intended to impress but often has the opposite effect. It was a standing joke among my colleagues . . . that the fanciest "packaging" usually indicated the poorest substantive quality. Everything from the binding of the report to special type fonts gets used to dress up shabby work, and it's quite easy to spot the analysts who are trying this ruse.

What is needed is balance between skills of communication and skills of substance. Obviously, one depends on the other. Without some substance, there will be nothing to communicate. Without clear and reasoned thinking, the analysis will be very difficult to communicate. Frequently, for example, what is put forth as a communication problem is in reality a thinking problem: fuzzy thinking makes for fuzzy communication. For our purposes, however, let us assume that the substance exists. In what other ways does an overemphasis on communication lead us astray?

One way is by encouraging us to omit embarrassing information. Analysts are often reluctant to admit the weaknesses in their work. In our desire for a tightly reasoned argument in which data and analysis directly support conclusions, we may be prone to leave out nonsupportive information. We believe

our conclusions to be correct; to ease the process of understanding and acceptance we want to leave out whatever detracts from that conclusion. Such a tactic may backfire, however; a third party to a communication may be delighted to point out the contrary evidence, thus undermining the main thrust of the study and calling into question the honesty of the analyst. My own practice is to sacrifice a little on the communication side by preempting the future critic. It often pays to include the adverse example and to admit the limitations of one's work, regardless of the minor difficulty in communication this causes. Such candor provides a measure of balance and softens an unnecessary tone of advocacy.

Another quite different way we can be led astray is by overemphasizing a particular medium. We become media freaks. First, we narrow our choice down to written reports or briefings and tend to slight other forms of communication—such as writing memoranda, and using the telephone, and taking advantage of social occasions. Then we put our energies into making sure our reports are neatly bound, with appropriate dividers in color. Or we develop briefings that make use of three projectors and artistic slides with numerous overlays allowing easy digestion of complex information.

These techniques work fairly well with many audiences. The problem arises when technique displaces what we were trying to do in the first place—communicate. Consider that one central aspect of effective communication is feedback. The analyst has to know whether his message is getting across to his audience and, if it is not, how to make suitable adjustments. The question then is not only which technique to use but how a particular technique enhances communication and provides the condition for feedback. Obviously, a face-to-face, informal situation would be best from this perspective. Both the analyst and his client could then engage the substance of the analysis and at the same time build mutual understanding and trust. Short of this ideal situation, the analyst has to ensure that his formal briefing, for example, is not so formal and simplistic that it entertains without engaging the audience or encouraging the audience's reponses to substance.

CONCLUSION

Concern about communication usually masks concern about the lack of one's influence: "The analysis is sound. Why is it that no one pays any attention to it?" Worse than the analysis not being used is the analysis being misused or used in unintended ways. What can be done about this? Some would argue that the analyst must be his own client and assume a political role. In this vision, the analyst can become a master coalition builder, seeking allies and neutralizing enemies in the pursuit of a study's recommendations. Yet I do not think most analysts have the stomach for the rigors of politics. Some more modest prescription is needed.

Here is where an appreciation of communication and its social context comes in. Helping to choose the audiences, designing arguments to fit those audiences, and employing effective means of communication will increase the use of analysis. To this end, analysts will have to improve their understanding of rhetoric and the psychology of communication. They will also have to accept the necessity of spending as much time on communication as they do on the analysis itself. Communication cannot be effective with meager resources or as an afterthought. Paying serious attention to communication is, so to speak, being political in the small, a sensible strategy consistent with present analytical practice and the analyst's limited resources. Analysts who then come to appreciate the art of effective communication will be in the good company of Aristotle, who pointed out the uses of rhetoric (*Rhetorica* I. 1. 1355[a]):

> Rhetoric is useful . . . because things that are true and things that are just have a natural tendency to prevail over their opposites, so that if the decisions of judges are not what they ought to be, the defeat must be due to the speakers themselves, and they must be blamed accordingly. Moreover, . . . before some audiences not even the possession of the exactest knowledge will make it easy for what we say to produce conviction. For argument based on knowledge implies instruction, and there are people whom one cannot instruct. Here, then, we must use, as our modes of persuasion and argument, notions possessed by everybody.

ACKNOWLEDGMENTS

I wish to thank S. Altman, E. Bardach, E. Beale, G. Chacko, C. Bellavita, G. Majone, M. Morris, D. Selby, A. Sindler, and P. Tannenbaum for their helpful comments.

REFERENCES

Berelson, B., and G. A. Steiner. 1964. Human Behavior: An Inventory of Scientific Findings. New York: Harcourt, Brace and World.

Brewer, G. D. 1977. What Happens After the Reports are Filed. Technical Report No. 4. New Haven, Connecticut: School of Organization and Management, Yale University.

Kahn, H., and I. Mann. 1957. Ten Common Pitfalls. RM-1937. Santa Monica, California: The Rand Corporation.

Meltsner, A. J. 1976. Policy Analysts in the Bureaucracy. Berkeley: University of California Press.

Mill, J. S. 1924. Autobiography of John Stuart Mill. New York: Columbia University Press.

Perelman, C., and L. Olbrechts-Tyteca. 1969. The New Rhetoric: A Treatise on Argumentation. Notre Dame, Indiana: University of Notre Dame Press.

Quade, E. S. 1975. Analysis for Public Decisions. New York: American Elsevier.

Rogers, E. M., with F. F. Shoemaker. 1971. Communication of Innovations: A Cross-Cultural Approach. 2nd ed. New York: The Free Press.

Toulmin, S. E. 1964. The Uses of Argument. Cambridge: Cambridge University Press.

Watson, B. (trans.). 1967. Basic Writings of Mo Tzu, Hsün Tzu, and Han Fei Tzu. New York: Columbia University Press.

9 On Designing Implementable Programs

Eugene Bardach

A well-known law of policy implementation holds that "whatever can go wrong will." This law is affirmed by universal experience, and there is no use protesting against it. Acceptance of this law need not imply defeatism, however. Our task must be to design policies with fewer things that "can" go wrong. This is not, alas, a goal easily attained. For one thing, we understand only imperfectly the theoretical limits of those institutions and processes we normally rely upon to implement policy—bureaucracies, grant-in-aid mechanisms, government procurement contracts, profession-dominated markets, legislative oversight committees, administrative audits, and so forth. Second, even when we have a fairly good grasp of the theoretical limits of implementation mechanisms, we are bound to be uncertain about what these limits are in any concrete and specific case. Finally, neither theory nor practical wisdom can obliterate the simple fact that sometimes we have little choice among implementation mechanisms. We can limit the number of things that "can" go wrong, or limit the damage they inflict when they eventually do go wrong, but in the end we are always stuck with certain irreducible risks posed by the implementation process.

SOME METHODOLOGICAL CONSIDERATIONS

TO WHAT SHOULD "IMPLEMENTATION" REFER?

Scholarly and professional interest in the problems of policy implementation has grown very rapidly in recent years, with the result that it is no longer clear (if it ever was) what the term "implementation" refers to. I would identify at least four referents:

- *Adoption* of a policy recommendation by an authoritative individual or institution, as in "The client has endorsed our analysis of airport landing fees and will soon promulgate regulations to implement it." This is typically an operations researcher's or planner's usage (Vertinsky; Huysmans, 1970; Alterman and Hill, 1974).
- The *empirical details* that reflect, or represent, the application of a policy principle, as in "Effluent taxes are fine in principle, but when it comes to implementing them, we see that they are often set too low." (It should be noted that "implementation" in this sense is bound to be somewhat imperfect, for the same logical reason that there is always a slippage between a theoretical construct and the operations that are used to measure it empirically.)
- The *operating routines* of an organization, or a network of organizations, that have been brought into being or have been modified by some policy mandate, as in "Over the years we have found it necessary to shift our implementation methods away from a reliance on regular audits to voluntary compliance and self-reporting."
- The *process* of rearranging patterns of social conduct so as to honor the prescriptions set forth in some policy mandate, as in "We are implementing the new workmen's compensation law just as fast as we can."

In this essay "implementation" will mean only the process indicated by the last definition. We shall attempt below to give greater conceptual clarity to this process, but at this point it is worth emphasizing where it fits into the larger process of policy development and execution. To put the matter simply: first there are words and then there are deeds. The implementation process is the *social activity that follows upon, and is stimulated by, an authoritatively adopted policy mandate,* which prior to implementation is only a collection of words.[1] This policy mandate might be a piece of legislation, a court decision, or a bureaucratic command. If the implementation process is bounded on its near side by the policy adoption process, it has an analogous boundary on its farther side: at some point in time it turns into a relatively fixed set of operating routines (one of our rejected definitions, above). At this farther point the mandated program is no longer in the process of "being implemented": it is established; it is in place.

The implementation process, therefore, is a transition process. It takes whatever social, political, and economic arrangements exist before the policy mandate is adopted and sets them in a different configuration. To the extent that this latter configuration resembles the arrangements intended by the policy mandate, the implementation process has been successful.

THE MEANING OF "SUCCESSFUL" IMPLEMENTATION

If success has a relationship to intentions, we should here note certain of the objectives that lie behind all policy designers' "intended arrangements." They

are (a) to achieve the goals defined by the policy; (b) to hold delay to a reasonable level; and (c) to hold financial costs to a reasonable level. The first two of these concerns are well documented in the seminal scholarly work on the implementation process, *Implementation*, by Pressman and Wildavsky (1973). They report that in 1966, the U.S. Economic Development Administration (EDA) announced that it would spend over $23 million in Oakland, California, and create 3,000 jobs for unemployed inner-city (mostly black) residents, but 3 years later only $4 million had been spent and 68 new jobs created. The two main projects were to be built by the Port of Oakland. One was an airport hangar and support facilities to be leased to World Airways, and the second was a marine terminal and access roads. EDA financing was contingent on the agency's approving minority-hiring plans to be submitted in advance by the prospective employer. But it proved immensely difficult for the many concerned parties to reach agreement over these plans.

Frustration of the third of our postulated major policy design intentions, holding costs to a reasonable level, is frequently observed in large-scale public works projects. Capital spending on the new San Francisco Bay Area Rapid Transit system was originally (in 1962) estimated at $792 million, but before it is done, the actual system cost will have doubled this estimate. In 1961 the new administration of President John F. Kennedy canceled research and development activity on a nuclear-powered plane. By the time of Kennedy's decision, however, the U.S. Government had spent more than $1 billion, over a 10-year period, on what many scientists and engineers had long believed to be a losing proposition.

If implementation success occurs when the intended results are achieved on time and within the original cost estimates, should we say that the reverse of any of these outcomes is to count as an "implementation failure"? Not necessarily. For one thing, cost estimates are often too low not because of implementation problems but because of systematic downward bias in the initial estimates. Moreover, if policy implementation runs into trouble with respect to goal attainment or delay, cost escalation is sometimes what bails the policy out. A bridge or tunnel only 80 percent complete does no one any good, after all, and it is often worth putting up with cost inflation to get the job done. As for failure to achieve the initially intended results, this too is not necessarily a failure of the implementation process. Indeed we should sometimes count it as a success, particularly if the original objectives were ill considered or the method of administrative control ill chosen. Rhetoric is cheap, after all, and policy mandates, which are rhetoric almost by definition, can drastically underestimate political as well as economic costs. Federal government officials in West Germany, for example, regularly mandate state officials to make rapid strides in controlling pollution. Federal officials have very little power or official authority in this policy area, however, and it seems inappropriate to attribute

the failure of the policy to "problems of implementation." (Mayntz, 1976, p. 584.)

Nor should we necessarily count as a failure of the implementation process a result that originates in a more fundamental conceptual defect in the policy design. If a policy of rent control, for instance, leads in the long run to housing shortages and deterioration of the housing stock, it would probably be mistaken to blame these outcomes on the way the policy was implemented. Similarly, if after a period of growing dissatisfaction with rent control it is decided to adopt a policy of rapid decontrol, whereupon rents rise steeply and drive middle-class families out of the city, no one should blame the implementation process. "Good" implementation cannot by itself offset the ill effect of "bad" policy any more than a more perfect compass and straight edge can help us square the circle.

To the extent that the implementation process "fails" under such conditions, it does so because it does not help the key actors in the policy process to recognize their conceptual mistakes and to take constructive measures to correct them. It is necessary to emphasize this special case of failure-as-success because policy making is often not a straightforward matter. There is inevitably a certain amount of trial and error, and sometimes the opportunity for trial and error is—or ought to be—built in. A successful implementation process, therefore, not only avoids known pitfalls but seeks better and perhaps unpredictable paths to new and perhaps unforeseen destinations.

THE IMPLEMENTATION PROCESS AS COLLECTIVE ACTION

Policy implies program, and program implies an ordered relation among interacting elements. Normally, the required programmatic elements are quite diverse and are controlled by many different parties. Although obviously every different policy entails a different set of program elements, we may suggest a list of some that commonly occur:

- Bureaucratic management functions
- Administrative regulations and guidelines[2]
- Administrative and financial accountability mechanisms
- Privately provided goods and services, e.g., professional services, land, office supplies
- Willing participation of presumptive clients or beneficiaries (in the case of service programs)
- Bureaucratic inspection and surveillance forces (in the case of regulatory programs)
- Permits issued by public regulatory agencies or elected officials

- Various political "clearances" from officials in levels of government or in public agencies not assigned principal management functions
- Innovative ideas in the realm of program conception and design
- Sources of funds
- Troubleshooting that irons out difficulties and assists in mediating negotiations
- Political support that sustains and protects the entire program and policy

Note that the above list—which is, of course, suggestive and not exhaustive—catalogues program elements or functions, not actors. *Who* is to furnish the requisite program elements is one of several "implementability" questions that a policy designer must answer. Sometimes the answers are obvious, e.g., "No agency other than the Tax Collector should be assigned the job of collecting a new sales tax." Sometimes the answers are not obvious but are unimportant, as when private goods and services are to be purchased from a competitive and largely undifferentiated set of suppliers. And in some cases the answers are highly problematic and uncertain. These last are the elements that a designer must attend to most closely, particularly if they are critical to a program's success.

We take it as axiomatic that no potential supplier of valued program elements will cooperate unless it is in his (or his organization's) interest to do so. Policy designers would do well to remember that they are launching a new mandate on a world already in motion. Their policy will almost surely be less important to others than it is to themselves. For many, it will be but a minor blip in their field of vision. For those who notice it at all, some will think it altogether objectionable and a good many others will view it primarily as a vehicle whereby they can advance policy and nonpolicy interests that have no special standing in the eyes of the mandate designers.

It is not clear to many potential suppliers, however, where, in fact, their interests lie. This uncertainty arises because, typically, so many other parties are potentially involved and so many of them are ready and able to bargain, maneuver, and procrastinate in order to extract the best possible terms for their cooperation. The result is often confusion and delay.

Agreements among the many diverse parties build up slowly and unevenly. Bilateral and multilateral negotiations have many ritualistic elements that add to the delay, like haggling over the exact words in planning documents that no one will pay attention to anyway, and often create as many difficulties as they resolve. As time passes, events overtake the implementation process and make certain tentative agreements obsolete or unacceptable to one or more of the parties. For example, elections shift the party in power, or inflation raises costs relative to budgets, or widespread governmental reorganization suddenly removes managers who had been counted on to assume major program responsibilities, or new civil service rules force a drastic slowdown in recruiting

technical specialists. In any case, even as some agreements build up, others decay, and evidence of the general tenuousness of things imperils even those agreements already reached. Inevitably, there are insinuations of negotiating in bad faith, of treachery and deception, of stupidity and incompetence. Such recriminations poison the atmosphere and make the implementation process even more chaotic and sluggish. In short, then, the process suffers from the well-known defects inherent in collective action, potentially to an extreme degree.

COLLECTIVE ACTION AS A LOOSE SYSTEM OF GAMES

Two metaphors capture the essential spirit of what we have said so far about the sources of implementation problems. One is "games," in that the implementation process is dominated by many actors, all maneuvering with and against each other both for end results and for strategic advantages. The second is "system," in that many of these games have some sort of connection with each other—the outcome in one, for example, affecting a party's strategic resources in the play of another—and all (or most) are related to each other more distinctly than they are related to games "outside" the policy area. To be sure, the "system" is quite loose, but there are enough interdependencies to warrant using the term.

Unfortunately, neither game theorists nor systems analysts are likely to find their standard tools relevant to understanding, much less controlling, such a complex system as an implementation process. The relevant human behavior is too subtle, the political texture too rich, the chances for actors to err or miscalculate too numerous. We must be satisfied with heuristic aids. I have therefore compiled an inventory of some of the most common implementation games and given them tag-line names. Each game is identified principally by the stakes for which the players compete. Our concern, of course, is only with those games that have adverse effects on the process of assembling needed program elements. These adverse effects are of four general types: (a) the diversion of resources, especially money, that ought properly to be used to obtain, or to create, certain program elements, (b) the deflection of policy goals stipulated in the original mandate, (c) resistance to explicit, and usually institutionalized, efforts to control behavior administratively, and (d) the dissipation of personal and political energies in game-playing that might otherwise be channeled into constructive, programmatic action. It should be emphasized, of course, that a game may have more than one type of adverse effect. The grouping of games in the following list is ultimately, therefore, somewhat arbitrary.

THE DIVERSION OF RESOURCES

Easy Money

Governments raise and spend vast sums of money. Government contracts are the mainstay of many firms, particularly in the defense industry; and the larger the public sector relative to the private, the more the government's likely procurement practices stand out as a salient feature of any new policy. Pressures for wide distribution of the patronage inherent in such procurement activity may be taken for granted and are not necessarily a cause for concern. The problem arises when unqualified, unmotivated, or overpriced contractors are given key responsibilities in the execution of a policy. Compounding the problem of the initial mistake in awarding the contract is the difficulty of getting the government out of it once the mistake is discovered. Program managers are always reluctant to admit error, and private contractors are often able to exercise a kind of blackmail power in the circumstances.

