RM-77-1

PITFALLS OF ANALYSIS AND THE ANALYSIS OF PITFALLS

G. Majone

January 1977

PREFACE

Like scientific research, applied systems analysis is essentially a craft activity; even though it does not operate on physical things and phenomena, but on intellectual constructs arising in the investigation of policy problems. Successful analytic work depends crucially on an intimate knowledge of methods and tools, and their limitations, and on a highly personal relationship between the analyst and his task.

The craft character of systems analysis can be seen most clearly in the concept of pitfall. A pitfall is the sort of error that destroys the solution of a problem and nullifies the validity of a policy recommendation.

Perhaps the most reliable way of assessing the maturity of a field of inquiry is the extent to which its common pitfalls are recognized. As a contribution to the methodological development of systems analysis, the IIASA Survey Project will publish a multi-authored volume on the limitations and pitfalls of the most commonly used analytic tools. The present paper attempts to provide a conceptual foundation for this volume. It is addressed to the practitioners as well as to the users of systems analysis.

ABSTRACT

The literature of applied systems has devoted considerable attention to the treatment of pitfalls. The present paper extends previous discussions in two ways: by introducing a new categorization of pitfalls: and by examining their epistemological, technical, and conceptual roots.

Analytic pitfalls are grouped around four rubrics that closely correspond to the four components of the analytic task: a) problem setting, data, and information; b) tools and methods; c) evidence and argument; d) conclusions, communication, and implementation.

A number of examples are discussed, and it is argued that analytic methods and techniques can be best understood in terms of the pitfalls they are designed to circumvent.

Pitfalls of Analysis and the Analysis of Pitfalls

Giandomenico Majone

"In analysis, the pitfalls are everywhere dense"*

INTRODUCTION

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" (Koopman, 1956) that may lead the unwary to absurd conclusions. A pitfall is for the nondemonstrative arguments used in the empirical sciences, in technology, and in systems analysis what the fallacy is for the deductive reasoning of logic and mathematics. In both cases, one has to be always on guard against hidden mistakes that have the power of destroying the entire validity of a conclusion.

Logicians distinguish between a fallacy and a simple falsity. A single statement may be false, but what is fallacious is the transition from a set of premises to a conclusion. Similarly, in systems analysis pitfalls should not be confused with blunders or errors that may affect, for instance, the numerical value of a solution but not the basic structure of the argument supporting it.

In logic there is a tradition of systematic discussion of fallacies that goes back to Aristotle's *De Sophisticis Elenchis*. John Stuart Mill devoted Book V of *A System of Logic* to an account and new classification of fallacies, and De Morgan, while rejecting previous attempts to produce exhaustive descriptions of all possible types of fallacy, still devoted an entire chapter of his *Formal Logic* to a penetrating analysis of many of the traditionally listed fallacies (Mackie, 1967).

---

*Dictum attributed to the distinguished British mathematician A.S. Besicovitch (Ravetz, 1973). The reference here is, of course, to mathematical analysis, where a set E is said to be everywhere dense if every point of the space containing E is a limit point of E.

More recent contributions, mainly devoted to a discussion of fallacies in everyday thinking, are Robert H. Thouless's *How to Think Straight* (1932, new edition 1947), and Susan Stebbing's *Thinking to Some Purpose* (1939).

Outside of logic and philosophy, the amount of attention devoted to the topic of pitfalls varies considerably among different disciplines. Very few natural sciences have standard literature on the possible pitfalls of their characteristic patterns of argument (Ravetz, 1973). On the other hand, the literature of statistics, a discipline specifically concerned with the logic of inductive reasoning and the weighing of evidence, contains many insightful discussions of pitfalls, both at the technical level (perhaps exemplified at their best by the published discussions of the Royal Statistical Society) and at the level of textbooks and popular expositions (Huff, 1954; Wallis and Roberts, 1956; Reichman, 1961). To some extent, this tradition has been carried over into the neighboring field of econometrics (Johnston, 1963; Cramer, 1969), but it does not seem to have penetrated deeply into actual econometric practice (Streissler, 1970). Social science literature reveals only a marginal awareness of the general significance of pitfalls; for instance, the only detailed discussion of the topic in the eight-volume *International Encyclopedia of the Social Sciences*, is the perceptive article on statistical fallacies by I.J. Good (1968). In systems analysis, a number of standard works include extensive treatment of pitfalls, and some of these discussions have attained the status of minor classics of the discipline (Koopman, 1956; Hitch, 1956, 1958; Kahn and Mann, 1957; Hitch and McKean, 1960; Quade, 1968, 1975).

Of course, these comparisons can be misleading if one is not careful to distinguish between explicit awareness of a problem and the inarticulated practical knowledge that results from long and successful experience. Thus, an explicit treatment of pitfalls is less important in the natural sciences than in other fields because it is possible to make practical tests of theoretical conclusions and because of the existence of effective professional mechanisms that impose quality control on results. Also, laboratory courses and similar devices help the student develop an intuitive feeling for the possibility of pitfalls in the standard procedures by which he verifies theoretical results. The situation prevailing in systems analysis is quite different. Here, direct verification of conclusions is seldom possible; professional mechanisms for the control of quality of analytic work are still in an embryonic stage; and the approach is too new for a widely shared tradition of critical thought to have developed. The significance of these factors as indicators of the need of discussing pitfalls in systems analysis is increased by two other characteristic features: its interdisciplinary character and the myopic pragmatism prevailing among many of its practitioners.

Systems analysts come from the most diverse disciplinary backgrounds, and students now acquiring specialized training are exposed to academic curricula that vary from school to school and, that represent tentative compromises among different intellectual traditions. Thus, a sort of jargon or *lingua franca* has evolved 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. For example, in this simplified pidgin, "cost" is often taken to mean just historical or sunk cost; "average" is only the arithmetic or sample average, even when other parameters are appropriate; ratios of benefits to costs become measures of effectiveness valid under all circumstances.

Most of the technical tools used in systems analysis have been developed by other disciplines, of which the average analyst cares to know only those limited parts that seem most directly applicable. But concepts and techniques removed from their broader disciplinary context tend to become stereotypes, and their limitations are not easily perceived by people interested only in immediate utility. Thus originate the pitfalls that B.O. Koopman (1956) has labeled linearitis, maximitis, and mechanitis. In the same way, all the subtlety of statistical reasoning is lost in ritualistic and often meaningless applications of significance levels.

The literature of systems analysis, as already indicated, has devoted considerable attention to the treatment of pitfalls. The present paper attempts to extend previous discussions in two ways: by categorizing pitfalls in terms of the four basic components of analysis to be discussed below; and by examining in some depth their epistemological, technical, and conceptual roots, and pointing out their interrelationships. It will be apparent that pitfalls discussed under one heading could also have been treated, in some of their aspects and ramifications, under other categories. This is unavoidable in any classification, but as long as the classification reflects the underlying structural differences reasonably well, the resulting redundancy is not too serious and may even be useful. Before entering into details, one further general observation should be added. A discussion centering on pitfalls is necessarily critical of widespread methods and practices. This may leave a negative impression upon the reader, tending to obscure the substantial accomplishments of systems analysis. For this reason, the positive aim of the present essay should be stressed; as Ravetz (1973, p. 100) writes:

> 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.

SYSTEMS ANALYSIS AS CRAFT WORK

The real significance of the category "pitfall" for systems analysis is best appreciated in relation to the craft aspects of the field. There is an undeniable similarity between the work of an analyst and that of the traditional craftsman. In both cases, successful performance depends crucially on an intimate knowledge of materials and tools and on a highly personal relationship between the agent and his task. Good analytic work cannot be produced mechanically any more than handicraft work 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 craftsman's work.

There are also obvious differences: the craftsman uses concrete materials in order to produce an object that has an appropriate shape and serves well-defined purposes; the analyst, on the other hand, operates with data, technical tools, and models to produce arguments and recommendations. In spite of these differences, the classical Aristotelian analysis of craft work can be usefully extended to systems analysis. The following treatment is patterned after Ravetz's (1973) penetrating discussion of the craft character of scientific inquiry.

Aristotle's scheme (described in his *Nichomachean Ethics)* involves four constituents (or "causes") of the craftsman's task: material, efficient, formal, and final. These refer, respectively, to the physical substance that is worked on; to the agent and the tools he uses in shaping it; to the shape acquired by the substance; and to the purpose of the activity--i.e., the creation of a specific object--or the functions served by the object itself. With suitable modifications, the same scheme can be applied to the analyst's work. To carry out this extension, the material component should be identified with the data, information, and conceptual constructs that are used in setting the problem to be analyzed. Tools, techniques, and models are the efficient components of the analyst's task. The "form" of the task 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.

A comparison between this description of the analyst's work and the more usual ones discussed in the literature would take us too far afield. For the purpose of the present discussion, it will suffice to point out that these other schemes are modeled on the structure of the decision process. Consequently, they rely primarily on categories like objectives and criteria, alternatives, benefits and costs, and choice. The two schemes are certainly not incompatible, but the one adopted here, relying as it does on the craft characteristics of analytic work, seems better suited to a discussion of pitfalls. Hence, the present treatment of pitfalls will be organized under four rubrics that closely correspond to the four components of the analytic task: (a) problem setting, data, and information; (b) tools and methods; (c) evidence and argument; (d) conclusions, communication, and implementation.