Budget

In general, bureau chiefs like large budgets.[3] Because incremental budgeting is practiced nearly universally, the more a bureau spends, the more it can hope to get. In addition, the political allies of a bureau in the legislative branch or in the higher levels of the executive branch tend to favor higher expenditures by the bureau and will reward the bureau for its ability to spend money with higher budgets. Such incentives always lead to unproductive expenditures, of course, but in no case more so than when a bureau's mission is almost solely to act as a financial intermediary, commonly as a distributor of grant-in-aid funds to other levels of government. In such cases the performance of the donor bureau is evaluated almost exclusively in terms of its ability to "move money," and so more of it moves to less effect than is ordinarily the case.

Funding

This is a game played by a grant-receiving bureau (or nonprofit organization) against a grant-giving bureau (although occasionally in covert collaboration with it). The object of Funding is to secure not only money but flexibility in its use. The recipient bureau seeks to maximize the size of its grant while minimizing the constraints on how it must spend the monies. Maximizing the grant level is done by padding requests for funds. Minimizing constraints is accomplished by foiling the surveillance and monitoring routines of the donor bureau. These routines are typically rather ineffective, primarily because programs in the public sector so often aim for effects that are hard to measure and therefore hard to hold agencies to account for.

Easy Life

Bureaucrats play this game in many different ways. One obvious way is to invent ways not to work very hard while appearing to do so; but this is not nearly so debilitating to a policy as working hard but only within convenient, and therefore habitual, modes. If successful policy implementation requires bureaucrats to make a daring break with customary practice, or to invent ingenious ways to carry out a new responsibility, the policy is probably in trouble.

Pork Barrel

We said above in our discussion of Easy Money that political pressures to spread patronage around are not necessarily detrimental, but one circumstance in which they are is when resources need to be concentrated in order to reach some threshold level of effectiveness. In the United States, because legislators are tightly bound to their individual geographically defined constituencies, there is usually enormous pressure to spread resources around. The original conception of the Model Cities program of the late 1960s, for example, was to saturate a few selected cities—as few as two and no more than ten—with federal dollars to assist in reinvigorating declining urban areas. By the time the Congress was finished, however, the number of target cities had risen to 150 while the available funds had shrunk to half the originally conceived amount.

THE DEFLECTION OF GOALS

The goals embodied in a policy mandate typically undergo some change during the implementation phase. They might have been ambiguous and therefore might have required, or at least permitted, further definition. Or they might have been based on a very weak consensus, hastily, and perhaps insincerely, contrived during the contest surrounding the adoption of the mandate. Indeed, interests opposed to the goals of the mandate might have stayed quiet during the adoption contest precisely because they counted on subsequent opportunities to achieve more decisive, and less publicized, victories during the struggle over implementation. In an important sense, implementation is the continuation of policy-adoption politics by other means. The politics of renegotiating goals can lead in several directions: trimming them back, distorting or perverting them, or even adding to them in a manner that eventually leads to an unsupportable political burden.

Keeping the Peace

The new wave of "social regulation" in the United States has created many new programs to extirpate real or imagined evils in respect to consumer

product safety, worker health and safety, environmental protection, fair hiring practices, and other areas as well. This new wave follows an older wave of economic regulation exemplified by agencies such as the Interstate Commerce Commission and the Civil Aeronautics Board. Such programs create opportunities for "zealots" to seize control of the regulatory apparatus. They also create incentives for "counterrevolutionaries" to move in and attempt to neutralize the legislative mandate during the implementation phase. The counterreformation usually proceeds through maneuvers aimed at (a) installing sympathizers at or near the top of the enforcement agency (Bernstein, 1955); (b) writing regulations, standards, rules, and guidelines that tend to condone existing practices either explicitly or through generous exemptions (Mayhew, 1968); (c) lightening the penalty structure for violations (Majone, 1976); (d) setting high standards of proof for violations, and throwing the entire burden of proof on the regulatory agency (Jones, 1975). Such maneuvers are often not vigorously resisted by political leaders in the executive branch or in the legislature, who do not relish continual turmoil, and many of whom, although they may have supported the new program initially, were more interested in making a political gesture than in seeing that much regulation is actually accomplished (Edelman, 1966). In many cases the regulated firms do not so much "neutralize" the agency as convert it into an instrument that helps them achieve cartelizing objectives (MacAvoy, 1970; Stigler, 1971).

Up for Grabs

What Keeping the Peace is for regulatory bureaucracies this game is for service-delivery bureaucracies. The conflict here is not between proponents and opponents, as it is in Keeping the Peace, but between two or more different sets of policy proponents, each of whom has a somewhat different conception of the desirable policy objectives for the agency. A case in point here is the U.S. Small Business Administration (SBA), which began in 1953 as a concession to a certain form of liberal and populist congressional opinion that wished to subsidize "small business" (Zeigler, 1961). The Eisenhower administration made sure that the agency was administered by "big business" sympathizers in the Commerce Department, however. In the later years of the Johnson administration, the SBA made more efforts to channel funds to minority businessmen, but this orientation was largely reversed by the more conservative Nixon administration (Blaustein and Faux, 1972).

Piling On

If a new program enjoys certain initial successes, it naturally expands its political support. It then becomes a target for interests who may have only minimal commitment to the program's objectives but who wish to capitalize on

its growing political assets. Such a program is vulnerable to Piling On in much the same way that a cash-rich corporation is vulnerable to being taken over by another firm through a merger or a tender offer to shareholders. By the time the Piling On process is over, the original program goal may have become greatly submerged or the supporting coalition may have collapsed under the weight of the new interests. For example, the concept of "affirmative action" in the United States once meant a commitment to give preferential treatment to job applicants from racial minorities when in other respects they were "equal" to applicants from nonminority backgrounds. Over the course of the last 10 or 15 years, however, the concept has come to mean quotas and a deemphasis on the "equality-in-other-respects" criterion. Many traditional liberals who supported the more restricted "affirmative action" concept have become resentful and frightened at its maximalist redefinition and, in effect, have withdrawn their energy and attention even from the initial goals.

In some sense, Piling On is just another name for the normal processes of incrementalist politics. In that sense, implementation processes and mandate-writing processes are all of a piece in the long run. Thus the United States over the last 15 years has moved from limited health insurance for the poor to the brink of comprehensive health insurance for all persons.

The anticipation of Piling On effects sometimes influences the political struggle over the adoption of the policy mandate. In the field of pollution control, many policy analysts would advocate effluent taxes rather than "legal orders" as a basic administrative device. But industry tends to resist the former more than the latter, it appears, partly because the latter can probably be more easily combated, and also because such a tax can too readily be transformed from a mere Pigovian price adjustment to a powerful revenue-raising device (Ackerman *et al.*, 1974; Kneese and Schultze, 1975; Majone, 1976).

THE DILEMMAS OF ADMINISTRATION

Many of the program elements at issue in an implementation process are assembled under the central direction of an administrative agency. Administration is a control process, however, and as such it inevitably sets in motion games of evasion. Anthony Downs (1967, p. 147) has stated the process as the "Law of Counter Control. . . . The greater the effort made by a sovereign or top-level official to control the behavior of subordinate officials, the greater the efforts made by those subordinates to evade or counteract such control." Counter-control is occasionally exercised through outright defiance—usually when it is possible to coordinate a resistance movement—but it usually takes the form of playing Tokenism or Procrastination, games that are familiar enough not to require description here. If the objects of control have near-monopoly power over important organizational resources, like direct linkages

to the clientele population, the would-be controllers may have a very hard time indeed against these games.

The literature on the administrative process and the exercise of authority in organizations is vast, however, and it would serve no purpose to recapitulate it here. Suffice it to say that because most of it refers to private business organizations, it does not convey the true magnitude of the control problem in public organizations. The essential difference is that most business firms are ultimately subject to the discipline of the marketplace, which is a sterner master than public institutions like legislatures or central budget offices. The reason for this difference, of course, is not that government somehow fends off market tests of its activities—though it does sometimes do so—but that it would not have undertaken many of its functions in the first place had they been carried on in the private market.

THE DISSIPATION OF ENERGIES

Finally, we may mention a number of different games that generally have the character of maneuvering for position in other games, an activity that creates delay and drains personal and organizational energies away from more constructive activities.

Tenacity

If participants in the implementation process have different preferences for how fast the program is to get under way, those with stronger preferences for speed are vulnerable to manipulation by others who are less concerned with speed. The great danger is that the holdout parties will overplay their hand and that everyone will become discouraged. The result can be total collapse (Pressman and Wildavsky, 1973, pp. 45–47, 110) or the erosion of financial and political support (Marris and Rein, 1967, pp. 94–119; Greenstone and Peterson, 1968, pp. 286–287).

Territory

Bureaucratic organizations normally wish to maintain and enhance their legal authority and political status (Banfield, 1961). Bureaucratic competition for this sort of "territory" can be constructive, but very often it leads to difficulties. For instance, it can interfere with efforts to coordinate different agency's responsibilities at the operating level. This occurs frequently in social service programs (Bardach, 1977, pp. 151–152; Rocheleau, 1976). Other documented cases of interest are in regional transportation systems design (Zwerling, 1974) and civil rights enforcement (Orfield, 1969; Radin, 1977).

Not Our Problem

This game is the mirror image of Territory. Instead of trying to grab more territorial responsibility, agencies shrink from it (Williams, 1975, pp. 549–552). Some responsibilities, after all, are so burdensome and so unglamorous that not even the most imperialistic program manager would want them. Three California state agencies spent almost a year trying to avoid the task of writing licensing regulations for community care facilities for the mentally ill, for example (Bardach, 1977, p. 161). Not Our Problem is particularly likely to occur if new assignments are being heaped on an agency without a concomitant increase in budgetary resources.

Odd Man Out

The inherent uncertainty that attends collective action makes it useful for actors in the implementation process to preserve their options to withdraw and cut their losses. This strategy sometimes entails maneuvering other parties into foregoing their own withdrawal options, typically by imposing "sunk costs" on them that they will hesitate to abandon.[4] Because parties resist being maneuvered in this way, the result can be stalemate. This is the scenario that seems to have occurred in connection with the development of the V/STOL (vertical and short take-off and landing) technology (Sayles and Chandler, 1971, p. 210).

Reputation

The games described so far are the public side of the implementation process. But the world of implementation games is populated by human beings with private as well as public lives. Their private games proceed simultaneously with their public games. Private needs and desires are, of course, manifold and diverse. We postulate, however, that the most significant private games are those that have to do with career aspirations and ambitions. And the crucial strategy in playing this game is to protect and augment one's reputation.

A reputation for what? That depends on the player and his personal situation. Presidents of the United States seek reputations for being sensitive to their power stakes (Neustadt, 1960). Reporters seek reputations for being "in the know." Policy analysts seek reputations for being sophisticated, "high-powered," and—let us now hope—for being able to avoid the common and uncommon "pitfalls of analysis." Highly placed policy makers, engaging in what Allison (1969) calls "bureaucratic politics," seek a reputation for being sound, discreet, and loyal, and for having been able to shape whatever consensus might have emerged.

Reputation games do not necessarily have more deleterious—or even

different—effects on implementation outcomes than other games. Because reputation is built largely on performance in relevant public games, any player's tactics in his private Reputation game are likely to be coincident with his tactics in the more public games. An exception to this rule, however, occurs when actors seek out means of persuading an audience that they are doing more or doing better than they really are, and more concretely still, when actors seek a reputation for boldness, toughness, or innovativeness. This sort of posturing can be extremely demoralizing to individuals doing more constructive but less visible work trying to implement a program (Gross *et al.*, 1971).

COPING WITH IMPLEMENTATION GAMES

We have identified, for heuristic purposes, over a dozen common implementation games. No doubt the list could be expanded, but these are trouble enough. Coping with them goes on in three somewhat different ways. First, there are strategies premised on field-level mediation and persuasion. Second, there are strategies premised on intermittent intervention by powerful agents who are spokesmen for the political legitimacy of the original policy mandate. Finally, these are preventive measures that can be taken at the policy design stage, principally by analysts and planners who have the foresight to design policies robust enough to withstand the buffeting of the implementation process.

FIELD-LEVEL MEDIATION AND PERSUASION

Organizational Development

An increasingly popular view among policy designers is that "if you can't beat 'em, join 'em." After all, people who "resist" your allegedly superior wisdom might be wise in their own way. They know their organization's strengths and weaknesses, its peer group alignments, its needs and priorities. Given the chance, and a little encouragement, they will successfully adapt whatever you have to offer to their own context and setting—in ways you could neither predict nor ordain. Such, at any rate, is the basic conclusion of a lengthy and very well documented study by researchers at the Rand Corporation of the impact in the classroom of federal programs supporting educational change and innovation (McLaughlin, 1976). Certainly there is much to be said for this point of view theoretically (Beckhard, 1969). It does run the risk, however, of legitimizing unacceptably large distortions of the goals enshrined in the policy mandate, particularly in policy areas in which (unlike education) we can count on considerably less consensus over basic program goals (e.g., penal institutions, unemployment insurance, worker safety regulation).

Negotiations

A very common means of attempting to reduce the delays, the misunderstandings, and the confusion attending the implementation process is bilateral or multilateral negotiations. Instead of simply maneuvering other parties at a distance through a whole succession of *faits accomplis*, people sit down and talk about what they might be able to do for each other. This is so commonplace, indeed, that we hardly notice it. We take it for granted that communication, persuasion, and face-to-face bargaining is more efficient than unilateralism and action without discussion.

On balance this is probably a sound assumption, supported by the broader principles of democracy as well as by the efficiency criterion. Yet we should note as well that the negotiating process has almost as much potential for creating problems as for solving them. If the process creates and sustains an atmosphere of mutual trust and cooperation, all will be well. If the reverse occurs, however, the outcome might be worse than if all parties took their chances on a less direct method of interaction (Crozier, 1964; Raven and Kruglanski, 1970; Schelling, 1963). Another problem with negotiations is that they can acquire a ritualistic autonomy. They become a substitute for constructive action rather than a prelude to it, as they serve important psychological functions for the participants, like relieving anxiety, symbolizing the "essential ambivalence" of all intimate relationships (Coser, 1956, p. 65), and releasing tension.

It is sometimes possible to reduce the cost of negotiations by using intermediaries. This is, of course, a common practice in the private sector, where there is a specialized profession skilled in "mediation and arbitration." By contrast, the use of mediation in the public sector is haphazard. One reason is that public sector negotiations are generally more multidimensional and political, and therefore lend themselves less readily to mediation; a second reason is that there is usually a scarcity of third parties who carry the requisite moral authority and who are willing to assume the politically unrewarding role of go-between (Bardach, 1977, p. 241).

USING THE POWER OF THE MANDATE

As we have seen, the implementation process is the continuation of "politics" by another name. Implementation politics have a somewhat different character from the politics of the policy-adoption process, however, for in the former case there exists an authoritative policy mandate that creates expectations—if nothing else—that "something" will be done. These expectations can be used to affect the "something" that eventually results.

Project Management

This concept comes from the history of large-scale government projects based on advanced technology, e.g., building the Polaris submarine fleet, sending men and satellites into outer space (Sapolsky, 1972; Sayles and Chandler, 1971). It represents a way of overcoming the limitations of the traditional division of organizational labor along "functional" lines (sales, production, research, for instance) when the organization must undertake a large, novel, and costly project requiring specialized and unpredictable inputs from several such departments and possibly from suppliers outside the organization. Although the idea of project management has come to be associated with particular technical paraphernalia like PERT charts and computerized information systems, the essence of the concept is in how it concentrates power and responsibility (Sapolsky, 1972, especially chapters 2, 3, 4, 8). In our terms, the project manager is the human expression of "the policy mandate." His authority rests on the presumption that he, and no one else, is its legitimate interpreter and spokesman.

There is some question about just how transferable project management might be from essentially technical projects to projects with a more "political" character, in which not only are the outputs laden with multiple and sometimes conflicting values but the inputs are acquired in political rather than economic markets. Under such circumstances project managers would tend to become just another set of players in the larger implementation game. They would have to play their own Budget, Territory, and Reputation games (especially the last).

The Political Fix

If project managers are too susceptible to being drawn into mere game playing, it is in part because they do not speak with enough authority. First, if their authority derives from the presumption that they are the sole legitimate interpreters of and spokesmen for the policy mandate, the presumption is highly rebuttable. After all, they are not the originators of the mandate, nor its political trustees, but mere agents, delegates. Secondly, whatever rights they exercise to distribute rewards and punishments (so as to induce cooperation in the collective enterprise), these rights are only derived from a higher source. Hence, while it is not easy to do so, it is possible to appeal from their verdicts to individuals even more powerful and authoritative—individuals in a position to play such a role I call "fixers."

Typically, fixers are influential legislators or top political appointees in the executive branch who were intimately associated with the adoption of the policy mandate. Such persons are ideal for playing the mediating role in the negotiating processes described above as well as for arbitrating controversies that result from the interplay of interests maneuvering against, and with, one

another out in the field. Moreover, fixers do not need to be exclusively reactive. They can, if they desire, play a "proactive" role as well: coaxing, bullying, troubleshooting, propagandizing, and so on.

The "fixer" does not and cannot act alone, particularly if he plays a proactive role. He needs staff who can handle the detail work, and he needs a widely dispersed network of informal "eyes and ears" out in the field that will register and report information on which he would wish to act (Bardach, 1974; Bardach, 1977, pp. 275–278; Kaufman, 1973).

The most severe limitation on "fixing" as a solution to the problems of the implementation process is that, for individuals with the requisite political resources, there are rarely enough incentives to do the necessary dirty work. It is behind-the-scenes work, for the most part, and legislators find this intrinsically unattractive as well as unrewarded either by constituents or by fellow legislators. Top political officials rarely stay in one spot long enough to master the details of the implementation process that need fixing, and in any case their own Reputation games are not particularly well served by investing much time and energy in "fixing" activities (Bardach, 1977, pp. 278–282; Hargrove, 1975, pp. 110–115; Levin, forthcoming).

PREVENTIVE STRATEGIES AT THE POLICY DESIGN STAGE

The policy analyst, and whoever else is involved in the design process, can take steps to design policies that have at least a fighting chance to survive the rigors of the implementation process. First, the basic social, economic, and political theory behind the policy must be reasonable and sophisticated: it will not do, for instance, to pretend that most people do not act most of the time in accord with a rather restricted notion of their self-interest; nor will it do to ignore inconvenient features of the world, like the sparse supply of managerial and technical competence or the enormous variety of local circumstances that policies must serve or the immense difficulty of coordinating large-scale activities on the basis of plans and promises rather than market signals (Devons, 1951; Goode, 1967; Wilson, 1967).

Second, a basic administrative strategy must be selected. Such a strategy should be simple. It should entail as little reliance on bureaucratic processes as possible, relying instead on actual or simulated markets, e.g., tax manipulations of existing price patterns (Kneese and Schultze, 1975; Levine, 1972; Pressman and Wildavsky, 1973). Bureaucracy cannot, of course, be avoided altogether; but its methods can be tailored as closely as possible to the task at hand. In particular, the heavy and usually uncritical reliance on audit-and-control mechanisms, extensive and detailed written procedures, and the like, should generally be avoided. In programs heavily dominated by professional service workers like doctors or teachers, indeed, such methods are likely to be irrelevant at best and destructive at worst (Bardach, 1977, pp. 109–141).

The third step is to make a list of requisite program elements and to note beside each one who—whether organizations, groups, or individuals—might be in a position to contribute them. With this list at hand, the fourth step is for the policy designer to go through the list of implementation games provided above (and elaborated in Bardach, 1977) or any other games that come to mind and ask: "Who, if anyone, is likely to play this game? With what effects? If the effects are unacceptable, is it possible to stimulate other actors whose games will neutralize these effects, either by preventing them from occurring or by compensating for those effects that are unavoidable?" For better or for worse, I do not believe there are any simple conceptual answers to these questions; and even if there were, it would require a great deal of effort and sophistication to derive their policy implications for the concrete problems at hand.

The fifth step is to think about how to establish some of the facilitative and fixing mechanisms described above, if appropriate, e.g., third-party mediators, project management, encouragement through agents of "organizational development," a fixer and his immediate staff, the fixer's network of "eyes and ears" in the field (see also Iglesias, 1976).

Finally, the policy designer should think about how to "phase in" a new policy so that it makes more friends than enemies, especially in its early and vulnerable months (or years, or weeks, as the case may be). More to the point, perhaps, it must make more influential friends than influential enemies; and the former must be led to mobilize still more support on behalf of the emerging policy or program while the latter must be combated, dispersed, and neutralized. At its highest and most comprehensive level, therefore, the political dynamics of the implementation process resemble the dynamics of the policy adoption process (Bardach, 1972, pp. 241–261), though in neither case are these dynamics well understood.

PITFALLS IN ANALYZING IMPLEMENTATION PROBLEMS

The single most important pitfall surrounding implementation analysis is the belief that it is not worth bothering about. Implementation issues sometimes seem to be of slight practical consequence compared to basic theoretical or political issues like "Do we really want to treat heroin use as a crime?" or "How much should we spend on exploring outer space?" or "Who should bear the financial burden of national health insurance?" Also, implementation issues tend to come up (when they do come up) toward the end of an analytical process, when professional and political investments in resolutions of more basic issues may already be in place and are very resistant to being disturbed by "mere" considerations of implementability. People who insist on raising these issues at this stage risk being called small-minded or being accused of

defeatism or of conspiring with the opposition. One way to avoid this pitfall is to make room for implementation analysis from the very earliest stages of the project. Although it probably helps to assign primary responsibility for implementation analysis to one individual, all participants in the project should become sensitized to the importance of implementation difficulties.

A related pitfall originates in the belief that implementation difficulties are usually minor, if they occur at all, and that, if not minor, they are in any event remediable. If this is so, then it follows that even justified skepticism about implementability should not be permitted to exclude a policy option that looks favorable on other grounds. Although this line of reasoning is sometimes sound, it more often leads to a treacherous pitfall. To a thoughtful implementation analyst, for instance, the wreck of the 1976 swine flu vaccination campaign in the United States was utterly predictable.

Implementation analysis is an exercise in concentrated pessimism. It is sometimes a psychologically difficult undertaking and is on that account shunned. There is no simple way to overcome this psychological obstacle, but a sense of humor, or better still a sense of irony, helps.

Although analytically it is often convenient to work with a model of a "closed system," in the real world of social, economic, and political action significant systems are almost always "open." Hence every policy "solution" potentially stimulates new and only vaguely predictable "problems" of its own, some of which merely impede the "solution" and some of which alter the essential character of the system itself. I have attempted to sketch some of the more common such problem-generating processes under the headings of Easy Money, Pork Barrel, Up for Grabs, Piling On, and Reputation, though there are surely many others as well. Precisely because these effects are new and unpredictable, analysts are tempted to ignore them.[5] Another reason analysts tend to ignore them is that it takes an active imagination coupled to a "dirty mind" to have a reasonable chance of foreseeing them. Analysts who, after due introspection, recognize that they are not thus equipped should find collaborators who are.

A fifth pitfall concerns the relationship of the policy adoption process to the implementation process. For numerous reasons, success or failure of the latter depends in part on how closely, and in what way, it is linked to the former. Too tight a linkage may have provided potential saboteurs with enough ammunition and allies to destroy the policy. Too loose a linkage may have put key implementers in an uncooperative frame of mind. In any case, once the policy is designed and authorized, it is too late to ponder the issue of whether and how to include or to exclude any implementers in the design process.

A final pitfall has to do with the methodology of implementation analysis itself. A certain amount of contemporary discussion of this topic prescribes overly formalistic and quantitative methods. It is sometimes recommended, for instance, that for all relevant implementers, coefficients that measure their

probable level of support, their degree of interaction, and so on be displayed. Similarly, there are recommendations to assess "the density of the interorganizational field," and hence "the need for coordination," by counting organizations and estimating their "interdependencies." Unfortunately, the key concepts in such a prescription are at worst meaningless and at best ambiguous. It should be recognized that the very nature of the implementation process—it is relatively unstructured, and its evaluation is very sensitive to the effects of errors in political judgment and defects in skill—requires that implementation analysis be handled as art rather than as science.

NOTES

1. Note that we do not say that the implementation process "translates the policy mandate into action" or "transforms prescriptions into results" or "realizes policy goals." All such locutions, common as they are, misleadingly suggest that words *become* deeds. Implementation is difficult, to be sure, but it need not depend on miracles.
2. Regulations and guidelines can be conceived as program elements that help to implement a higher-level policy mandate, like a statute. But it should be remembered that, like statutes, regulations and guidelines are only words. In keeping with our usage of the term "implementation," these words are merely stimuli to a subsequent, and conceptually distinguishable, implementation process or, more precisely, a subprocess within the larger process. To put the point another way, the implementation of regulations and guidelines can be just as problematic as the implementation of statutes or judicial decisions or any other set of words stating a policy mandate. (Cf. Rabinovitz *et al.*, 1976.)
3. The main exceptions, I believe, are (a) when rapid budgetary growth would destabilize existing intrabureau power and status relationships, (b) when the corollary of budget growth is greater political visibility and vulnerability to attack from outside, and (c) when the condition of accepting the larger budget increment is to undertake a difficult and unpleasant task.
4. No matter what economists say, people are often reluctant to treat sunk costs as irrelevant. In certain contexts this is not as irrational as it might appear, since writing off a sunk cost is an admission of failure that could carry independent penalties.
5. In another context Aaron Wildavsky has observed that most "new" policies these days represent attempts to straighten out problems created by "old" policies from the not-too-distant past.

REFERENCES

Ackerman, B. A., S. Rose-Ackerman, J. W. Sawyer, Jr., and D. W. Henderson. 1974. The Uncertain Search for Environmental Quality. New York: The Free Press.

Allison, G. T. 1969. Conceptual models and the Cuban missile crisis. American Political Science Review 63: 689.

Alterman, R., and M. Hill. 1974. Implementation of Plans. Draft manuscript. Haifa: Technion.

Banfield, E. 1961. Political Influence. New York: The Free Press.

Bardach, E. 1972. The Skill Factor in Politics: Repealing the Mental Commitment Laws in California. Berkeley: University of California Press.

Bardach, E. 1974. Subformal warning systems in the species *Homo politicus*. Policy Sciences 5: 415–431.

Bardach, E. 1977. The Implementation Game: What Happens After a Bill Becomes a Law. Cambridge, Massachusetts: MIT Press.

Beckhard, R. 1969. Organization Development: Strategies and Models. Reading, Massachusetts: Addison-Wesley.

Bernstein, M. 1955. Regulating Business by Independent Commission. Princeton, New Jersey: Princeton University Press.

Blaustein, A. I., and G. Faux. 1972. The Star-Spangled Hustle. Garden City, New York: Doubleday.

Coser, L. A. 1956. The Functions of Social Conflict. London: Collier-Macmillan.

Crozier, M. 1964. The Bureaucratic Phenomenon. Chicago: University of Chicago Press.

Devons, E. 1951. The problem of co-ordination in aircraft production, pp. 102–121. *In* Lessons of the British War Economy (D. N. Chester, ed.). Cambridge: Cambridge University Press.

Downs, A. 1967. Inside Bureaucracy. Boston: Little, Brown.

Edelman, M. 1964. The Symbolic Uses of Politics. Urbana: University of Illinois Press.

Elmore, R. F. 1976. Follow through planned variation, pp. 101–123. *In* Social Program Implementation. (W. Williams and R. F. Elmore, ed.). New York: Academic.

Goode, W. J. 1967. The protection of the inept. American Sociological Review 32(1): 5–19.

Greenstone, J. D., and P. E. Peterson. 1973. Race and Authority in Urban Politics. New York: Russell Sage Foundation.

Gross, N., J. B. Giancquinta, and M. Berstein. 1971. Implementing Organizational Innovations: A Sociological Analysis of Planned Educational Change. New York: Basic Books.

Hargrove, E. C. 1975. The Missing Link: The Study of the Implementation of Social Policy. Washington, D.C.: The Urban Institute.

Huysmans, J. H. B. M. 1970. The Implementation of Operations Research. New York: Wiley.

Iglesias, G. U. (ed.). 1976. Implementation: The Problem of Achieving Results (A Casebook on Asian Experiences). Manila: Eastern Regional Organization for Public Administration.

Jones, C. O. 1975. Clean Air: The Policies and Politics of Pollution Control. Pittsburgh: University of Pittsburgh Press.

Kaufman, H. 1973. Administrative Feedback. Washington, D.C.: The Brookings Institution.

Kneese, A. V., and C. L. Schultze. 1975. Pollution, Prices, and Public Policy. Washington, D.C.: The Brookings Institution.

Levin, M. A. Forthcoming. The Political Dilemmas of Social Policymaking.

Levine, R. A. 1972. Public Planning. New York: Basic Books.

MacAvoy, P. W. (ed.). 1970. The Crisis of the Regulatory Commission. New York: Norton.

Majone, G. 1976. Choice among policy instruments for pollution control. Policy Analysis 2: 589–613.

Marris, P. and M. Rein. 1967. The Dilemmas of Social Reform. New York: Atheneum.

Mayhew, L. 1968. Law and Equal Opportunity. Cambridge, Massachusetts: Harvard University Press.

Mayntz, R. 1976. Environmental policy conflicts: the case of the German Federal Republic. Policy Analysis 2: 577–587.

McLaughlin, M. 1976. Implementation as mutual adaptation: change in classroom organization, pp. 167–180. *In* Social Program Implementation (W. Williams and R. F. Elmore, ed.). New York: Academic.

Neustadt, R. E. 1960. Presidential Power. New York: Wiley.

Orfield, G. 1969. The Reconstruction of Southern Education: The Schools and the 1964 Civil Rights Act. New York: Wiley.

Pressman, J. L., and A. Wildavsky. 1973. Implementation. Berkeley: University of California Press.

Rabinovitz, F., J. Pressman, and M. Rein. 1976. Guidelines: a plethora of forms, authors, and functions. Policy Sciences 7: 399–416.

Radin, B. A. 1977. Implementation, Change, and the Federal Bureaucracy: School Desegregation Policy in HEW. New York: Teachers College Press.

Raven, B. H., and Kruglanski, A. W. 1970. Conflict and power, pp. 103–104. *In* The Structure of Conflict (P. Swingle, ed.). New York: Academic.

Rocheleau, B. 1976. Implementing quality control in the human services, or Why is implementation usually unsuccessful? Paper delivered at the 1976 annual meeting of the Midwest Political Science Association, Chicago, Illinois.

Sapolsky, H. M. 1972. The Polaris System Development: Bureaucratic and Programmatic Success in Government. Cambridge, Massachusetts: Harvard University Press.

Sayles, L. E., and M. K. Chandler. 1971. Managing Large Systems. New York: Harper and Row.

Schelling, T. C. 1963. The Strategy of Conflict. Cambridge, Massachusetts: Harvard University Press.

Stigler, G. J. 1971. The theory of economic regulation. Bell Journal of Economics and Management Sciences 2: 3–21.

Vertinsky, I. Investigating Implementation: a Multi-Paradigm Approach. Draft manuscript.

Williams, W. 1975. Implementation analysis and assessment. Policy Analysis 1: 531–566.

Wilson, J. Q. 1967. The bureaucracy problem. The Public Interest 7: 3–9.

Zeigler, H. 1961. The Politics of Small Business. Washington, D.C.: Public Affairs Press.

Zwerling, S. 1974. Mass Transit and the Politics of Technology: A Study of BART and the San Francisco Bay Area. New York: Praeger.

10 Pitfalls of Evaluation

Harry P. Hatry

Program evaluation is a term that has been commonly used in the United States to refer to an assessment of a government program's past or current performance, usually to determine the impact, or effectivness, of that program.

Program evaluation should be considered a vital element of systems analysis when public officials are attempting to determine whether a current or new program (or a variation of a program) should be continued, modified, expanded, cut back, or terminated. The term "program" evaluation has not typically been used in the private sector. Private firms, however, routinely evaluate past performance by examining sales, profits, and share of the market, and many of the procedures sometimes used in program evaluation, such as various forms of experimentation, are used by the private sector, for example, in marketing research. This chapter, however, focuses on program evaluation of government-sponsored programs.

The key issues in program evaluation are (a) to attempt to determine what the effects (impacts) of the program have been over a specific period of time and (b) to determine the extent to which the specific program, rather than other factors, has caused the effects. Both issues are typically subject to considerable uncertainty, particularly in the great majority of government programs that are not undertaken under controlled laboratory conditions. Program effects are often unclear, are often ill defined, and can be quite messy to measure.

Dealing with the second issue—to what extent the effects identified can be attributed to the specific program—also presents considerable difficulties, because outcomes can be affected by numerous other factors in addition to the program itself, and the effects of these external factors will generally be difficult to determine without careful analysis.

Throughout this chapter the term "program" will be used to encompass both relatively narrow activities and broad-scale major programs. Examples com-

mon to most nations and for which program evaluation principles apply are programs for use in such basic governmental services as police and fire protection, solid waste management, the delivery of physical and mental health services, provision of transportation services, water supply and treatment, parks and recreation, library services, employment and training programs, and education.

Evaluations may be of ongoing programs—programs that have been in place for some time and have been operating at full scale. Or evaluations may be of trials of new programs (or program variations)—perhaps where the scope of that trial is limited to a portion of the ultimate target population.

Many of the pitfalls discussed in earlier chapters also apply to the form of analysis called program evaluation. Most of these do not need to be dealt with again here, though a few will be, where they are particularly applicable to program evaluation.

The intent of this chapter is not to discourage those who want to undertake program evaluation, but rather to encourage evaluators to make their work as meaningful as possible, and to keep both their own and their customers' expectations within reasonable bounds. As with systems analysis in general, it is seldom that evaluators will be able to provide conclusive findings about the impacts of current programs. And, as with systems analysis in general, it is almost never possible for the evaluators to say that a particular program will continue to have the same consequences.

TYPES OF PROGRAM EVALUATIONS

In the past few years there have been numerous publications about various aspects of program evaluation. This chapter is not the place to describe in depth the content of program evaluation. We hope that most readers will at least be somewhat familiar with the concepts of program evaluation.[1] Probably the form of evaluation most written about is the "controlled experiment," the Cadillac (or Mercedes–Benz) of the evaluation world. It is also, however, the form of evaluation that is probably the least often undertaken because of its special requirements. In the controlled experiment, each member of the target population (or a random sample) is randomly assigned to either a group that receives the treatment or to other groups that receive no treatment or some alternative treatment. Each unit in the population of interest should, to the extent possible, have the same probability as all other units in that population of being assigned to each group.[2] The program intervention (treatment) is then applied for a period of time, and the impacts on the groups are compared to determine the extent to which the groups differed. The random assignment is intended to ensure that the groups have similar characteristics at the beginning of the experiment.

There are less demanding forms of program evaluation. These, in general, provide less conclusive evidence that any impacts observed were actually caused by the program intervention. These include such variations as (a) merely looking at the values of the evaluation criteria before the intervention is made and then looking at them again some time thereafter to see what changes have occurred; (b) deriving a historical time series based on several years of past data, extrapolating the trend into the future, and comparing that extrapolation to what actually happens after treatment (this controls for certain types of underlying long-term conditions); (c) using "natural" comparison groups when there is a group that did not get the treatment that is sufficiently similar to the treated group that different outcomes are probably due to the intervention. Evaluations that are not controlled experiments are likely to be used considerably more often, because of their less demanding requirements.[3]

Some of the pitfalls presented below apply more to certain of these types of evaluations than to others. This will be indicated where appropriate.

PITFALLS OF PROGRAM EVALUATION

The pitfalls briefly discussed below are arranged according to the time at which the pitfall generally occurs: before the beginning of actual data collection for the evaluation, during the process of data collection, or after the data have been collected (when findings are to be presented and use is to be made of the findings). Numerous categories and groupings of pitfalls could be used. What follows is merely one person's perspective.