PROBLEM SETTING, DATA, AND INFORMATION

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; Ravetz, 1973). For the same reason, the data and information used in specifying the problem conceal even more pitfalls than is the case in subsequent stages of the work. The kind of information obtained to identify social "needs"--for instance, through opinion polls and attitude surveys--depends very much on the way the questions have been framed. Even census "facts" are susceptible to different interpretations according to the skill used in devising the questioning techniques. As Edwards Deming (1969, p. 656) reminds us,

> 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.

It has been said that the first pitfall of experimental research is that of too easily accepting readings that are stable as reports which are sound. Similarly, stability of replies is no test of a meaningful question. As Payne (1951, p. 17) points out, "the more meaningless a question is, the more likely it is to produce consistent percentages when repeated."

A more subtle type of pitfall is quite common in attitude surveys. It consists in attempting to measure total, rather than marginal, utilities, even when only marginal valuations can provide guidance to policymakers. The general mistake here is thinking that by measuring total utilities (at existing prices), one can infer something about what is needed to change behavior (at unspecified prices). For further discussion of this pitfall, with examples, see Lipsey (1975, pp. 171-172).

Once the policy problem has been specified (and a good deal of judgment is clearly required to avoid specifications that are either too narrow or too broad), new types of data will come into play, and with them come new types of possible pitfalls. Perhaps the most serious and widespread error at this stage is the failure to recognize the large margin of error surrounding all socioeconomic statistics. For example, a reported drop of one or two percentage points in a country's gross national product is generally viewed as a most significant indication that immediate government action is required. In fact, national income figures cannot probably be known without an error of ±10 to ±15 percent, and comparable uncertainties are present in foreign trade, price, unemployment, and growth statistics (Morgenstern, 1963).

This statistical pitfall has a conceptual counterpart. All too often, analysts and policymakers tend to forget the conventional and institutionally dependent character of the categories used in the description of social phenomena. For instance, unemployment is usually determined in terms of qualification for the relevant social security benefits, and "crime" clearly depends on the particular legal system and prevailing social conventions. Health indices, levels of education, public expenditures for research: these and all other basic categories of social policy must be interpreted in relation to a particular institutional context before they acquire any kind of operational meaning; their usefulness depends entirely on a clear recognition of their conventional character.

Even the best data are much too raw to be used in an analytic argument without being refined into a more reliable and useful form. This transformation requires craft skills that are different from those needed in problem setting and in gathering of data by sampling, experimentation, or utilization of already existing material. This new phase of the analyst's work, the production of information, can be illustrated by a number of examples: calculating averages and other statistical parameters; fitting a curve to a set of points; reducing data by means of some multivariate technique; devising systems of equations or inequalities to represent functional interdependencies. The operations performed on the original data may be technically involved or quite simple, but they always represent a crucial step. Through these operations, the data are transformed into information, and from this point on the analysis, except for an occasional check, is carried out exclusively in terms of intellectual artifacts.

The transformation of data involves three basic judgments, all of which present the risk of serious pitfalls. The first is that the data reduction does not involve too great a loss of information, relative to the problem under discussion (generally speaking, the existence of "sufficient statistics," containing the same amount of information as the original sample, cannot be assumed). The second is a judgment of the goodness of fit of the model to the original data. The third basic judgment is that this particular transformation of the data, among the many possibilities, is the significant one.

Quade (1975, p. 299) gives an entertaining example of a pitfall involving the third type of judgment: the use of the arithmetic instead of the harmonic mean in computing turnarounds of troop and cargo ships of different speeds during World War I. This is, admittedly, a rather trivial mistake (though a frequent one; see Reichman, 1961, Chapter 5), but it is precisely its elementary character that shows how easy it is to stumble into pitfalls in even the simplest aggregation of data. Another example of a pitfall of aggregation at a somewhat more advanced level, is the so-called "ecological fallacy," which has received a good deal of attention in the statistical and sociological literature (Robinson, 1950; Goodman, 1959; Allardt, 1969). The pitfall consists in using ecological correlations (i.e., statistical correlations involving properties of *groups* of individuals) as substitutes for individual correlations, in which the correlates are properties of individuals. Robinson has cast strong doubts on the validity of a number of empirical studies by showing that the two correlations are in general different (they may even differ in sign), and that the values of ecological correlations strongly depend on the type of grouping used.

Invalid or meaningless inferences about individual behavior from aggregate relationships abound in the planning literature (Polanyi, 1951). In particular, the gravity models used in urban planning, while reasonably reliable at the scale of a metropolitan area, have no explanatory power at the neighborhood level to which they are sometimes applied (Brewer, 1973; Lee, 1973).