PITFALLS BEFORE DATA COLLECTION

PITFALL 1 *Not doing an assessment of the potential utility and evaluability of evaluation candidates to assure that it is likely that the program can be evaluated in sufficient time for the evaluation findings to be useful and within available resources.*

Not assessing the potential utility and evaluability of programs may be one of the major causes of frustration in program evaluation in the United States (see, e.g., Horst *et al.*, 1974). Programs probably should not be the subject of substantial program evaluation effort when (a) the program has vague objectives; (b) even though the objectives are reasonably clear, the current state of the art in measurement does not seem to permit meaningful measurement of impacts, or (c) the program's major impacts cannot be expected to show up until many years into the future, by which time the information is not likely to be useful (when, for example, it seems likely that, by the time the evaluation is completed, all relevant important decisions will already have been made and

whatever is found cannot be acted on). An illustrative list of criteria for assessing candidates for evaluation is presented in Table 1.

In many instances, evaluability problems can be alleviated. The evaluability assessment should be careful not to overreact to apparent hurdles. For example, with persistence it is often possible to identify major specific objectives for

TABLE 1 Criteria for Selecting Issues for Program Evaluations

Can the results of the evaluation influence decisions regarding the program?

- Programs for which a decision regarding continuation, modification, or termination need to be made are obvious candidates.
- Poor candidates are those where decision-makers have strong preconceptions of program value or where there is considerable support by influential vested interest groups—if these circumstances make it very unlikely that the program would be altered, regardless of evaluation findings. (However, in some cases the program may be of such great importance to a community that government officials may proceed with an evaluation and be prepared to seek changes despite political risks.)
- Can the evaluation be completed in time to be helpful to decision-makers? Evaluations which are completed after public officials become committed to a decision are likely to be useless.

Can the evaluation be done?

- Is the particular innovation sufficiently defined, including both what the innovation is and its purposes? To what extent are there clear linkages between the innovation and its desired impacts? Note that it is not necessary that everything about the project be spelled out completely in advance, but only that they be sufficiently specific so that evaluators could determine whether a project actually had been implemented and have enough information to identify adequate evaluation criteria.
- Are sufficiently reliable and valid data obtainable on important effects of the program?
- Can sufficient resources be obtained to meet the time schedule and technical requirements of the evaluation?
- Has the program been stable enough so that an evaluation will provide relevant information? If the program is constantly changing, or is about to change, in significant ways, it is not a good candidate for evaluation.

Is the program significant enough to merit the evaluation effort?

- Programs which involve a large amount of resources or those programs which have important benefits and possible important negative consequences to the public, should be given higher evaluation priority. Thus, the likely cost of the evaluation should be compared to the possible decreased cost or improved effectiveness that could result from evaluation findings.
- Is the program suspected by the government of being marginal in its performance? If so, there may be opportunities for making major improvements or cost reductions.
  Is it a new program whose potential benefits and costs are highly uncertain?

SOURCE: Matry *et al.* (1973).

programs that seem vague at first; it is often possible to obtain rough but adequate impact information about characteristics that at first glance appear to be too subjective (such as by using structured interviewing of systematic samples of clients on various aspects of program services); and often even evaluations conducted over prolonged periods may be useful for decisions in later years even though they are not useful to the current government administration.

Proper evaluation requires adequate staff, money, and time, and the evaluation plan clearly needs to be compatible with the resources available. However, though some corner-cutting and less sophisticated approaches can often be used when resources are scarce, too many such compromises can weaken an evaluation to the point where it is not worth doing.

Seldom discussed in the literature is the need to distinguish whether the program to be evaluated is under development or is operational: Evaluations of projects in a developmental stage, in general, seek less definite information on impacts and are likely to be more concerned with determining the characteristics of the preferred program and its basic feasibility. Ignoring this distinction appears to have resulted in inappropriate expectations and inappropriate evaluation designs in some instances, especially for U.S. federal agency evaluations. Technology-oriented programs are particularly likely to go through a formal developmental stage. (Perhaps planning for "human service" projects should also include both a development and operational test stage, though this prolongs the period before an innovation is introduced into operation.)

PITFALL 2 *Starting data collection for the evaluation too early, e.g., prior to shakedown of the program, and not allowing enough time to assess a stable operation.*

This pitfall is encountered particularly in the evaluation of new programs. There seems to be a chronic temptation to begin collecting data for evaluation as soon as the initial attempt at implementation begins. For many new or revised programs, however, the shakedown period may last many months, during which program procedures stabilize, new people become adjusted to the new procedures, and the new program begins to operate under reasonably normal conditions. Thus, enough time must be allowed before beginning collection of the post-program data and enough time thereafter for an adequate test of the new program. Evaluation periods of less than a year, in addition to often not providing enough program experience, can also be affected by seasonal differences; for example, tests of a new street repair procedure might not uncover the special effects of the bad weather season. The appropriate timing will, of course, depend on the nature of the program and the setting into which it is introduced. To illustrate a likely typical timing: A minimum period of perhaps 6 months might be appropriate before the program is assumed to have been im-

plemented, and at least 1 year of subsequent program operation should be covered by the evaluation.

PITFALL 3 *Not planning the evaluation so as to try to help program managers improve the program in addition to merely measuring its impact.*

Some evaluators may not agree that this is a pitfall. Trying to avoid it will probably require extra effort that could dilute the impact evaluation. It appears, however, to this author that one of the major obstacles to the undertaking and use of evaluation is that often too little attention is paid to identifying constructive ways in which programs could be improved. Unfortunately, this is easier said than done. There are usually a large number of factors, in addition to the program procedures, that can affect program success. These include such things as the quantity and quality of the staffing used to operate the program, the success in motivating the employees who will implement it, and the organizational structure in which the new program operates. Evaluators often are in a particularly good position to obtain insights into reasons for problems. If the evaluators can take advantage of this and act as a constructive force for program improvement, over the long run the credibility and utility of the evaluation field will increase, and perhaps the innate hostility of program managers to being evaluated will be diminished.

PITFALL 4 *Expecting too much from evaluations and not, from the beginning, making clear to potential users the likely limitations.*

Evaluations of any kind seldom give definitive, conclusive, unambiguous evidence of program success (or failure). Even with experimental designs, as will be noted later (see, especially, Pitfalls 5 and 14), there are inevitably numerous problems in keeping the experiment uncontaminated and, subsequently, in extrapolating and generalizing the results beyond the experimental scope and time period. The evaluators should be careful to make it clear to their customers and to potential users of the evaluation findings that such limitations exist. Unrealistic expectations by customers can cause rapid disillusionment with the immediate evaluation, which may discourage future evaluation support.

PITFALLS DURING EVALUATION

PITFALL 5 *Not knowing what the intervention has really been . . . what has really been implemented.*

The best laid plans of evaluators can come a cropper when they are placed in real-life settings. Particularly for long experimental periods, the chances of deviation from the original intentions increase. For example, in evaluations of neighborhood police teams, the assignment of teams to specific neighborhoods may depart from the plan if dispatchers assign those police officers too

frequently to other neighborhoods. In some cases, if the program planners and evaluators watch carefully, such deviations can be corrected, but in other situations this may not be possible. Another example involves the difficulties of maintaining random assignment procedures in an experiment when assignments have to be made throughout the period by personnel other than the evaluation team. In an experiment to test the effects of requiring appearances for moving traffic violations before a judge, the court clerk who had responsibility for the random assignments had firm ideas about the need for a court appearance for young drivers and did not adhere to the random assignment procedure (Conner, 1977; other examples are also provided in this paper). Random assignments of clients in controlled experiments may initially be done appropriately but may subsequently, under the pressure of heavy workload, be altered.

Because of such potential problems it is important that the evaluators carefully monitor the program over the period of the evaluation. At the least, they should check periodically to ascertain that there have been no major departures from the plan during implementation. When substantial deviations occur, adjustments should be made (such as, in effect, beginning a "new" evaluation if a major overhaul of the program occurs *during* the evaluation). If such adjustments cannot be made satisfactorily, and if the changes are of major importance, the evaluation should be terminated or at least the alterations should be explicitly considered when assessing the findings.

PITFALL 6 *Use of inadequate measures of program effects.*

The credibility and usefulness of an evaluation can be called into considerable doubt if inadequate measures are used. Variations of this pitfall include limiting the assessment to only one criterion or a very few criteria when others are also relevant (perhaps because a decision is made to evaluate only those criteria agreed on ahead of time with program officials), and neglecting possible unintended, sometimes beneficial and sometimes detrimental, consequences of the program. For example, an evaluation of a program for placing mental patients in their own homes and communities rather than in large government institutions should consider not only changes in the condition of the clients but also the effects on clients' families and the community into which the clients are transferred.

This criterion problem is a well-known one and applies to all systems analysis; it already has been discussed in an earlier chapter. Its great importance, however, requires it to be singled out in a discussion of evaluation as well. Before establishing the final evaluation criteria, evaluators should review the objectives of the program from the viewpoint of the agency installing the program and of the clients of the program and should look for significant effects that were not initially anticipated. The danger that evaluations will be overly affected by the evaluator's own preferences (see, e.g., Shaver and

Staines, 1972) suggests that evaluators should strive to obtain various perspectives on the objectives (both explicit and implicit) of the program to be evaluated. For example, opinions might be sought from supporters and opponents, program operators and clients, budget officials and program managers. (This assumes that evaluators will, in general, have sufficient leeway from the sponsors of the evaluation to try to be comprehensive.)

An important variation of this pitfall is failure to assess the impact of the program on the various major client groups. Inevitably, programs have differing effects on various groups, helping some groups significantly more than others and, perhaps, harming other groups. Aggregating all groups together will hide such differences and prevent users of the evaluation findings from considering equity issues. Evaluators should look for, and report on, any significant differential impacts on the various population groups affected by the program.

Finally, the lack of an assessment of program financial costs can also be an important omission. Evaluations often neglect costs, but such information can be of considerable use to public officials.

PITFALL 7 *Failure to identify, and adjust for, changes in data collection procedures that occur between pre- and post-program data collection.*

This problem typically occurs when the evaluation depends on data that had been collected before the evaluation began or when it depends on the use of existing government records (rather than when both pre- and post-program data are collected especially for the evaluation). Government data definitions and data collection procedures change periodically, and such changes can cause important differences in the meaning of such data. Evaluators using data for which they themselves have not determined the data collection procedures should be careful to look for, and adjust for (if they can), such occurrences. As has been noted (Riecken and Boruch, 1974), "Too often a new program is accompanied by changes in record-keeping."

PITFALL 8 *Letting those whose work is being evaluated, in effect, evaluate themselves.*

This is a well-known pitfall. Nevertheless, there are often temptations to ignore it. Government agencies with tight resources (especially subnational governments such as state and local governments in the United States) are particularly tempted. It is desirable for any agency as a matter of good management to undertake some internal evaluation of its own programs. For example, mental health and social service agencies frequently use caseworkers' ratings to assess the progress of the caseworkers' clients. This procedure is reasonable when the information is solely for internal purposes, such as for use by the caseworkers themselves and their immediate supervisors. Such procedures, however, do not provide data on client improvement after clients have left the

programs (to determine the longer-term effects of the services), and such procedures seem to be expecting too much of human nature (asking employees to provide objective information that will be used to make judgments about continuation of their own programs).

PITFALL 9 *Not looking for other plausible explanations for the outcomes.*

After outcomes have been measured, the key dilemma in program evaluation is to determine the extent to which the outcomes can be attributed to the program itself. When controlled-experiment designs are used, this problem is substantially reduced, though not fully eliminated. For noncontrolled-experiment design, this problem is of major importance. Inappropriate attribution is a major pitfall for nonexperimental evaluation designs. A number of variations of this pitfall have been the subject of frequent discussion in publications on program evaluation[4]:

- Initial conditions may be so extreme that, even without program intervention, a natural shift toward improvement (or deterioration) can be expected even without the new program (this is sometimes called "regression toward the mean"). Clients who at the outset have the greatest need (e.g., are in crisis situations) are likely to show a greater amount of improvement than others. Conversely, the less needy, most able clients, will tend to show little, no, or negative improvement, regardless of the program (Conner, 1977). For example, a program initiated because of a very recent rash of problems [e.g., a high traffic fatality rate (Campbell and Ross, 1968)] might show an improvement merely because the chances of recurrence are small.
- Similarly, in some cases there might be a normal "maturation," in which the group acted on would improve normally even without program intervention, perhaps because of aging. For example, as criminals, alcoholics, or drug addicts age, there often appears to be a reduction in their adverse behavior even without treatment programs. Another example: An evaluation of community alcoholic treatment centers included a follow-up 18 months after intake of a sample of persons who had an intake record but for whom only nominal treatment was provided (Armor *et al.*, 1976). Of this group, 54 percent were identified as being in remission. Neglecting dropouts from programs aimed at helping clients can result in overestimating the results of the programs.
- The lack of a comparison group, or the use of an inappropriate comparison group, can distort the interpretation of the findings. Even when controlled experiments are not utilized, examining groups that were not the object of the program can often be helpful in providing evidence about whether the outcomes were due to the program. Thus, in the classic evaluation of the Connecticut highway speeding crackdown, the large reduction in fatalities in Connecticut was compared with other nearby, and presumably similar, states to see if similar reductions had occurred (Campbell and Ross, 1968). Such

a comparison helped rule out some other possible causes of reduced fatalities in Connecticut, such as special weather conditions in the region during the period or the introduction of safer automobiles.

An example of a situation where the effects of a program would be understated if no comparison group were considered occurred in a study of the treatment of arthritis (Boruch, 1976). The study indicated that the severity of arthritis *increased* after the introduction of a treatment program. However, in this case there was a control group. The experiment showed that the control group's conditions worsened even more than the treated group, indicating that the treatment program was a success rather than the failure which might otherwise have been reported.

The example cited earlier, the evaluation of community alcoholism treatment centers, followed up not only those who received significant amounts of treatment but samples of two comparison groups: (a) persons that had made only one visit to a treatment center and who received no further treatment and (b) clients who had received minimal services (usually detoxification) but who then left the center and never resumed contact (Armor *et al.*, 1976). The evaluators identified 67 percent of all treated clients as being in remission at the time of an 18-month follow-up. But, as noted above, even the nominal-treatment group showed a 54 percent remission rate. Thus, it appears possible that a substantial portion of the remissions in the treatment group might have occurred with only nominal help. Considering only the treatment group remissions might lead one to overstate the effects of the treatment.

When comparison groups are used, however, they should be used with considerable care to avoid misinterpretations. If the programs' clients differ from the comparison group in some potentially critical characteristic (e.g., persons entering the program may have considerably greater motivation than those with apparently similar problems who did not enter the program), differences in impact could be due to that characteristic and not to the program.[5] In using comparison groups (even in controlled experiments), it is important to check the comparison groups for similarity of key characteristics, where this is possible. Unfortunately, adequate data are often not available for such comparison groups. (A further discussion of problems with comparison groups is given in Edwards *et al.*, 1975.)

- "Hawthorne effects," where receivers of the service (or program personnel) become aware of the evaluation and act in special ways, can make the outcomes of the test unrepresentative and nongeneralizable. This applies to experimental as well as nonexperimental designs.[6]
- Similarly, program personnel who are handpicked to operate the new program can make the outcomes of the test unrepresentative and nongeneralizable. Using specially chosen personnel may be appropriate in the developmental stages of a program, but it is to be avoided when the program is to be evaluated for its generalizability. For a representative test, personnel of

the type that would ordinarily be operating the program after the evaluation period should be used. Otherwise, any observed advantage to the treatment group might be due to the use of the special personnel.

- Not explicitly considering, and controlling for, "workload/client difficulty" when assessing the results can lead to misinterpretation of what has occurred. The difficulty of the incoming workload can cause success (or failure) rates to be misleading. Higher apparent success rates for programs that had a mix of clients who were easier to help compared to programs with a more difficult workload should not be necessarily labeled as being more effective. Consider the following hypothetical outcomes (Millar *et al.*, 1977):

	Total Cases	Number and Percentage of Cases in Which Dependency Was Reduced
High client difficulty	400	100 (25%)
Low client difficulty	100	100 (100%)
Total, sample 1	500	200 (40%)
High client difficulty	100	0 (0%)
Low client difficulty	400	300 (75%)
Total, sample 2	500	300 (60%)

Based on the totals alone, the results for the second sample appear superior because success was achieved in 60 percent of the cases as contrasted with 40 percent for the first sample. Sample 1, however, shows a higher success rate for *both* "high-difficulty" clients (25 percent, compared to 0 percent for sample 2) and for "low-difficulty" clients (100 percent, compared to 75 percent for sample 2). The overall higher success rate for the second sample stems from its having a large proportion of clients with low difficulty.

Thus the difficulty of the incoming workload can, on occasion, be a major explanation for observed effects. In controlled experiments, even if workload difficulty is not explicitly controlled in making random assignments (such as by stratifying the sample), randomization would likely result in assigning similar proportions to each of the groups. Nevertheless, the control and treated groups should be examined after they are chosen to determine if they are indeed sufficiently similar on "difficulty."

- When applying statistical analysis, correlations can be mistakenly assumed to represent causality. For example, a regression analysis might indicate that lower performance is positively correlated with the amount of tutoring received. Concluding that tutoring has a negative effect on performance would, of course, overlook the possibility that those requiring the most

tutoring would likely be substantially poorer performers to begin with; their subsequent performance, though higher than before, might still not reach the levels of brighter pupils.

- In general, not explicitly looking for other plausible explanations for observed differences in effects—for a wide variety of possible circumstances or factors can cause unrepresentative and misleading findings—is a pitfall. Another example: Changes in employment status of persons given training programs might occur because of improved general economic conditions, even in the absence of the training programs (Boruch, 1976). (If control groups are used, the evaluation should be able to detect advantages of the training program that occur despite these changing external conditions.)

PITFALL 10 *Various pitfalls in undertaking sample surveys.*

A frequently used procedure in evaluations is to survey samples of program clients or comparison groups to obtain needed feedback about certain aspects of the program. Because of its importance as a program evaluation procedure, some of its major pitfalls are briefly listed below.[7]

Unclear, misleading, or biased wording of the questions asked of respondents.

Poor interviewers or interviewing procedures.

Lack of pretesting of the question wording and the order of questions, followed by correction of any difficulties found during the pretest.

Inadequate sample selection, such as sampling in a way that omits or undersamples a key group (such as persons without telephones, those with unlisted telephone numbers, those living in trailers, or those living in multiple dwelling units when the list from which the sample is drawn includes only individual structures and not the individual units in each structure).