A METHODOLOGICAL DIGRESSION

Pitfalls of aggregation and disaggregation are closely related to the logical fallacies of diversion and composition. The fallacy of division consists in arguing from the premise that something is true of some set or class considered collectively to the conclusion that the same is true of the parts or individual elements. In the fallacy of composition one argues from the properties of the parts to the properties of the whole.

It should be noted that both types of inference are perfectly legitimate in some circumstances. The pitfall consists in a failure to realize that difficulties *may* arise and that arguments must, therefore, be supplied in order to show that a particular inference is in fact legitimate. This is true in general: insufficient attention to the specific conditions under which some particular inference or action is or is not permissible is the common characteristic of all pitfalls. A few more examples will make this point clear.

Some very common pitfalls in mathematics and statistics are connected with the interchange of analytic operations: interchange of the order of partial differentiation for a function of several variables, termwise integration of differentiation of a series, or what B.O. Koopman (1956) has called the "fabulous law of averages":

$$E[f(X_1,\ldots,X_n)] = f(E[X_1],\ldots,E[X_n]) \quad , \qquad (1)$$

where $X_1,\ldots,X_n$ are random variables, f is a nonlinear function, and E is the expected value operator (for a numerical illustraion of this pitfall in the case of a networking problem, see Wagner, 1975, pp. 664-665). The special case

$$E[X^2] = (E[X])^2 \qquad (2)$$

has trapped legions of undergraduate students, and even some popularizers of statistics (e.g., Moroney, 1951, p. 250).

However, relations like (1) and (2) can occasionally be correct. Even more frequently, they can be *approximately* right, and the approximation may be good enough for the problem under discussion. Thus, if f(X) is approximately linear over the entire range of variation of X, a Taylor series expansion about the mean shows that f(E[X]) is indeed a good approximation to the exact value E[f(X)]. Analogous considerations apply in the case of the second moment or the variance.

The interchange of limiting processes is also valid under some conditions, namely when the convergence is uniform. Indeed, uniform convergence is a good example of a concept consciously developed (by outstanding mathematicians of the early nineteenth century like Abel) as a safe path through a region full of well-concealed pitfalls. The moral of these examples is clear: pitfalls result from disregarding the range of valid applicability of concepts, methods, or theories that, in themselves, are neither correct nor incorrect. Like the good craftsman, the analyst must be able to recognize and accept the limitations of his tools.

TOOLS AND METHODS

The tools of systems analysis may be roughly classified according to their function in data production, manipulation, or interpretation. The category of interpretive tools includes tool disciplines, like mathematics, statistics, or economics, which the analyst has to master to some extent in order to do competent work. The pitfalls attendant upon the use of particular results taken out of their broader disciplinary context have already been referred to in the introduction. The danger is made particularly acute by what Ravetz (1973) calls the "prevailing methaphysics," according to which the scientific character of a field is assumed to be in direct proportion to the degree of its mathematical formalization. As a result, the analyst is sometimes tempted to use formal tools that exceed the level of his mathematical or statistical sophistication and whose range of meaningful applicability he is therefore incapable of assessing.

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, on the other hand, we often witness the phenomenon of "new-toolism," a disease to which operations researchers and systems analysts seem to be particularly predisposed. Those affected by this disease "come 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" (Wohlstetter, 1970, p. 106).

In the preceding pages we have seen how difficult it is to obtain information that is both reliable and relevant. The difficulties are compounded when data are processed by means of formal techniques and models. For example: are the results derived from a particular model more sensitive to changes in the model and in the methods used to estimate its parameters, or to changes in the data? No general answer to this crucial question is available, and the limited evidence is conflicting. Thus, one study (Holden, 1969) comparing two estimation procedures (ordinary least squares and limited-information maximum-likelihood two-stage least squares) for a five-equation econometric model using two sets of data comes to the conclusion that

> The 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.

On the other hand, the authors (Denton and Kuiper, 1965) of a study comparing ordinary least squares and two-stage least squares for a much larger model (the Canadian Econometric Model) and using three sets of data find that:

> Variations 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.

The following example exhibits another aspect of the complex interrelationship between data and model. Consider a linear regression model

$$y_i = \beta_o + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i \quad , \qquad i = 1,\dots,n \quad (3)$$

subject to the usual assumptions concerning the error terms u_i. The x's can represent either planned or unplanned ("passive") observations; the same formal data manipulations are carried out in either case. But as Box (1966) has pointed out, the significance of the fitted regression equation is quite different in the two cases. If we are passively observing a system, rather than actively experimenting with it (under controlled conditions), the error terms typically represent the effect of some "latent" variables $x_{k+1},\dots,x_m$ that cannot be observed. Because of this, the well-known phenomenon of "nonsense" correlation may arise; we can, nevertheless, produce a meaningful estimate of y by means of the fitted equation $b_1x_1 + \dots + b_kx_k$, at least as long as the system continues to behave as it did when the observations were taken. On the other hand, if y actually depends on latent variables $x_{k+1},\dots,x_m$, rather than on the observed $x_1,\dots,x_k$, it would be quite misleading to interpret b_i as the effect on y of a unit change in x_i, as we are entitled to do when the observations come from a planned experiment.

This is only one of the several pitfalls into which the analyst can fall when he disregards the distinction between planned and unplanned data (more on this point can be found in Ravetz, 1973, pp. 78-80).

EVIDENCE AND ARGUMENT

The argument is the link connecting data and information with the conclusions of an analytic study. The structure of the argument will typically be a complex blend of factual statements and subjective evaluations. Along with mathematical and logical deductions, it will include statistical, empirical, and analogical inferences, references to expert opinion, estimates of benefits and costs, and caveats and provisos of different kinds. This unavoidable complexity makes any formal testing of the argument quite impossible. Whatever testing is done must rely on a variety of professional standards, corresponding to the different analytic methods employed; on the plausibility and robustness of the results; and on the assumed criteria of adequacy of the client.

The nature of the evidence plays a crucial role here, 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 of fact. An inappropriate selection of data or models, their placement at a wrong point in the argument, a style of presentation inappropriate for the audience to 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 abstract 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. 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 (1973) 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.

The same cannot be said of systems analysis, even though complexity of arguments and large amounts of data of doubtful reliability and relevance characterize analytic studies as well.

Among the most widespread pitfalls encountered in analytical studies in connection with argument and evidence, three deserve special notice. The first one originates in the contemporary fashion of using mathematical formalizations on every possible occasion. Kahn and Mann (1957, p. 47) observe that:

> The analyst often dresses up his results and attempts, either consciously or unconsciously, to hide fairly elementary notions in extreme mathematical and technical language. Though it is probably not possible to condense the most esoteric results of modern mathematics and physics into the language of the newspapers, this is just not true of any applied operations analyses that we have seen.

It should be added that an overly formalized style of presentation not only obscures the real issues and impedes assessment of the plausibility of the conclusions; it also induces a tendency to accept statistical information or the results of mathematical calculations as facts rather than evidence.

The second group of pitfalls is encountered when existing information is taken over for use in an analytic argument. All kinds of distortions occur when data gathered by one organization for broadly defined purposes are used by others to support specific conclusions (for some of the organizational problems involved, see Morgenstern, 1963, pp. 11-12). Whether such material is of sufficient strength and fit for its function in the argument depends on the mode of its original production, and this is often difficult for the analyst to assess and usually impossible for him to change.

Finally, questions concerning the acceptable degree of approximation of a numerical result or the acceptable level of precision of a set of data acquire their full meaning, for systems analysts at any rate, in connection with the use of evidence. Two pitfalls should be mentioned in this context: the belief that there is an absolute standard of adequacy, and the rejection of items of information or opinions for which consensus among the experts is lacking. The belief in absolute standards overlooks the fact that even the physical sciences use several degrees of acceptable levels of precision for their data simultaneously. For example, some physical constants are known with an accuracy of 10^{-14}, while the age of the earth can only be estimated with an error of billions of years. Because of the diversity of the data used in a typical analytic study, the acceptable margins of error may have to be even larger than those the economist or the sociologist must realistically accept. This does not mean, of course, that the systems analyst should not have high standards of quality for his evidence. The pitfall consists in setting the standards so high that they become self-defeating.

MODELS AS EVIDENCE

Large-scale, policy-oriented models have come under increasing criticism in recent years. Examples of conceptual, technical, and institutional pitfalls in model construction and utilization can be found in a number of perceptive review papers (Shubik and Brewer, 1972; Lee, 1973; Allen, 1975; Brewer, 1975) and in book-length case studies (Brewer, 1973; Ackerman et al., 1974; Feiveson et al., 1976).

Among the modeling efforts discussed by these writers, the Delaware Estuary Comprehensive Study examined by Bruce Ackerman and coauthors is probably the most thorough and competent. Yet even this study is marred by a number of serious pitfalls ranging from use of dissolved oxygen as the sole indicator of water quality and neglect of peak discharges of sewage when heavy rains overload the sewers, to incorrect estimation of recreational benefits and failure to call attention to the great uncertainties of these and other estimates.

The main conclusion to be derived from these critical reviews can perhaps be summarized by the statement that models do not, and cannot, provide *answers* to policy questions. Under suitable analytic and institutional conditions they can, however, supply *evidence* that, together with other information, intellectual constructs, and subjective judgments, can be used in arguments supporting a policy conclusion.

The path from model to conclusions is long, involved, and beset by difficulties of all kinds as Lee (1973, p. 167) has pointed out:

> 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.