Inadequate number of callbacks, so that those persons not generally at home are not heard from. A variation of this is not providing for interviews at various times throughout the week, so that persons that tend not to be home during certain hours (for example, during the day) are systematically excluded from the sample.

Too many persons in the sample who cannot be located or too many refusals, leading to low completion rates. (This is particularly a problem when mail surveys are used, especially those without provision for follow-up of nonrespondents—but it applies to all modes of interviewing. Mail return percentages are typically about 25 percent or less; in such cases, even though the *number* of mail returns may be quite large, the results cannot be said to be representative of the population and therefore are not reliable for deriving generalizations about the population.)

PITFALL 11 *Collecting too much data and not allowing adequate time for analysis of the data collected.*

These two problems go hand in hand. They are all too prevalent when tight timetables exist for evaluations—as usually seems to be the case. The temptation seems to be prevalent to collect data on any characteristic of the client or situation that conceivably could be relevant, and then not allow enough time for analysis of the data. The temptation to collect data is a difficult one to overcome, particularly since in evaluations it is often difficult at the beginning to know which data will be of importance in the study. The argument is often advanced that you can always exclude data later. However, once collected, data pile up and this has a pyramiding effect in terms of data processing and analysis effort (as well as adding to the costs of data collection).

A particularly good place for restraint is the interview questionnaire. It seems common that the questionnaire becomes longer and longer as those on the evaluation team think more and more about what information should be collected. Suggestions for questions should have to be justified by identifying the use to which the answers to each question would be put and the likelihood that the information obtained will shed light on relevant hypotheses. Allowing enough time for data analysis is complicated by the tendency to impose overly tight deadlines for evaluations. When implementation difficulties delay the start of the program, when data comes in later than anticipated, and when computer processing is later than promised, these all lead to squeezing the amount of time available for analysis before the deadline for the evaluation. Better scheduling to allow for unforeseen contingencies and the inclusion of fewer data to be processed will help alleviate these problems.

PITFALLS AFTER DATA COLLECTION AND INITIAL ANALYSIS

PITFALL 12 *Overdependence on statistical significance, particularly at the expense of practical significance.*

Too narrow a focus on too much precision and too much reliance on statistical significance can lead to excessive costs in resource allocation (such as by encouraging the use of larger samples than needed at the expense of other evaluation tasks) and to misleading findings. Statistical significance levels at the 95–99 percent significance levels will often be overkill for programs other than those that involve important safety or health elements. Typically, the information involved in evaluations and other factors involved in making program decisions is not precise, and for most government decisions precision is not needed. "It doesn't pay to lavish time and money on being extremely precise in one feature if this is out of proportion with the exactness of the rest" (Herzog, 1959). Even if lower significance levels are used (90 percent and even 80 percent), the use of statistical significance as the only criterion for detecting differences can be misleading to officials using the information. What may be a statistically significant difference (at a given significance level) can, particularly where very large samples are involved, suggest that there are important

differences, although the differences may be small in practical terms and may be unimportant to most public officials. With large sample sizes, differences of even two or three percentage points between the outcomes of the treatment and comparison groups can be statistically significant, but they may often not be significant to officials making decisions based on that information.

A good practice is to present both the actual differences and the level of statistical significance, so that users of the information can judge for themselves. All too often, summaries of findings indicate whether findings are statistically significant without indicating the actual amount of the differences.

PITFALL 13 *Focusing only on the overall (average) results and not considering the component effects.*

Examination of the aggregate data is, of course, useful for assessing the program's aggregate effect. However, in general, the analysis should not be limited to the aggregate effects. It may often be more useful to examine subsets of the data. For example, when a number of different projects are included in the evaluation, the evaluators should consider whether certain projects or groups of projects tended to have greater effects than others. Variations in conditions among projects are likely, and an examination may be able to shed light on possible reasons for variations, possibly suggesting variations that should be considered further even though the overall program does not appear successful. As another example, some types of clients served by the program may be more (or less) successfully served than others, even though such differences were not anticipated in the original evaluation design. Therefore, in general, the evaluators should examine various subgroups to detect whether there may be groups that were substantially better (or worse) served than indicated by the aggregate figures. For example, a particular type of program may work well with more severe cases than with less severe cases, or with older clients than with younger clients, or with female clients rather than male clients. Sometimes, samples to be followed up in the evaluation may be stratified at the beginning to ensure adequate consideration of different characteristics. If this is not done, an after-the-fact analysis of outcomes for various types of client would need to be undertaken.

PITFALL 14 *Overly generalizing (extrapolating) the results beyond the evaluation.*

Even when the evaluation is well done and very well controlled, there are numerous pitfalls in trying to generalize results to future, possibly more widespread, usage. "Too many social scientists expect single experiments to settle issues once and for all" (Campbell, 1969). "The particular sample from which control and experimental group members are drawn . . . may be idiosyncratic in that other potential target populations are not represented. If the conditions . . . in the experiment . . . differ markedly from conditions which prevail

in other populations, then it is reasonable to believe that additional testing of the program is required . . ." (Riecken and Boruch, 1974). There are several variations of this pitfall; recognizing them should temper statements about the generalizability of findings:

- The trial's results may in effect represent only one sample point—that is, one trial under one set of conditions. Replication may be needed in other sites, and at other times, before one can state with confidence the general effectiveness of the program. Of course, to the extent that the initial trial covers a variety of sites and the evaluation is of a program that covers the entire target population, this will be of less concern. Often, however, there will be limitations on the size and coverage of the trial. Not all locations, not all potential client groups, not all other potentially important conditions are likely to be covered. Such limitations of the evaluation should be clearly stated in the findings. An example: The New Jersey Graduated Work Incentive experiment examined only one type of geographical location—the U.S. urban East Coast—it covered only male-headed households, and it varied only the level of income guaranteed and the tax rate (Roos, 1975). The findings' applicability to other conditions would need to be judged accordingly. Another example: If a test of a new street-patching material happens to be undertaken during a year with an unusually low amount of rain, the validity of the findings would be in question for periods of normal rainfall.
- A special variation of this overgeneralizing pitfall can occur when explicit or implicit statements are made about the particular characteristics of the intervention that caused the observed impacts. This problem arises particularly where only one site (and one set of program intervention characteristics) is used in the trial of the program. As discussed under Pitfall 5, it is vital that evaluators know what was actually implemented and that they be alert for features of the trial that appear to be significant in the program's apparent success or lack of it, even though they were not initially intended to be tested during the trial. For example, in evaluations of the effectiveness of social service casework, such characteristics as the particular technique used, the caseworker's personality, the amount of time spent with the client, and the caseworker's style could all affect the outcomes (Fischer, 1976). Unless the evaluation procedures in some way attempted to isolate these characteristics in the test, the evaluators would be unable to generalize about the extent to which these characteristics affect the outcomes. They would not be able to state whether, for example, apparent successes (or failures) were the result of the techniques used or the caseworkers' style and personality. This might be less of a problem if a large number of sites and many different caseworkers were involved in the test. Otherwise there would be substantial ambiguity about what was driving the observed outcomes and what should be done about the program. The conclusion might be reached that casework is (or is not) effec-

tive, whereas what was actually evaluated was only one combination of casework and caseworker characteristics.

- Behavior may change when the novelty of a new program wears off (for either program operators or their clients). And client behavior may alter from that in trials undertaken on only part of the population when the program is established so that everyone can receive the program. For example, a program to determine the effects of the use of group homes rather than large institutions for caring for children with juvenile delinquency records might be tested in one or two locations. The findings might not be representative of other settings if the program were expanded—for example, citizens might become antagonistic to a large number of homes in their community. (This is especially likely to happen if the locations are chosen because of the communities' willingness to test the group homes.)
- Some groups may turn out not to have been covered by the test. In some instances, this may have been part of the test plan. In others it may be unintentional. The evaluators should determine which types of clients were included and which were not. They should avoid attributing observed effects to those not covered in the test unless a logical case can be made for it. Many tests will not be able to cover all the major target groups that were initially intended for coverage and that are intended to be covered by the program after it goes into full-scale operation. If this is found to be the case, the findings should be qualified. The New Jersey Graduated Work Incentive experiment, as noted earlier, was limited to male-headed households and those located in only one geographical location.

PITFALL 15 *Not permitting program personnel to review the findings before finalization and promulgation.*

Permitting program personnel to review the findings before promulgation is generally a matter of courtesy and good practice. It also, however, has an important technical purpose: it will provide a review of the data and the findings from a different perspective. This practice appears to be regularly followed by audit agencies in the United States, but it seems to be less common in special evaluations undertaken by administrative agencies. Program people may be aware of situations and factors that the evaluators have missed, and they can often add considerable insight into the interpretations of the data, sometimes identifying misinterpretations and misunderstandings by the evaluators. Even when program managers are defensive and hostile, they may offer comments that will indicate that the evaluators have indeed made misinterpretations or even errors that should be corrected. In one evaluation in which the author was involved, drug treatment program personnel reviewing the draft report pointed out to the evaluation team that an important group of program clients had been left out of the evaluation—requiring the evaluators to follow up what would otherwise have been a neglected group of clients. Finally, the opportuni-

ty to suggest modifications may reduce defensiveness by program personnel, thereby encouraging subsequent implementation of recommendations based on the evaluation findings.

PITFALL 16 *Not indicating the assumptions, uncertainties, and other limitations of the evaluation when presenting the findings.*

For example, limitations of the evaluation criteria used, especially those that may have been identified too late during the process of evaluation, and limitations of the measuring instruments should be made clear. The amount of uncertainty in the findings should be identified not only when statistical analysis is used, but in other instances as well. Information about the magnitude of uncertainty should be presented, even if only in the form of the evaluators' subjective judgments.

PITFALL 17 *Not separating facts from opinion.*

In presenting the findings of an evaluation, whether orally or in writing, evaluators should be careful to distinguish those elements that are opinions from those that are based on objective findings and objective information. (This pitfall, of course, applies to any form of systems analysis, including program evaluation.) It is easy for users of evaluation reports to be misled about the nature of the supporting evidence for findings, especially when a series of findings is provided.

This caveat also applies to recommendations: the basis of each recommendation, indicating whether it is based on strong evidence or not, should be identified. Also, when evaluators attempt to provide insights into why programs are not as effective as they might be and then provide recommendations to improve the program, there is a tendency not to distinguish those recommendations which follow from the major technical examination from those recommendations that have emerged from the more subjective, qualitative insights the evaluators obtained during the technical evaluation. Under Pitfall 3, providing insights into reasons for program problems was encouraged. Preferably, such insights would be obtained through special technical analyses. However, even where these are obtained through more qualitative means, it is important that the basis of the evidence be clearly presented.

PITFALL 18 *Poor presentation of findings, so that potential users cannot readily understand them.*

As discussed in the chapter on communications, this tends to be a chronic problem with analysts. Program evaluation findings, whether presented orally or in writing, should be clear, concise, and intelligible to the mix of users for whom the report is intended. This should not, however, be used as an excuse for not providing adequate technical backup (documentation) for findings. The technical evidence should be made available in writing, either in the body of the

text, in appendixes, or in a separate volume, so that technical staffs of the various report users and other reviewers can have the opportunity to examine for themselves the technical basis of the findings.

ACKNOWLEDGMENTS

The author wishes to thank Carol H. Weiss of Columbia University and Sumner Clarren, Joseph Nay, and John Waller of the Urban Institute for their suggestions.

NOTES

1. For those who are not. references such as the following are suggested to obtain details on program evaluation: Campbell and Stanley (1966), Glaser (1973), Hatry *et al.* (1973), Herzog (1959), Riecken and Boruch (1974), Suchman (1967), Weiss (1972a, 1972b), Wholey *et al.* (1970).
2. The groups might be groups of people, groups of locations, or even groups of equipment (for example, when a new type of vehicle-mounted device is being compared against the existing device). In practice, pure random assignment often is not possible nor appropriate and has to be approximated. (See, for example, Riecken and Boruch, 1974).
3. Two illustrations of noncontrolled experiment evaluations are the evaluation of impacts of community alcoholism treatment centers (David J. Armor *et al.*, *Alcoholism and Treatment* (California: The Rand Corporation, June 1976) and the evaluation of automatic vehicle monitoring equipment systems for police use (Richard C. Larson *et al.*, "Evaluating a Police-Implemented AVM System: The St. Louis Experience (Phase I)" *IEEE Transactions on Vehicular Technology*, February 1977, pp. 60–70. The first evaluation illustrates what can be done for a "human service" oriented program; the second for a technology-oriented innovation.
4. For example, discussions of such circumstances (sometimes referred to as "threats to validity") are included in Campbell (1969), Campbell and Stanley (1966), Bernstein *et al.* (1975), Nunnally and Wilson (1975) and Fischer (1976). Not all of the threats to validity listed in these references have been included here. The list that follows has been selected to highlight those believed most relevant for a discussion of "pitfalls."
5. Programs whose procedures are such that the clients determine for themselves whether they will enter a program (and no provision is made for random assignment from this particular group) raise the question of whether the added motivation is an important characteristic in determining program success; highly motivated persons might be able to achieve more progress than others even without the program. Volunteers for experiments tend to be more innovative or self-confident than the average eligible participant. See Hamrin (1974).
6. However, this particular problem does not apply to some types of evaluations. As Roos (1975) points out, if the evaluation is *retrospective*, not requiring new data to be collected, the trial cannot be affected by the evaluation effort.
7. There are numerous references that discuss survey procedures in greater detail. For example, see Abramson (1974), Earl R. Babbie (1973), Campbell *et al.* (1969), Duckworth (1973), Levy (1972), Survey Research Center (1976), Webb and Hatry (1973), and Weiss and Hatry (1971). In addition, the *Public Opinion Quarterly* frequently contains articles on the problems of surveys. The U.S. National Center for Health Statistics Series 2 reports (on "data

evaluation and methods research") provide occasional studies on the problems of survey research, among them Gleeson (1972).

REFERENCES AND SELECTED BIBLIOGRAPHY

Abramson, J. H. 1974. Survey Methods in Community Medicine. New York: Longman.

Armor, D. J., J. M. Polich, and H. B. Stambul. Alcoholism and Treatment. Santa Monica, California: The Rand Corporation.

Babbie, E. R. 1973. Survey Research Methods. Belmont, California: Wadsworth.

Bernstein, I. N., G. W. Bohrnstedt, and E. F. Borgatta. 1975. External validity and evaluation research, a codification of problems. Sociological Methods and Research 4: 101–128.

Boruch, R. F. 1976. On common contentions about randomized field experiments. Evaluation Studies Review Annual 1: 158–194.

Campbell, D. T. 1969. Reforms as experiments. American Psychologist 24: 409–429.

Campbell, D. T., and H. L. Ross. 1968. The Connecticut crackdown on speeding. Law and Society Review 8: 33–53.

Campbell, D. T., and J. C. Stanley. 1960. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally.

Campbell, J. R., *et al.* 1969. Household Survey Manual, 1969. Washington, D.C.: Executive Office of the President, Bureau of the Budget.

Conner, R. F. 1977. Selecting a control group: an analysis of the randomization process in twelve social reform programs. Evaluation Quarterly 1: 195–244.

Duckworth, P. A. 1973. Construction of Questionnaires, Technical Study. Washington, D.C.: U.S. Civil Service Commission.

Edwards, W., M. Guttentag, and K. Snapper. 1975. A decision-theoretic approach to evaluation research, pp. 139–181. *In* Handbook of Evaluation Research, Volume 1 (E. L. Struening and M. Guttentag, ed.). Beverly Hills, California: Sage.

Fischer, J. 1976. The Effectiveness of Social Casework. Springfield, Illinois: Charles C. Thomas.

Glaser, D. 1973. Routinizing Evaluation: Getting Feedback on Effectiveness of Crime and Delinquency Programs. Bethesda, Maryland: National Institute of Mental Health, Center for Studies of Crime and Delinquency.

Gleeson, G. A. 1972. Interviewing methods in the Health Interview Survey. Vital and Health Statistics 2: 1–86.

Hamrin, R. D. 1974. OEO's performance-contracting project: evaluation bias in a social experiment. Public Policy 22: 467–488.

Hatry, H. P., R. E. Winnie, and D. M. Fisk. 1973. Practical Program Evaluation for State and Local Government Officials. Washington, D.C.: The Urban Institute.

Herzog, E. 1959. Some Guide Lines for Evaluative Research. Washington, D.C.: U.S. Department of Health, Education, and Welfare, Welfare Administration, Children's Bureau.

Horst, P., J. N. Nay, and J. W. Scanlon. 1974. Program management and the federal evaluator. Public Administration Review.

Larson, R. C., K. W. Colton, and G. C. Larson. 1977. Evaluating a police-implemented AVM system: the St. Louis experience (Phase I). IEEE Transactions on Vehicular Technology VT-26: 60–70.

Lenihan, K. J., and F. M. Casey. 1977. Unlocking the Second Gate. R & D Monograph 45. Washington, D.C.: U.S. Department of Labor and Employment and Training Administration.

Levy, C. V. 1972. A Primer for Community Research. California: Far West Research, Inc.

Millar, A. P., M. P. Koss, and H. P. Hatry. 1977. Monitoring the Outcomes of Social Services. Volume 1: Preliminary Suggestions. Washington, D.C.: The Urban Institute.

Nay, J. N., J. D. Waller, and J. W. Scanlon. 1976. Evaluation: The Acid Test. Washington, D.C.: The Urban Institute.

Nunnally, J. C., and W. H. Wilson. 1975. Method and theory for developing measures in evaluation research, pp. 227–232. *In* Handbook of Evaluation Research, Volume I (E. L. Struening and M. Guttentag (ed.). Beverly Hills, California: Sage.

Riecken, H. W., and R. F. Boruch, (ed.). 1974. Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York: Academic.

Roos, N. P. 1975. Contrasting social experimentation with retrospective evaluation: a health care perspective. Public Policy 23: 241–257.

Suchman, E. A. 1967. Evaluative Research: Principles and Practice in Public Service and Social Action Programs. New York: Russell Sage Foundation.

Shaver, P., and G. Staines. 1972. Problems facing Campbell's "experimenting society." Urban Affairs Quarterly 7: 173–186.