To produce acceptable evidence, models must satisfy certain *procedural* requirements that make assessment of their output possible; they have to be "in proper form," as lawyers would say. Actually, the legal analogy is quite appropriate in this context, for it reminds us that the assessment of arguments is an activity necessarily involving formalities. It is not enough for an effective argument to have a particular shape, like the mathematical garb of a formal model. After all, a mathematical style of presentation is not incompatible with a "black box" approach (it may even encourage it), and black box models, as Lee points out (1973, p. 175), "will never have an impact on policy other than through mystique, and this will be short lived and self-defeating."

Brewer (1975) and Ackermann and coauthors (1974) among others, have stressed the need for developing institutional mechanisms for assessing modeling activities in terms of internal validity as well as policy relevance. But an effective appraisal function requires that analytic arguments be set out and presented in a sequence of steps conforming to certain basic rules of procedure, like those that hold when questions of law are debated in a court. 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 bear on the soundness of a legal claim?" Without the large number of distinctions that he introduces (statements of claim, evidence of different types, testimony, proof of facts, presumptions, interpretation of a statute or discussion of its validity, verdicts, sentences), it is impossible to understand the nature of the legal process (Toulmin, 1964).

The intrinsic complexity of policy analysis certainly matches that of legal argumentation. Hence, while it is important to insist that models be transparent and as simple as possible, more detailed procedural guidelines will have to be developed if models are to play their limited but potentially useful role in the policy process. It is ironic that while data generation absorbs so much of the modelers' time and ingenuity, the transition from these data to conclusions should be accomplished by arguments that often do not bear close scrutiny. 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.

CONCLUSION, COMMUNICATION, IMPLEMENTATION

The conclusion of an analytic study may be a forecast, an issue paper, a recommended course of action, or an assessment of ongoing policies. Whatever its nature, a conclusion is never concerned with "things themselves" but with those intellectually constructed concepts and categories that can serve as the objects of an argument. The contact with the external world of economic, social, technical, and political phenomena is always indirect and elusive. A different conceptualization of the problem situation, different tools or models, or a few different subjective judgments made at crucial points of the analysis can lead to quite different conclusions. This is true of any form 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.

Paradoxically, uncritical acceptance of the "scientific method" as the model that systems analysts should strive to imitate only generates disillusionment when it is realized that the conclusions of analysis suffer from the same limitations as those of science, without sharing their strengths. In fact, 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. They attempt to increase the collective confidence of a group of new disciplines striving for academic and social recognition; they do not direct attention to any deep intellectual affinities. But in order to be used as an ideology, science has to be viewed as the incessant accumulation of indisputable facts and timeless truths. The obscurity of its foundations, the artificiality of its objects of inquiry, the tentative character of its conclusions must all be forgotten. What remains--those finished products that find their way into textbooks and strike the popular imagination--is no longer science but, rather, folk-science, i.e., "a part of a general worldview, or ideology, which is given special articulation so that it may provide comfort and reassurance in the face of crucial uncertainties of the world of experience" (Ravetz, 1973, p. 386).

Basically, lack of appreciation of the problems of communication and persuasion--the pitfall that Kahn and Mann (1957) have called hermitism--stems from a misconception of both science and systems analysis. For only the believer in apodictic certainties can scorn the techniques of persuasion. Fallible conclusions must fight for acceptance before they can affect the course of events.

The natural and social sciences offer many examples of theories whose correctness, according to the usual scientific standards, was established only gradually, and only *after* people had become convinced of their values through more "irrational" methods of support. Thus, according to modern interpretations (Feyerabend, 1975, p. 141), Galileo's advocacy of Copernicanism was eventually successful not because of any decisive proof he could adduce (even his telescope would produce only conflicting evidence), but

> ...because of his style and his clever techniques of persuasion, because he writes in Italian rather than in Latin, and because he appeals to people who are temperamentally opposed to the old ideas and the standards of learning connected with them.

Discussing Adam Smith's principles of division of labor and free exchange, Alchian and Allen (1974, p. 200) 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.

Stigler (1955) also mentions the use of techniques of persuasion in the realm of ideas and gives other examples of famous economists who "have employed the techniques of the huckster."

If communication and persuasion are important for the acceptance of theoretical results, they become crucial in systems analysis. In fact, to avoid pitfalls of communication that may endanger the practical relevance of the entire study (some examples are discussed in Quade, 1975, pp. 312-317), 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 recommendations convincing "even to a hostile and disbelieving, but intelligent audience" (Kahn and Mann, 1956). Notice that the two stages must be viewed as closely interdependent, rather than as discrete and separate analytic components. As we have seen, the plausibility of a conclusion depends on the structure of the supporting argument and on the strength and fit of the evidence used in it.