Struening, E. L., and M. Guttentag, (ed.). 1975. Handbook of Evaluation Research, Volume 1. California: Sage.

Survey Research Center. 1976. Interviewer's Manual. Revised edition. Ann Arbor, Michigan: Institute for Social Research, The University of Michigan.

Webb, K., and H. P. Hatry. 1973. Obtaining Citizen Feedback: The Application of Citizen Surveys to Local Governments. Washington, D.C.: The Urban Institute.

Weiss, C. H. 1972. Evaluating Action Programs: Readings in Social Action and Education. Boston: Allyn and Bacon.

Weiss, C. H. 1972. Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, New Jersey: Prentice-Hall.

Weiss, C. H., and H. P. Hatry. 1971. An Introduction to Sample Surveys for Government Managers. Washington, D.C.: The Urban Institute.

Wholey, J. S., J. W. Scanlon, H. G. Duffy, J. S. Fukomoto, L. M. Vogt. 1970. Federal Evaluation Policy. Washington, D. C.: The Urban Institute.

11 The Pitfalls of Language, or Analysis through the Looking-Glass

Kathleen A. Archibald

He who does not expect the unexpected will not detect it.
HERACLITUS, B 18.*

Majone reminds us in his introductory paper that policy analysis deals largely with intellectual constructs rather than physical phenomena. We start from words, we work with words, we report in words—or to be more exact, in words and other symbols.

Pitfalls in verbal communication are, in Hogben's view,

due to [the] inadequacy of many words we use to fulfill their ostensible function as signals, i.e. as meaningful symbols. To speak here of symbols, uttered or written, as meaningful, signifies that the recipient (listener or reader) can decode the message of the transmitter (speaker or writer) in terms consistent with the intention of the latter, a proceeding which would entail no hazards if each of the words we use had only one meaning in a particular context. This is probably true of no natural language.[1]

Policy analysts use a natural language with bits and pieces from a variety of more specialized languages added. This addition of concepts and technical terms from a variety of disciplines is intended to avoid some of the pitfalls Hogben refers to by introducing words that have a single, specified meaning in a particular context, that is, *univocality in context*. This does not always work, however, even when careful definitions of the terms introduced are provided, because the meaning of a word often depends on more than the immediate context in which it appears. Words have long roots. When transplanted from the neat, well-tended garden of one discipline to the communal plot of policy analysis, large chunks of the root system are likely to be left behind.

These problems are exacerbated when work must be translated from one natural language into another. There are few translators familiar with the polyglot nomenclature of policy analysis. The first time an article of mine was translated into a language foreign to me, I was sent a copy of the translation to

* As translated from Diels-Kranz, *Fragmente der Vorsokratiker*, I, 1960, by Karl R. Popper.

check. I then had to find another translator to translate the Spanish words back into English—like the White King in *Through the Looking-Glass*, who needed two messengers, one to fetch and one to carry.[2] The English words my translator fetched for me from the Spanish conveyed at many points meanings my original words were never meant to carry. As I remember, Lindblom's "disjointed incrementalism" underwent the most peculiar transformation. (On top of this, or perhaps on its bottom since it is a pitfall, the journal published the piece before receiving the corrections we so laboriously worked out.) I now circumvent such pitfalls in transmission by assuming it's the recipient's problem.

To avoid pitfalls when I myself am the recipient of translated work, I mostly pray. A dictionary is not of much help. It is difficult to know you are trapped in a bad translation while you are trapped. Only when a good translator comes along and sets you free can you see the pitfall. The most stunning example of this I have come across is in Shapiro's translation of Habermas' *Knowledge and Human Interests*. In a translator's note Shapiro explains:

> The official translation of Freud's concepts *das Ich* and *das Es* as the ego and the id was a serious mistake that both reflects and has contributed to the scientistic self-misunderstanding of metapsychology. *Das Ich* means the I and *das Es* the it. That is, they refer to the antithesis between reflexive, personal subjectivity and reified, impersonal objectivity. Freud's famous statement of the goal of psychoanalysis "Wo Es war, soll Ich werden," should read in English, "Where it was, I shall become," or perhaps, "Where it-ness was, I-ness shall come into being." The choice of scientific Latin to render these terms in English reifies the I into an object by making it into an "ego," which is not the word used to express reflexivity, self-consciousness, and agency in English. As an "ego," the I is already an it, and the qualitative distinction between I and it is not visible in comparing the terms ego and id. At the same time, this scientistic terminology obscures the connection between the models of reflection of psychoanalysis and German Idealism. All this notwithstanding, the weight of recent tradition has made it seem advisable to retain "ego" and "id" in this translation.[3]

"The weight of recent tradition" that leaves us with ego and id weighs heavily on policy analysis too. Quade has written:

> Because the term "policy analysis" does not imply anything about quality, rigor, or comprehensiveness, but merely the purpose and the context of the work, it may seem desirable to adopt a new name for the type of analysis that incorporates the improvements we call for. But we have too many qualifying terms already and another would simply add confusion. Hence, we continue to use the term policy analysis.[4]

The qualifying terms that have been put in front of analysis include systems, systematic, and applied systems. "Analysis" remains the one constant symbol in the name. This is the one we are stuck with by weight of recent tradition. The adjectives come and go and multiply, but analysis carries on. There have been contenders—the systems *approach*, policy *sciences*—but they have not caught on. "Analysis" has staying power, and that's a useful quality when it comes to names, but is it an appropriate name?

WHAT'S IN A NAME?

It was on her journey by train through Queen's 3d (*Through the Looking-Glass* is, among other things, a chess problem) that Alice first met the Gnat. On the train it is the size one would expect a gnat to be, but when the train jumps a brook and Alice finds herself sitting quietly under a tree, she is being fanned by the wings of "a *very* large Gnat." Not one hundred feet tall like Carlos Castaneda's Gnat,[5] but big: " 'About the size of a chicken' Alice thought." She didn't feel nervous with it because they had been talking together on the train. The Gnat, however, is a little nervous about Alice's attitudes:

"—then you don't like *all* insects?" the Gnat went on, as quietly as if nothing had happened.

"I like them when they can talk," Alice said. "None of them ever talk, where *I* come from."

"What sort of insects do you rejoice in, where *you* come from?" the Gnat inquired.

"I don't *rejoice* in insects at all," Alice explained, "because I'm rather afraid of them—at least the large kinds. But I can tell you the names of some of them."

"Of course they answer to their names?" the Gnat remarked carelessly.

"I never knew them do it."

"What's the use of their having names," the Gnat said, "if they wo'n't answer to them?"

"No use to *them*," said Alice; "but it's useful to the people that name them, I suppose. If not, why do things have names at all?"

"Ca'n't say," the Gnat replied.[6]

Since Analysis does not answer to her name, it is reasonable to ask what it is about her name that might be useful to the people who call her that. Our opinion of what "analysis" is, of what the word may mean, is formed from all our past encounters with the word. If we have not met the word before, or have not been properly introduced, we may turn to a dictionary. *Webster's New Collegiate* gives seven meanings for "analysis," the first two of which are generally applicable.

1. separation of a whole into its component parts; 2a: an examination of a complex, its elements, and their relations; b: a statement of such an analysis.

Other definitions apply to specialized fields.

"Systems analysis" is included in the dictionary, and "systems analyst" is even included as a synonymous cross-reference under "analyst"—the field has come of age! Here's how the craft is introduced to a collegiate newcomer:

the act, process or profession of studying an activity (as a procedure, a business, or a physiological function) *typically by mathematical means* in order to define its goals or purposes and to discover operations and procedures for accomplishing them most efficiently. [Emphasis added.]

The "*typically* by mathematical means" is a characterization that Rand-style systems analysts have been trying to establish some distance from, at least since the early 1960s. Quade, in *Analysis for Military Decisions*, distinguished

systems analysis from operations research by pointing out that the broader problems it dealt with were not susceptible to the precise formalism and full quantification that operations research was striving for as a hopefully nascent branch of mathematics. In systems analysis, economics had become almost indispensable, and other more qualitative social sciences were making significant contributions.[7]

In the 1968 *Systems Analysis and Policy Planning*, systems analysis was loosened up another notch. Quade explained that it is broader than "the application of quantitative economic analysis and scientific methods to such matters as weapon design and the determination of force composition and deployment." He went on to say systems analysis is not "a method," not "a set of techniques," and not "a type of report." It is

> *a research strategy, a perspective* on the proper use of the available tools, *a practical philosophy* of how best to aid a decisionmaker with complex problems of choice under uncertainty. In the absence of a good brief definition, *systems analysis*, as the term is intended to be understood in this book, can be characterized as a *systematic* approach to helping a decisionmaker choose a course of action by investigating *his full problem, searching out* objectives and alternatives, and comparing them in the light of their consequences, using an *appropriate* framework—*in so far as possible analytic*—to bring *expert judgment and intuition* to bear on the problem.[8]

Quade's most recent book, *Analysis for Public Decisions* (1975), is about policy analysis, a term that is likely to be used, he says, rather than systems analysis "when political and social factors predominate."[9] Policy analysis is introduced as one among several functional substitutes for the market mechanism: what "the market does for business in areas where competition is relatively free",[10] elections, press criticism, scholarly critics, public scandals, and analysis can do for government. Thus the company Analysis keeps appears to change with her first name. Also the way she is defined.

We have a *looser, ana*logical introduction now—as is meet and proper in introducing ANA (from Greek *'ανα-*, meaning up, back, or again) LYSIS (from Greek *λύειν*, to loose, set free)[11]—and a wider definition:

> In a broad sense *policy analysis* may be defined as any type of *analysis* that generates and presents information in such a way as *to improve the basis* for policy-makers to exercise their *judgment*. . . . In *policy analysis*, the word *analysis* is used in its most general sense; *it implies the use of intuition and judgment and encompasses* not only the examination of policy by decomposition into its components but also the *design and synthesis* of new alternatives.[12]

Now this *is* a loosening up. Analysis implies the use of intuition and encompasses design and synthesis. Analysis is often thought to be related to the antonyms of intuition: information, elaboration, induction, experience.[13] And synthesis is an out-and-out antonym for analysis, in its everyday sense—"the composition or combination of parts or elements so as to form a whole"

(*Webster's*)—and in its philosophical sense. Quade here has analysis become so capacious that it now encompasses its antonym, synthesis.

Majone takes a tack similar to Quade's. He has analysis as a weak sister to science, suffering "from the same limitations . . . without sharing [its] strengths." He goes on to suggest that "the references to science that figure so prominently, for instance, in the official definitions of operations research are not so much methodological indications as they are ideological props." Their latent function is "to increase the collective confidence of a group of new disciplines striving for academic and social recognition."[14]

"Don't stand chattering to yourself like that," Operations Research said, looking at her for the first time, "but tell me your name and your business."

"My *name* is Analysis, but ———"

"It's a stupid name enough!" Operations Research interrupted impatiently. "What does it mean?"

"*Must* a name mean something?" Analysis asked doubtfully.

"Of course it must," Operations Research said with a short laugh: "*my* name means the shape I am—and a good scientific shape it is, too. With a name like yours, you might be any shape almost."[15]

Majone, in suggesting why problems of implementation still give us so much difficulty, points to "the excessive *reliance of many analysts* on normative decision models and the pure logic of choice, *on analysis* rather than *synthesis and design*."[16] If the reliance of analysts on analysis is a fallacy, then the pitfalls of analysis perhaps include the pitfalls of Analysis, the name.

"But 'glory' doesn't mean 'a nice knock-down argument'," Alice objected.

"When *I* use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean—neither more nor less."

"The question is," said Alice, "whether you *can* make words mean so many different things."

"The question is," said Humpty Dumpty, "which is to be master—that's all."[17]

Humpty Dumpty's question can be taken in several ways. The question is turned into "a known fact by linguists" by one commentator on Carroll. "When new words or usages are introduced into a language, their continued use often depends on the status of the individual or group introducing them. One *can* use words as he likes, but if his usage is to be of significance, it must be accepted by others. Consequently, he must, in his linguistic usage, be the master."[18] But Humpty Dumpty said "*which* is to be master" not *who*, suggesting the issue is whether words shall be the master or the persons using them.

Pitfalls in language derive from a lack of mastery in this latter sense. In calling our craft Analysis, have we put ourselves on a pitfall-strewn path?

What is meant by "analysis" has undeniably changed over the 30-odd years since Rand started doing something called "systems analysis." The obeisance to mathematics and to science has decreased considerably. But honor is still paid to being as analytic, as systematic, and as objective as possible. And the embracing of intuition has a strangely ambivalent quality, considering that many in the field are presumably familiar with the feats of intuition in mathematics, logic, and science.[19]

"INTUITION" AS AN EXAMPLE

Quade mentions intuition frequently and defines analysis, in *Analysis for Public Decisions*, as implying intuition. His treatment of it, however, is under "Alternative Sources of Advice." It is brief and ambivalent.

> One [alternative] is pure intuition, with or without divine guidance. The intuitive process is to learn everything possible about the problem, to live with it, and let the subconscious provide the solution. This approach can in no sense be considered analytic because *no effort is made to structure the problem* or to establish cause-and-effect relationships and operate on them to arrive at a solution. But one may use help; read tea leaves or inspect the entrails of sheep as the Romans did, or even support intuition with analysis. Later we shall see that analysis *must* be supported by intuition.
>
> Between pure intuition on the one hand and policy analysis on the other, there are intermediate sources of advice that are at least partially analytic, although the activities involved are ordinarily less systematic, explicit, and quantitative.[20]

This is a friend of intuition speaking, but one might not suspect it without being told.

What is characteristic of intuition is that there is no *evident* step-by-step process of reasoning involved (the conscious, systematic steps come later, to test the soundness of the intuition and to put it in communicable form). There may not even be symbols involved in the critical flash of insight; Einstein said his theory of relativity came to him first as a kinesthetic feeling. To say that "no effort is made to structure the problem" is a little off. Intuition often provides the idea of how to structure a problem that does not fit standard molds and only after considerable time has been spent struggling with structuring.[21]

My intent is not to criticize Quade's treatment *per se*; he gives more attention to intuition than most authors. I wish rather to indicate the awkwardness and ambivalence surrounding the concept of intuition and to suggest that this is perhaps to be expected among persons who identify themselves as "analysts" and rejoice in the systematic, the explicit, the quantitative, and the objective.

A craftsman who rejoiced in intuition might well cultivate devices that from an analytic perspective look a lot like pitfalls. Charles Lutwidge Dodgson, the Oxford mathematician and logician who told the Alice stories, originally

extempore to Alice Liddell, could be said to have fallen into a pitfall when, "in a desperate attempt to strike out some new line of fairy-lore, I had sent my heroine straight down a rabbit-hole, to begin with, without the least idea what was to happen afterwards."[22] Dodgson, or Lewis Carroll, put himself in a situation where he had to run, or rave, on pure intuition.

The rabbit-hole as an extemporaneous device is a pitfall if and only if you want to plan ahead systematically and thus avoid saying whatever comes into your head. If you want to stay close to pure intuition, on the other hand, find a figurative rabbit-hole and jump in. Find a scenario that permits you to expect only the unexpected.

One of the interesting things about the Alice tales, for those unfamiliar with them, is that they show Carroll to have been a better mathematician and logician than Dodgson.[23] Thus they show the strengths of pure intuition or active imagination.

I am not saying that policy analysts are against intuition; it is part of the common wisdom that we are dependent on intuition. Hitch has said:

> *It is utterly wrong to look upon intuition and analysis ... as rivals or alternatives.* Properly used, they complement each other. We have seen that every systems analysis is shot through with intuition and judgment.

But Hitch introduced his discussion of intuition, as did Quade, by referring to it as *an alternative*: "Concentrating on the second word 'analysis,' the alternative is, I suppose, intuition."

The problem is, I suggest, one of language. Analysis and intuition are in one sense opposites. To be analytic is to not be intuitive. People who call themselves designers recognize intuition and analysis as different phases in the design process. People who call themselves analysts, to be consistent, must consider all of what they do as analysis. So intuition and design come to be considered as aspects of analysis. Semantic confusion follows: in the space of three pages, Hitch says intuition is and is not an alternative to analysis, is a complement of analysis, and "is a species of logical analysis."[24]

This ambivalence and ambiguity surrounding intuition is not peculiar to systems analysis or policy analysis. It is shared by other fields that value both systematic reasoning and creativity. Popper has said of science:

> Intuition undoubtedly plays a great part in the life of a scientist, just as it does in the life of a poet. It leads him to his discoveries. But it may also lead him to his failures. And *it always remains his private affair*, as it were. Science does not ask how he has got his ideas, it is only interested in arguments that can be tested by everybody.[25]

To the extent that our focus on analysis puts intuition out of focus, blurs our appreciation of it, and makes it a "private affair," we can expect to have more difficulties with the creative aspects of our work—problem-setting and program design, for instance—than might otherwise be the case. A major difference between those who call themselves designers and those who call

themselves analysts is that the former do not relegate intuition solely to the realm of the private. They are interested in devices, like Carroll's rabbit-hole, that invite a group of problem-solvers to exercise their intuition jointly. The connections underlying an intuition may be nonevident; nevertheless, persons working together in the intuitive mode often stimulate each other.[26]

Hitch concludes the piece from which I have quoted by saying that systems analysis provides a framework that permits the intuitive judgments of experts in many fields to be combined.[27] This is all to the good as long as we do not confuse this framework with those devices that facilitate intuition, devices that often purposely turn away from the actual problem being considered.

LANGUAGE CREATION AND PROBLEM SOLVING

Anatol Holt has argued that "language creation" and problem-solving are indissoluble activities" and that this indissolubility hampers the adaptive capacity of advanced technological societies. His statements were made at a conference organized by Gregory Bateson on "The Effects of Conscious Purpose on Human Adaptation."

"We have depended through history," [Holt said], "on a beautifully subtle method, you might say, of semantic and formal drift in our symbolic apparatus. . . . The semantics gradually slip, old meanings decay [and new ones grow] and the formal apparatus—only very gradually and sufficiently slowly—changes. Well, now of course, the investments in certain symbolic structures have become enormous; our symbolic methods have become embodied in enormous technological complexes."

"Therefore, hard-programmed," Gregory [Bateson] put in.

"Therefore very difficult to change."[28]

On the heels of this interchange, Bateson explains his use of the term "consciousness":

that which is easily communicable between persons, *especially by language*, and faces certain filtering operations or something of the kind, which cut out many attributes of [the sea of mind]. It was that subtraction that I felt needed discussion.[29]

The author of the volume, the linguist Mary Catherine Bateson, then inserts a segment from her father's earlier writing to clarify further his view of consciousness.

A peculiar sociological phenomenon has arisen in the last one hundred years, which perhaps threatens to isolate conscious purpose from many corrective processes that might come out of less conscious parts of the mind. The social scene is nowadays characterized by the existence of a large number of self-maximizing entities, which, in law, have something like the status of "persons"—trusts, companies, political parties, unions, commercial and financial agencies, nations, and the like. In biological fact, these entities are precisely *not* persons and are not even aggregates of whole persons. They are aggregates of *parts* of persons. When Mr. Smith enters the boardroom of his company, he is expected to limit his thinking narrowly to the specific purposes of the

company or to those of that part of the company which he "represents." Mercifully it is not entirely possible for him to do this, and some company decisions are influenced by considerations that spring from older and wiser parts of the mind. But ideally, Mr. Smith is expected to act as a pure, uncorrected consciousness—a dehumanized creature.[30]

I return now to Holt's idea of hard-programming of significant parts of our symbolic apparatus via their embodiment in interlocking technological complexes. Science and technology, and large formal organizations as well, thrive on symbols that are univocal—"the reduction of meaning to a single meaning."[31] I would go further and say that our culture, what may be called Western industrial culture, favors univocality. It is, more often than not, set up as the ideal we should strive for in our communications.

A common response to a disagreement entangled with the use of words is "It's only semantics." *What* is? And why is "semantics" qualified by *only*? The presupposition underlying this is that there exists one meaning you are both focusing on and if you could but agree on a suitable name for this one meaning, then there would be no disagreement about *it*. Meaning, in this view, is a simple thing that exists "out there." Not only can be expressed univocally, but it is *best* expressed univocally. Or as Hogben put it in the quotation I opened with, from a chapter entitled "Semantics in Overalls" from his *Mother Tongue*, here given in extended form:

The keyword of the title of this chapter comes from a Greek one which signifies a *sign*, whence *signal* as in *semaphore*. Today, it most usually refers to the *recognition of pitfalls* of verbal communication *due to [the] inadequacy of many words* we use to fulfill their ostensible function as signals, i.e. as meaningful symbols. To speak here of symbols, uttered or written, as meaningful, signifies that the recipient (listener or reader) can decode the message of the transmitter (speaker or writer) in terms consistent with the intention of the latter, a proceeding which would entail no hazards if each of the words we use had only one meaning in a particular context. This is probably true of no natural language.[32]

As a craft that uses words, we need *some* "semantics in overalls," that is, a certain set of signals with univocal meanings. The title of this volume, however, is not phrased in such a working language. I have demonstrated that "analysis" is far from univocal. It not only has a plurality of meanings (*plurivocality*) so diverse as to include its opposites; it is also defined *analogically*. This puts "analysis" way out of the league of univocality.

"Pitfalls" was evidently out there to start with. Univocal meaning is literal meaning and this volume is surely not about literal pitfalls—"pit[s] flimsily covered or camouflaged and used to capture and hold animals or men." A pitfall of analysis is a figurative pitfall—"a hidden or not easily recognized danger or difficulty."[33] So we are, in this volume, considering a metaphor of a plurivocal analogy.

We have "of" still to consider. "Of" is one of those words that is so plurivocal that out of context it does no more than specify some connection.

Webster's New Collegiate Dictionary gives 12 meanings for "of"; the *Oxford English Dictionary* gives 17 main uses and 63 senses, most with subsenses, going on for 6 pages. We also learn from the *OED* that the primary sense of "of" (from Old English, as is 'pitfall') was *away, away from.* All current uses of "of" are derivatives.

If we take univocality as our ideal for *all* meaningful verbal communication, we would be forced to say that there are a number of pitfalls of verbal communication in any phrase that includes the word "analysis." I have used the topic of this volume, "pitfalls of analysis," as an example of such a phrase to indicate that words that ostensibly serve to qualify and thus limit "analysis" (for example, this volume is not about all aspects of analysis; it is limited to the pitfalls of analysis and related issues), may do just the opposite. To talk of the pitfalls of analysis, *if* "analysis" is construed to mean that which the linguistic opinion leaders of the Anglo-American speech community (e.g., *Webster's, OED*) say it may mean, is to talk about many things that "analysis" is not. It is to talk about synthesis and design and about rhetoric and implementation.[34] Thus "analysis" qualified by "pitfalls of" seems to become broader than "analysis" unqualified.

THE SYMBOLIC METHOD AND THE INTERPRETIVE DISCIPLINES

If we believe that any lack of univocality in our working language is a pitfall to be avoided, we had best retire Analysis and adopt a new name. There are crafts, however, that value plurivocality, and perhaps policy analysis could learn from them.

Mary Catherine Bateson mentions poetry and psychotherapy:

> Poets and clinicians bring to their listening a particular kind of attention [Theodore Reik calls it "listening with the third ear" in a book of that title], sensitive not only to the single logical path between ideas, but to the multiple bonds between symbols. Unconsciously we all swim in a sea of contexts. . . . Underlying all our conversation of logic and abstraction, each symbol or word used is fitted by each person into multiple contexts deeper than any argument.[35]

There is a strong symbolist tradition within theology, a field not quite as distant from policy analysis as one might expect. Augustine was within this tradition. Taking policy license with his words:

> What more liberal and fruitful provision could [our legislators] have made in regard to [their pronouncements] than that the same words might be understood in several senses, all of which are sanctioned by the concurring testimony of other passages equally [enlightened]?[36]

Luther, on the other hand, did not appreciate plurivocality. He believed that Scripture had a "single, simple, solid and stable meaning":

> I acknowledge that [legislation] is a most rich and inexhaustible fountain of wisdom;

but I deny that its fertility consists in the various meanings which any man at his pleasure may assign. Let us know that the true meaning of [legislation] is the natural and obvious meaning, and let us embrace and abide by it resolutely.[37]

What can the symbolist approach, which values words with "excess" meaning, with plurivocality and long roots, do that the univocal approach cannot do? Quite simply, it facilitates the making of new connections and associations, some of which may break the barrier that systematic rationality often puts in the way of creativity.

Majone has, fortunately for my purposes, introduced legal argumentation as a mode analogous to policy analysis in interesting and useful ways.[38] Unger, in a masterful critique of the "liberal tradition" in Western thought, groups "the reasonings of lawyers and judges" with those of theologians and grammarians as comprising "the ancient humanistic arts," all of which incorporate "pretheoretical experience into theory itself."

"Compare with this," he writes,

the approach of linguistics, or of the sociology of religion or law, when they adopt the modern view of science. . . . They want to become the science of language, of religion, or of law in general, rather than the doctrines of a particular language, religion, or legal order. As empirical sciences, they treat views about language, religion and law prevalent in specific societies as part of what has to be explained instead of as part of the explanation. The greater the independence a science achieves from the pretheoretical views of any one group, the more universal and objective it supposedly becomes.

Another characteristic of these humanistic arts or crafts "is their indifference to the contrast of description and evaluation." They are "both descriptive and prescriptive." In adjudicative practice, Unger points out, "every understanding of how the law stands on a given point depends on a conception of its purposes, on a view of what the law ought to accomplish." The understandings of the theologian and the grammarian are similar in accepting and developing a "unity of fact and value." Compare this approach with "the empirical sciences of language, religion, and law, [which] for their part, pride themselves on the strictness with which they hold the line between [explanation] and evaluation."

The third characteristic these "interpretive disciplines" share is that they must deal with the kind of ambiguity of meaning the earlier (and later) sections of this paper demonstrate.

[The ancient humanistic arts or] interpretive disciplines must also deal with the ambiguity of meaning peculiar to the phenomena of consciousness. Every act of speech, or religious ritual or belief, or rule of law has a meaning or a purpose given to it by the intention of the original speaker, holy man, or lawmaker. Yet this meaning must always be rediscovered by an interpreter, who has his own purposes and his own form of existence. What then guarantees that interpreted and interpreter can communicate? The resolution of the ambiguity of meaning is possible to the extent that interpreter and interpreted participate in the same community or tradition of shared beliefs, beliefs that are both understandings and values. Thus, for example, the judge defines the purpose

of the laws he applies by a combination of his view of the intentions of legislators or earlier judges and his sense of the demands of his own social situation at the time the case is decided. This combination can be achieved only because the judge is able to see himself as a collaborator of the lawmakers who preceded him, and as one whose ideas and values are not starkly at variance with theirs. Similarly, the philologist, faced with a text from an earlier age, can interpret it symbolically if he is able to bear in mind the literary tradition that leads back from his own perspective to the author's.

Unger points out there is a "wonderful symmetry" in the contrast between empirical sciences and the interpretive disciplines. "The former distinguish clearly the dimension of their subject matter [*facts, data*] from the dimension of the theory with which they describe it [*theory* or *model*]." They treat the subject matter itself, however, without depth: the subject matter is not set against a "background of intentions and values." The interpretive disciplines, on the other hand, place their subject matter and their account of it "within a single, continuous" network. They give depth to the subject matter by distinguishing within it "the two aspects of outward symbol and inward intent or value. The second gives the first its meaning."[39]

POLICY ANALYSIS AS AN INTERPRETIVE DISCIPLINE

It appears from this characterization of the distinction between empirical sciences and interpretive disciplines that policy analysis is more like law, grammar, and theology than it is like an empirical science. Majone has written of the kinship to law and rhetoric. The logic of choice is a close relative of grammar, as a small taste of a Kenneth Burke or a Lewis Carroll easily suggests. And to bring the theological analogy closer to home, let us remember Webber and Rittel on "wicked problems."[40] As summarized by Quade:

[Wicked problems] have no definite formulation and no stopping rule to tell the problem-solver when he has a solution. Moreover a proposed solution is not true or false but good or bad. There may be neither an immediate nor even an ultimate test of a solution—the set of potential solutions is not enumerable; every such problem is essentially unique and is a symptom of another problem. . . . The problem has to be "structured" before the traditional methods can apply and in structuring it may no longer remain the same problem.[41]

Wicked problems look very much like the age-old problems of ethics and morality that theologians deal with.

We *recognize* policy analysis as having interpretive aspects in many of our communications to each other and to the users of analysis. We do not, however, *identify* policy analysis as an interpretive discipline, i.e., as a humanistic art, nor should we, since analysis is other things as well.

While it is good not to identify analysis as one of the humanistic arts—not to say out loud that analysis is more like theology than it is like empirical science—it may not be so good to disown, or at least not to make our own, the

tools and methods of these disciplines. Several of these methods are in fact tools of language-creation designed for solving "wicked problems." (In case my reader has forgotten—it's been a long way round—this ties back to Holt's view of language-creation and problem-solving as indissoluble activities.)

To illustrate what I mean by recognizing analysis as an interpretive discipline while virtually ignoring interpretive methods, I again focus on the editors of this volume (on the presupposition that they are among the opinion leaders in the semantics of analysis).

Quade, in talking about "Some Difficulties with Objectives," writes:

> Legally, there are few requirements for policy-makers to make clear statements of their goals. On the other hand, in many circumstances, there are clear political advantages in being ambiguous about, and even silent concerning, goals and objectives. . . . In stating objectives, officials do not always reveal what they really want, sometimes because they feel they must maintain a position or front, sometimes because it might mean loss of support, sometimes (maybe most often) because they do not know what they want.

Quade then asks, "How does an analyst help identify objectives?" Despite the difficulties described in knowing what the objectives are, the analyst somehow "must obtain one or at most a limited set [of goals] to use in his analysis." How is this to be done? "By interpretation," you might say if you considered yourself a participant in an interpretive discipline. But Quade does not say that. He says that in helping to set goals the analyst "has only one tool—analysis, supported by his judgment," and goes on to quote Hitch saying that "nothing but rigorous, quantitative analysis" can tell us what we need to know about objectives.[42] This may be so, but it assumes that the objectives are already identified, and the problem Quade started with was how to identify them.

Majone moves in the opposite direction, *from* something not wanted, i.e., a problem situation, rather than *to* something wanted, i.e., objectives:

> Systems analysis is concerned with problem solving, but it usually begins with something less structured than a problem, namely a problem situation. This is an awareness that things are not as they should be, but without a clear idea of how they might be put right. Problem setting refers to the translation of a problem situation into a policy problem, expressing the goals to be achieved and a strategy for their accomplishment. Overlooking the importance (or, indeed, the very existence) of this stage of analysis is a pitfall whose seriousness can be inferred from the following quotation (Rein and Schon, 1976): "Policy development is essentially about a process of *problem-setting*; it is concerned with developing new purposes and new interpretations of the inchoate signs of stress in the system which derive from the past. . . . Problem-setting is important not only because it is difficult but because the questions we ask shape the answers we get."
>
> The amount of detail that is useful at the stage of problem setting is different from what is needed in the phase of actual problem solving. The appropriate styles of thinking will also differ in the two situations. Because of the unstructured character of a problem situation, imagination, judgment, and analogical and associative thinking play a bigger role than rigor and technical skills. (Rein and Schon, 1976, and Ravetz, 1973).[43]

Majone then switches to an empirical science perspective and goes on to talk of "data and information."

Majone does refer here to interpretive tools—to "imagination, judgment, and analogical and associative thinking." And the Hitch quotation Quade turns to does refer to "inventive and ingenious analysis." But where in the literature, and where in the training of analysts, is attention given to methods of improving imagination, judgment, analogical and associative thinking, inventiveness and ingenuity? What proportion of books and articles and classes are devoted to such topics, compared to the space and time devoted to improving rigor and technical skills? It is a minute fraction of the total verbiage on policy analysis.

Now that we recognize the nature of the problems we are dealing with and the importance of synthesis and design, oughtn't we to explore all the methods that have been used for dealing with such matters? To take on such a task presents an enormous challenge. But now that the disjunction between policy analysis and mathematics and logic, on the one hand, and between policy analysis and empirical science, on the other, has been recognized, what other alternatives are there? By responding to the problems confronting policy makers and being willing to break through disciplinary boundaries, policy analysis has come face to face with those phenomena of culture and consciousness that do not fit into either a logical or a causal mode of explanation. Yet the language of policy analysis, I suggest, still derives largely from these two modes of explanation. There is an awareness that policy analysis lacks theoretical foundations and that many of the pitfalls of analysis are attributable to this lack.[44] There is another way of looking at this problem, however. Analysis cannot build securely upon the foundations provided by the logical and causal modes of explanation; yet it has, not surprisingly, borrowed from the language and thus the metaphysics of disciplines which rely on these modes.

This is not a pitfall of analysis *per se*. It is a pitfall of the Western intellectual tradition, certainly as it has developed over the last two or three hundred years, and parts of the problem can be traced back further.[45] To get outside a system of thought that dominates one's whole culture is no mean feat. Yet any system of thought is a kind of trap: it influences the kinds of belief persons consider themselves entitled to hold and the kinds of social order they take for granted.[46] It is a particularly insidious kind of trap because it is a net woven with words. As a system of thought becomes dominant, it dominates the natural language so that the patterning of the symbolic apparatus is the trap. Yet the only way out of the trap, and also the only way to come to a shared recognition of the existence of the trap, is via the use of the very words that are knotted into the net. If all words had univocal meanings, perhaps there would be no way out.

It is easier to understand why devoting insufficient time to problem formulation is a pitfall of analysis if one understands how the principle of analysis fits

into the tradition of Western thought. Quade suggests that the leaders of a project should devote as much as 50 percent of their time to exploring broadly and deeply all aspects of the problem and of policy objectives. "The pitfall is to give in to the tendency to get started without having devoted a lot of thought to the problem."[47] Where does this "tendency" to get started on the analysis of particulars come from? According to Unger, it comes from the principle of analysis, a principle that holds that "the procedures of analysis and combination exhaust the ways in which we can gain knowledge." Since you cannot combine until you have parts to combine, you start with an analysis of parts.

> The analytical view of knowledge completes the dismemberment of the classical conception of theory. . . . Its general consequence is to demoralize all attempts to construct conceptual systems before the patient investigation of particulars, for it is in these particulars rather than in the system which pretends to describe them that reality lies. Consequently, the principle of analysis furnishes a priceless defense of the specialization of the sciences. . . . It also turns us away from any attempt to understand our social situation as a whole, a whole that could never be elucidated by the study of parts alone. . . .
>
> The principle of analysis may owe much of its appeal to a particular form of social order, to whose perpetuation it in turn contributes: the situation in which each man's social existence is divided into a diversity of roles. Every role calls for the exercise of special skills and the satisfaction of special expectations. . . . When all interests are viewed as either private and subjective themselves or as combinations of subjective and private interests, no activity can be seen as the expression of a universal interest whose realization demands a universal knowledge of society. A fractured social existence can only produce a fractured knowledge of the social order. The principle of analysis is therefore the natural limit to its self-consciousness.[48]

There is no way in a paper such as this to summarize the carefully constructed "total critique" that Unger sets forth nor to consider the work of other authors who have attempted to demonstrate that we are caught in a metaphysical belief system that severely limits our vision of alternatives.[49] I wish only to sound the warning again and, more important, to suggest that policy analysis *has already* made a number of moves that will contribute to a disentangling of this net.

The willingness of the best policy analysts to follow a problem wherever it leads, through disciplinary walls as necessary and even beyond the boundaries of established disciplines, has done much to bring into question the metaphysical tenets of the social sciences. The involvement of policy analysis in grave problems and issues has led the field to recognize, albeit hesitantly and tentatively, the need for interpretation and the impossibility of separating fact and value in airtight compartments.

As policy analysis has come of age, however, and attempts are made to describe what it is and where its boundaries shall be drawn, a confusion of identity is evident. Rigor and technical virtuosity are admired and often even set up as the sole ideal, but when it comes to examining pitfalls, we find that the most serious pitfalls will not be circumvented by greater rigor or improved

technical skills. Competencies usually considered "softer"—imagination, judgment, interpretive skills—are just as important.