The same need to consider simultaneously the different aspects of the analytic process arises in connection with problems of implementation. For all its obvious importance, implementation remains the *terra incognita* of systems analysis, and the few explorers who have ventured into it have accomplished little more than a first reconnaissance of the terrain. Without sufficient empirical material for a detailed discussion of pitfalls of implementation, we shall have to limit ourselves to suggesting an explanation of why problems of implementation have not yet found a satisfactory formulation, let alone practical solution, in systems analysis.

I shall argue that the main fallacy responsible for this state of affairs is the excessive reliance of many analysts on normative decision models and the pure logic of choice, on analysis rather than synthesis and design. This is particularly evident in the work of analysts raised in the normative spirit of contemporary economics and decision theory (Majone, 1975). Of course, this criticism is relevant only for applied studies. The 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 (Lindley, 1971). But this is a purely formal solution, of no help whatsoever for understanding the relationship between decision and action.

What, then, is the intrinsic limitation of normative models? Basically, it is this: the prescriptions and operating rules provided by such models set standards by which behavior can be evaluated, but they cannot explain deviations from the norms, nor can they indicate the means of correcting them. Suppose that the behavior of the system under investigation does not conform to the conditions of the model, perhaps because some previously ignored constraints must now be taken into consideration. Is it possible, within the context of the model, to understand the causes of the deviations, or to deduce the conditions for a second best solution from those for the original optimum? In general, the answer is no, and this is proved, at least for problems of economic policy, by the second-best theorem of welfare economics (Lipsey and Lancaster, 1956-1957).

This theorem shows that the normative rules for a Pareto optimum are not valid policy criteria in a situation where they are not all simultaneously satisfied. To achieve a second-best position it may, in fact, be necessary to violate even those rules that could have been satisfied. In other words, the common assumption that it is better to fulfill at least some of the optimum conditions rather than none turns out to be false. Hence, while the normative model gives precise conditions for an optimal or best position, there is no corresponding set of rules for the achievement of a second-best, or even a better position, in a world where the optimum is unattainable.

The situation described here is by no means peculiar to welfare economics. The same is true, for example, in the field of technology. The operational principles of a machine set a standard: the ideal of a machine in good working order. The principles become "rules of rightness" that account for the successful working of a machine but leave its occasional failures unexplained (Polanyi, 1962). The following remarks by two control theorists (Gumowski and Mira, 1968) also point in the same direction:

> If a mathematical model of a practical system is based on highly idealized elements, this model will apply, only if the design of the practical system has been carried to a successful completion. In other words, it will apply only *after* the designer has

> eliminated all possible causes of trouble. If the practical system does not yet operate properly, its mathematical model composed of idealized elements will not offer any clue permitting to locate the cause of trouble... . Even a superficial analysis will show that a very substantial part of contemporary control theory ... has relatively little to offer to the practical designer. In particular, it is not of much help in the selection of practical components permitting one to attain a given design goal.

Perhaps the problem of implementation is so difficult to formulate simply because it does not exist as something that can be factored out of all the other aspects of the policy process. Belief in its separate existence results from the normative tenet that full articulation of a plan should precede its institutional realization. In fact, "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" (Feyerabend, 1975, p. 26). Uncritical acceptance of the metaphysical distinction between idea and action is probably the most serious pitfall of applied systems analysis.

REFERENCES

Ackermann, B.A., S.R. Ackermann, J.W. Sawyer, Jr., and D.W. Henderson (1974). *The Uncertain Search for Environmental Quality*, New York, The Free Press.

Alchian, A.A., and W.R. Allen (1974), *University Economics*, London: Prentice/Hall International.

Allardt, E. (1969), Aggregate Analysis: The Problem of its Informative Value, in *Quantitative Ecological Analysis in the Social Sciences*, M. Dogan and S. Rokkan, eds., Cambridge, Massachusetts, M.I.T. Press.

Allen, W. (1975). *On Accuracy, Precision and the Real World: Some Thoughts on Systems Dynamics as a Policy Tool*, Santa Monica, The Rand Corporation, P-5540.

Box, G.E.P. (1966), Use and Abuse of Regression, *Technometrics*, vol. 8, no. 4, pp. 625-629.

Brewer, G.D. (1973), *Politicians, Bureaucrats, and the Consultant*, New York, Basic Books.

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

Cramer, J.S. (1969), *Empirical Econometrics*, Amsterdam, North Holland Publishing Co.

Deming, W.E. (1969), Boundaries of Statistical Inference, in N.L. Johnson and H. Smith Jr., editors, *New Developments in Survey Sampling*, New York and London, Wiley- Interscience.

Denton, F.J., and J. 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*, vol. 47, pp. 198-206.

Feiveson, H.A., F.W. Sinden, and R.H. Socolow, editors (1976) *Boundaries of Analysis*, Cambridge, Massachusetts, Ballinger Publishing Company.