WORDS ON THE WAY TO DESIGN

If what we call Analysis is to be about synthesis and design, then it needs a working vocabulary to deal with problems of aesthetics and ethics, with issues of "adequacy, harmony and appositeness."[50] And if it is not to remain enmeshed in inappropriate metaphysical belief systems, it will also need a language of metaphysical design. The obvious place to look for help in developing such languages is outside the community of scientific and logical discourse.

Such a search would lead analysis into unfamiliar territory; it would, in particular, lead analysis into the wonderland of symbolic and hermeneutic methods, old and new. Not everything in this wonderland would be unfamiliar. Cost analysis of a rudimentary form still exists just over the border on the other side of the looking-glass.

> "Now then! Show your ticket, child!" the Guard went on, looking angrily at Alice. And a great many voices all said together ("like the chorus of a song," thought Alice) "Don't keep him waiting, child! Why, his time is worth a thousand pounds a minute!"
>
> "I'm afraid I haven't got one," Alice said in a frightened tone: "there wasn't a ticket-office where I came from." And again the chorus of voices went on. "There wasn't room for one where she came from. The land there is worth a thousand pounds an inch!"
>
> "Don't make excuses," said the Guard: "you should have bought one from the engine-driver." And once more the chorus of voices went on with "The man that drives the engine. Why, the smoke alone is worth a thousand pounds a puff!"
>
> Alice thought to herself "Then there's no use in speaking." The voices didn't join in, *this* time, as she hadn't spoken, but, to her great surprise, they all *thought* in chorus (I hope you understand what *thinking in chorus* means—for I must confess that *I* don't), "Better say nothing at all. Language is worth a thousand pounds a word!"[51]

What is known in policy analysis as iteration—that is, a circular reasoning from objectives to alternatives to data and back again as many times as necessary—finds a more hospitable domicile in hermeneutic territory than on the other side of the boundary where circular reasoning is sometimes frowned upon.

Through the looking-glass is no idle metaphor; what better way to imply self-referentiality than going through a mirror? Self-referentiality is a characteristic of both culture and consciousness. It can be demonstrated that this self-referentiality generates a basic indeterminancy[52] that in turn gives rise to (a) the presence of uncertainty and those serious pitfalls associated with it that are inherent in analysis[53]; (b) the possibility of novelty, of creative solutions; and (c) the indissolubility of language-creation and problem-solving and thus the need to use old words in new ways.

Compare the task of the analyst facing wicked problems with that of the artist. Unlike the elements in logical and causal connections, the elements of a policy or a painting "are not connected with one another by necessity," although as Unger puts it, "their coexistence is ordered."

> The concept of style helps illustrate these points. The distinctive elements of a style in painting do not cause one another, nor are they logically entailed by each other. They cannot be ordered serially, for there is no feature that comes first. . . . The different attributes of the style, as displayed in a work of art, are not necessary conclusions or effects of one another, but they stand in a relationship of appositeness. What precisely is the nature of this harmony, and by what standards do we establish and measure it?[54]

The elements of a policy or a painting, or more generally of a structure or a system, "cause" one another reciprocally. *To explain* such a structure or system involves an iterative circularity, the parts being explained by each other and the whole, and the whole being explained by the parts. If you put the policy or painting in a broader context and relate it to its historical time and place, you do not avoid such circularity. You merely confront it again on a different level. The explanation, if such it is, is one of mutual self-consistency; the reasoning is circular.

To create a new style in painting or *to design* a novel policy solution, who can say what comes first? Does an idea of the whole bring the parts into being, are the parts arranged to bring a new whole into being, or do parts and whole come into being together iteratively and dialectically? Whichever, we have a circularity or, more precisely, a spiraling that produces novelty.

Policy analysis engages of necessity in such circular or spiraling reasoning. Problems, objectives, alternatives, data, and information are mutually interdependent. And all are set within and mutually interdependent with the culture that produces them and the consciousness that attempts to grasp their meaning. Self-referentiality means simply that an answer is fed back into the expression from whence it came. A tentative statement of objectives is used to derive alternatives. An attempt is made to understand the consequences of the alternatives. Then these consequences are compared with the first tentative statement of objectives in order to revise it in light of what is now known about alternatives. And so on, in a bootstrapping fashion.

This spiraling self-referentiality, or iteration, is accepted within policy analysis, but there has been no attempt in the analytic literature, as far as I know, to provide a theoretical justification. Such a justification can be found once we move into the territory of those disciplines that define themselves as interpretive. Within this territory, the method is known as the hermeneutic circle. As described by Dilthey:

> [It] starts from the apprehension of indefinite–definite parts and proceeds to the attempt to grasp the meaning of the whole, alternating with the attempt to take this meaning as a basis for defining the parts more clearly. Failure makes itself known when individual parts cannot be understood in this way. This then creates the need to

redefine the meaning so that it will take account of these parts. This attempt goes on until the entire meaning has been grasped.[55]

The justification of the apparent circularity of the interpretive process lies in what Habermas refers to as "the peculiar double status" of the words and actions one is trying to understand. They are both *symbols* and *facts*.[56]

The hermeneutic circle called iteration in policy analysis is designed to circumvent a pitfall: the pitfall of inappropriate objectivity. This is the pitfall of treating meaning (or value) and fact as independent when dealing with phenomena of culture and consciousness.

Other characteristics of the symbolic and hermeneutic territory will be less familiar to policy analysis, and it is precisely at these points that the interpretive methods may be most useful in bringing to our attention pitfalls we are not yet aware of. For instance, the facile assumption that people will behave selfishly and egotistically, so common in policy analysis, and the equally facile assumption that people will behave altruistically, more common among reformers "outside" the system, both ignore the interdependence of individual behavior and institutions. Such presuppositions also ignore the openness of both consciousness and culture to reinterpretation and thus to redesign. Policy analysis is intimately involved in changing both institutions and individual behavior, both culture and consciousness; thus, to ignore the interdependencies and to take as fixed that which is in fact open to redesign could be a serious pitfall, particularly in an era when we are pressing on the limits of our institutions and our planet.

In a paper of this length I can do no more than suggest the kinds of pitfalls that a reconnaissance of the symbolic and hermeneutic territory would help policy analysis detect. But does a craft called Analysis have any right to move into this territory and claim its methods for its own? It does, just as soon as it steps over the boundary.

On the other side of the boundary, the plurivocality of "analysis" is no longer an embarrassment but rather a resource to be conjured with. "Analysis" means etymologically a loosening up or setting free. How would one loosen up a problem composed of intellectual constructs, or a metaphysical net of words, other than by loosening up one's use of words? That is, by accepting and using some of the words in one's working language as symbols swimming in a sea of contexts, as symbols whose value resides in their multiplicity of meanings and associations—a multiplicity that permits the creation of new meanings and thus the creation of a new language more capable of coping with wicked problems.

Kahn and Mann gave us the pitfall.[57] Vickers gave us the lobster pot.

Lobster pots are designed to catch lobsters. A man entering a man-sized lobster pot would become suspicious of the narrowing tunnel, he would shrink from the drop at the end; and if he fell in, he would recognize the entrance as a possible exit and climb out again—even if he were the shape of a lobster.

A trap is a trap only for creatures which cannot solve the problems that it sets. Man-traps are dangerous only in relation to the limitations on what men can see and value and do. The nature of the trap is a function of the nature of the trapped. To describe either is to imply the other.[58]

Man-traps are most often made of words. If analysis is not to be entrapped by its own name, it may be helpful to remember that, among other things, "analysis" is

> An untangling, a loosening up, a setting free,
> With an iterative sense built into the *ana-*.
> It is not a cutting up. That's anatomy.
> Nor being beyond feeling—that's anaesthesia.
> It's a word of feminine gender with claims to the soft.
> And it has for a homonym, An Alice Is.

ACKNOWLEDGMENTS

I would like to thank Robert E. Bickner, Anne Highsmith Dale, Giandomenico Majone, Esther D. Nisewanger, and Edward S. Quade for helpful comments.

NOTES AND REFERENCES

1. Hogben, L. 1964. The Mother Tongue. London: Secker and Warburg, p. 274.
2. Green, R. L. (ed.). 1965. The Works of Lewis Carroll. Feltham, Middlesex: Hamlyn, p. 182.
3. Habermas, J. 1971. Knowledge and Human Interests (J. J. Shapiro, trans.). Boston: Beacon Press, p. 344.
4. Quade, E. S. 1975. Analysis for Public Decisions. New York: American Elsevier, p. 5.
5. Castaneda, C. 1971. A Separate Reality. New York: Simon and Schuster, pp. 143–148.
6. Carroll[2] (p. 6).
7. Quade, E. S. 1964. Orientation, p. 7. *In* Analysis for Military Decisions (E. S. Quade, ed.). Chicago: Rand-McNally.
8. Quade, E. S., and W. I. Boucher. 1968. Introduction, p. 2. *In* Systems Analysis and Policy Planning: Applications in Defense. R-439-PR (abridged). Santa Monica, California. The Rand Corporation. Emphasis added and deleted.
9. Quade[7] (p. 24).
10. Quade[7] (p. 4).
11. This and subsequent etymological references from
    Partridge, E. 1966. Origins. 4th edition. London: Routledge and Kegan Paul.
    Shipley, J. T. 1967. Dictionary of Word Origins. Totowa, New Jersey: Littlefield, Adams.
12. Quade.[7] Emphasis added.
13. Webster's Synonyms, Antonyms and Homonyms. 1962. Ottenhelmer Publishers, distributed by Crown Publishers, New York, p. 97.
14. Majone, G. 1977. Pitfalls of analysis and the analysis of pitfalls. Journal of Urban Analysis 4: 248.
15. Borrowed from Carroll[2] (p. 171).
16. Majone[14] (p. 250). Emphasis added.

17. Carroll[2] (p. 174).
18. Kirk, D. F. 1962. Charles Dodgson, Semeiotician. Monograph, Humanities, No. 11. Gainesville: University of Florida, p. 56.
19. See, for instance,
    Hadamard, J. 1954. An Essay on the Psychology of Invention in the Mathematical Field. New York: Dover. and
    Koestler, A. 1967. The Act of Creation. New York: Dell.
20. Quade[4] (p. 27). Emphasis added.
21. Hadamard[19]; Koestler[19].
22. Carroll[2] (p. 235).
23. Holmes, R. W. 1971. The philosopher's Alice in Wonderland, pp. 159–174. *In* Aspects of Alice (R. Phillips, ed.). New York: Vanguard.
    Huxley, F. 1976. The Raven and the Writing Desk. London: Thames and Hudson.
    Kirk[18].
24. Hitch, C. 1973. An appreciation of systems analysis, pp. 31–33. *In* Systems Analysis (S. L. Optner, ed.). Harmondsworth: Penguin.
    A slightly different version of the paper is also in Quade[7] (pp. 13–23).
25. Popper, K. R. 1966. The Open Society and Its Enemies. Volume II, 5th edition. London: Routledge and Kegan Paul, p. 16. Emphasis added.
26. See
    Adams, J. L. 1974. Conceptual Blockbusting. San Francisco: Freeman.
    Koberg, D. and J. Bagnall. 1973. The Universal Traveler. Los Altos, California: William Kaufmann.
    and the excellent bibliographies in each.
27. Hitch[24] (p. 36).
28. Bateson, M. C. 1972. Our Own Metaphor. New York: Knopf, pp. 159–160.
29. Ibid., p. 160. Emphasis added.
30. Ibid., pp. 160–161.
31. Brown, N. O. 1966. Love's Body. New York: Random House, p. 192.
32. Hogben[1]. Emphasis added.
33. Webster's New Collegiate Dictionary. 1974. Springfield, Massachusetts: Merriam.
34. Majone[14].
35. Bateson[28].
36. Quoted in Brown[31] (p. 192).
37. Quoted in Brown[31] (p. 192).
38. Majone[14] (p. 247).
39. Unger, R. M. 1975. Knowledge and Politics. New York: The Free Press, pp. 112–113.
40. Webber, M. M., and H. W. J. Rittell. 1973. Dilemmas in general theory of planning. Policy Sciences 4: 155–169.
41. Quade[4] (p. 8).
42. Quade[4] (pp. 85–87).
43. Majone[14] (p. 239). The references cited are
    Ravetz, J. R. 1973. Scientific Knowledge and its Social Problems, Harmondsworth: Penguin.
    Rein, M., and D. A. Schon. 1976. Problem Setting in Policy Research. Cambridge, Massachusetts: MIT Department of Urban Studies and Planning, mimeo.
44. Quade[4] (p. 299 and elsewhere).
45. See
    Becker, E. 1968. The Sructure of Evil. New York: The Free Press. and
    Unger[39] for the relevant intellectual history of the past two hundred or so years. The stress in Western thought on what various authors have referred to as righthanded thinking, masculine

consciousness, or Apollonic rationality favors the logical and causal modes of explanation and can be traced all the way back to the early Greeks. See, for instance,
Hillman, J. 1970. First Adam, then Eve: fantasies of female inferiority in changing consciousness. Art International 14: 30–42.

46. Unger[39] (p. 47).
47. Quade[4] (p. 306).
48. Unger[39] (pp. 48–49).
49. For a sampling of such work see Becker,[45] as well as
Berry, W. 1977. The Unsettling of America. San Francisco: Sierra Club Books.
Nozick, R. 1974. Anarchy, State, and Utopia. New York: Basic Books.
Schon, D. A. 1971. Beyond the Stable State. New York: Random House.
Wolin, S. S. 1960. Politics and Vision. Boston: Little, Brown.
50. Unger[39] (p. 14).
51. Carroll[2] (p. 139).
52. See
Brown, G. S. 1969. Laws of Form. London: Allen and Unwin.
Varela, F. J. 1975. A calculus for self-reference. International Journal of General Systems 2: 5–24.
53. Quade[4] (p. 300).
54. Unger.[39]
55. Dilthey, W. Complete Works, Vol. 7. p. 227, translated by Shapiro and quoted in Habermas[3] (p. 170).
56. Habermas[3] (p. 171).
57. Kahn, H. and I. Mann. 1957. Ten Common Pitfalls. RM-1937. Santa Monica: The Rand Corporation.
58. Vickers, G. 1970. Freedom in a Rocking Boat. Harmondsworth, Middlesex: Penguin, p. 15.

The Authors

STANLEY M. ALTMAN
W. Averell Harriman College of Urban and Policy Sciences
State University of New York at Stony Brook, Stony Brook, New York

KATHLEEN A. ARCHIBALD
School of Public Administration
University of Victoria
Victoria, British Columbia

EUGENE BARDACH
Graduate School of Public Policy
University of California, Berkeley
Berkeley, California

E. M. L. BEALE
Scientific Control Systems Ltd. and Scicon Computer Services Ltd.
Milton Keynes, United Kingdom

ROBERT E. BICKNER
Academic Programs
University of California, Los Angeles
Los Angeles, California

HARRY P. HATRY
The Urban Institute
Washington, D.C.

LAURENCE E. LYNN, JR.
John F. Kennedy School of Government
Harvard University
Cambridge, Massachusetts

GIANDOMENICO MAJONE
International Institute for Applied Systems Analysis
Laxenburg, Austria

ARNOLD J. MELTSNER
Graduate School of Public Policy
University of California, Berkeley
Berkeley, California

EDWARD S. QUADE
International Institute for Applied Systems Analysis
Laxenburg, Austria

Index

Edward Quade

Ed Quade received his first and Masters degree from the University of Florida, and his Ph.D. in Mathematics from Brown University. He taught Mathematics at both Universities and became Professor at the University of Florida in 1948. He was with the Rand Corporation as a mathematician ending up as Senior Mathematician. He retired in 1973 after which he served as a consultant and is now Executive Editor at the International Institute for Applied Systems Analysis, Laxenburg, Austria.

Contents

THE WILEY IIASA INTERNATIONAL SERIES ON APPLIED SYSTEMS ANALYSIS

Conflicting Objectives in Decisions
Edited by David E. Bell, *University of Cambridge;* Ralph L. Keeney, *Woodward-Clyde Consultants, San Francisco; and* Howard Raiffa, *Harvard University*

Material Accountability: Theory, Verification, and Applications
Rudolf Avenhaus, *Nuclear Center Karlsruhe and University of Mannheim*

Adaptive Environmental Assessment and Management
Edited by G.S. Holling, *Institute of Animal Resource Ecology, University of British Columbia*

Organization for Forecasting and Planning: Experience in the Soviet Union and the United States
Edited by William R. Dill, *New York University; and* G. Kh. Popov, *Moscow State University*

Management of Energy/Environment Systems: Methods and Case Studies
Edited by Wesley K. Foell, *University of Wisconsin-Madison*

Systems Analysis by Multilevel Methods: With Applications to Economics and Management
Yvo Dirickx, *Twente University of Technology; and* L. Peter Jennergren, *Odense University*

Connectivity, Complexity, and Catastrophe in Large-Scale Systems
John L. Casti, *New York University*

Pitfalls of Analysis
Edited by G. Majone *and* E.S. Quade, *International Institute for Applied Systems Analysis*

Control and Coordination in Hierarchical Systems
W. Findeisen, *Institute of Automatic Control, Technical University, Warsaw;* F.N. Bailey, *University of Minnesota;* M. Brdyś, K. Malinowski, P. Tatjewski, *and* A. Woźniak, *Institute of Automatic Control, Technical Institute, Warsaw*

Forthcoming

Computer Controlled Urban Transportation
Edited by H. Strobel *(GDR)*

Mathematical Modeling of Water Quality
G. Orlob *et al.*

The Appraisal of Costs for Major Decisions
R. Bickner *(USA)*

The Electronic Oracle
D. Meadows *and* J. Robinson *(both of USA)*

Ecological Policy Design: A Case Study of Forests, Insects, and Managers
G.L. Baskerville, W.C. Clark, C.S. Holling, D.D. Jones, *and* C.A. Miller *(all of Canada)*

JOHN WILEY & SONS
Chichester · New York · Brisbane · Toronto
A Wiley–Interscience Publication

ISBN 0 471 27746 0