Feyerabend, P. (1975), *Against Method*, London, NLB.

Good, I.J. (1968), Fallacies, Statistical, in *International Encyclopedia of the Social Sciences*, D.L. Sills, ed., New York, The Macmillan Company and The Free Press, vol. 5.

Goodman, L.A. (1959), Some Alternatives to Ecological Correlation, *American Journal of Sociology*, 64, pp. 610-625.

Gumowski, I., and C. Mira (1968), *Optimization in Control Theory and Practice*, Cambridge, U.K., Cambridge University Press.

Hitch, C.J. (1956), Professor Koopman on Fallacies: A Comment, *Operations Research*, vol. 4, no. 4, pp. 426-430.

Hitch, C.J. (1958), Economics and Military Operations Research, *The Review of Economics and Statistics*, pp. 199-209.

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

Holden, K. (1969), The Effect of Revision of Data on Two Econometric Studies, *The Manchester School*, pp. 23-27, quoted in Ravetz (1973), p. 84.

Huff, D. (1954), *How to Lie with Statistics*, New York, Norton.

Johnston, J. (1973), *Econometric Methods*, New York, McGraw-Hill, Inc.

Kahn, H., and I. Mann (1956), *Techniques of Systems Analysis*, Santa Monica, California, The Rand Corporation, RM-1829.

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

Koopman, B.O. (1956), Fallacies in Operations Research, *Operations Research*, vol. 4, no. 4, pp. 422-426.

Lee, D.B. Jr. (1973), Requiem for Large-Scale Models, *AIP Journal*, pp. 163-178.

Lindley, D. (1971), *Making Decisions*, London, Wiley-Interscience.

Lipsey, R.G., and K. Lancaster (1956-57), The General Theory of Second Best, *Review of Economic Studies*, vol. 26.

Lipsey, R.G. (1975), *An Introduction to Positive Economics*, 4th Edition, London, Weidenfels and Nicholson.

Mackie, J.L. (1967), Fallacies, in *The Encyclopedia of Philosophy*, P. Edwards, ed., New York, MacMillan Publishing Company, and The Free Press, vol. 3, pp. 169-179.

Majone, G. (1975), The Use of Decision Analysis in the Public Sector, in *Utility, Probability, and Human Decision-making*, K. Wendt, and J. Vlek, eds., Dordrecht, Holland, D. Reidel Publishing Company, pp. 397-407.

Morgenstern, O. (1963), *On the Accuracy of Economic Observations*, 2nd Edition, Princeton, N.J., Princeton University Press.

Moroney, M.J. (1951), *Facts from Figures*, Harmondsworth, U.K., Penguin Books.

Payne, S.L. (1951), *The Art of Asking Questions*, Princeton, New Jersey, Princeton University Press.

Polanyi, M. (1951), The Span of Central Direction, in *The Logic of Liberty*, London, Routledge and Kegan Paul.

Polanyi, M. (1962), *Personal Knowledge*, London, Routledge and Kegan Paul.

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 Publishing Company Inc., pp. 345-363.

Quade, E.S. (1975), *Analysis for Public Decisions*, New York, American Elsevier Publishing Co., Inc.

Ravetz, J.R. (1973), *Scientific Knowledge and its Social Problems*, Harmondsworth, U.K., Penguin Books.

Reichman, W.J. (1961), *Use and Abuse of Statistics*, Harmondsworth, U.K., Penguin Books.

Rein, M., and D.A. Schon (1976), Problem Setting in Policy Research, Cambridge, Massachusetts, M.I.T. Department of Urban Studies and Planning, mimeo.

Robinson, W.S. (1950), Ecological Correlations and the Behavior of Individuals, *American Sociological Review*, vol. 15, pp. 351-357.

Shubik, M., and G.D. Brewer (1972), *Models, Simulations, and Games -- A Survey*, Santa Monica: The Rand Corporation, R-1060-ARPA/RC.

Stebbing, S. (1939), *Thinking to Some Purpose*, Harmondsworth, U.K., Penguin Books.

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.

Streissler, E.W. (1970), *Pitfalls in Econometric Forecasting*, London, The Institute of Economic Affairs.

Thouless, R.H. (1947), *How to Think Straight*, New York, Simon and Schuster.

Toulmin, S. (1964), *The Uses of Argument*, Cambridge, Cambridge University Press.

Wagner, H.M. (1975), *Principles of Operations Research*, 2nd edition, Englewood Cliffs, New Jersey, Prentice Hall Inc.

Wallis, W.A., and H.V. Roberts (1956), *Statistics: A New Approach*, Glencoe, Illinois, The Free Press.

Wohlstetter, A. (1970), Analysis and Design of Conflict Systems, in *Analysis for Military Decisions*, E.S. Quade, ed., Amsterdam, North Holland Publishing Company.