VOLUME II


EDITED BY
G. CHEN
J. M. JAMIESON
L. L. SCHKADE
C. H. SMITH

THE SYSTEMS INQUIRY SERIES
INTERSYSTEMS PUBLICATIONS
\[
\frac{40}{14} \times \frac{1}{4} = \frac{160}{56}
\]

VOLUME II

EDITED BY
G. CHEN
J. M. JAMIESON
L. L. SCHKADE
C. H. SMITH

PUBLISHED BY
INTERSYSTEMS in the
SYSTEMS INQUIRY SERIES
Systems inquiry is grounded in a philosophical base of a systems view of the world. It has formulated theoretical postulates, conceptual images and paradigms, and developed strategies and tools of systems technology. Systems inquiry is both conclusion oriented (knowledge production) and decision oriented (knowledge utilization). It uses both analytic and synthetic modes of thinking and it enables us to understand and work with ever increasing complexities that surround us and which we are part of.

The series aims to encompass all three domains of systems inquiry: systems philosophy, systems theory and systems technology. Contributions introduced in the series may focus on any one or combinations of these domains or develop and explain relationships among domains and thus portray the systemic nature of systems inquiry.

Five kinds of presentations are considered in the series:
1. original work by single author,
2. edited compendium organized around a common theme,
3. edited proceedings of symposia or colloquy,
4. translations from the original works, and
5. out of print works of special significance.

Appearing in high quality paperback format, books in the series will be very moderately priced in order to make them accessible to the various publics who have an interest in or are involved in the systems movement.
# CONTENTS

## VOLUME I

### PREFACE
1

### SECTION I  THE GENERAL SYSTEMS APPROACH

A. INTRODUCTION
1

B. GENERAL SYSTEMS THEORY

C. INTERDISCIPLINARY SYSTEMS ANALYSIS AND RESEARCH

D. LEVELS OF ANALYSIS

### SECTION II  ON THE NATURE OF SYSTEMS

A. INTRODUCTION
69

B. SYSTEMS DEFINITION AND CONCEPTS
   - Hall, A.D. and Fagen, R.E.: *Definition of a System*, Vol. 1. / 75

C. THE PURPOSES, STRUCTURES AND PROCESSES OF SYSTEMS
Maruyama, Magoroh: *Metaorganization of Information*, Vol. 11. / 257

SECTION III ON THE NATURE OF SOCIAL SYSTEMS

A. INTRODUCTION / 279

B. SOCIAL SYSTEM DEFINITION AND CONCEPTS
   Bakke, E. Wight: *Concept of the Social Organization*, Vol. 4. / 289

C. THE PURPOSES, STRUCTURES AND PROCESSES OF SOCIAL SYSTEMS
   Haberstroh, Chadwick J. and Gerwin, Donald: *Climate Factors and the Decision Process*, Vol. 17. / 335
   Marschak, Jacob: *Efficient & Viable Organizational Forms*, Vol. 4. / 375

VOLUME II

SECTION IV THE ANALYSIS, DESIGN AND MANAGEMENT OF COMPLEX SYSTEMS

A. INTRODUCTION / 399

B. APPROACHES TO ANALYSIS

C. APPLICABILITY AND LIMITATIONS OF ANALYSIS
D. APPROACHES TO ORGANIZATIONAL DESIGN

E. STRESS IN ORGANIZATIONS

F. HEALTH AND STABILITY OF ORGANIZATIONS

SECTION V EXAMPLES OF APPLICATION OF SYSTEMS THEORY TO COMPLEX SOCIAL SYSTEMS

A. INTRODUCTION / 553

B. HUMAN BEHAVIOR IN LARGE ORGANIZATIONS
Argyris, Chris: *Understanding Organizational Change*, Vol. 4. / 559

C. APPLICATION OF SYSTEMS THEORY TO INSTITUTIONS
Howland, Daniel: *A Hospital System Model*, Vol. 9. / 597

D. APPLICATION OF SYSTEMS THEORY TO REGIONAL CONCERNS

E. APPLICATION OF SYSTEMS THEORY TO NATIONAL PROBLEMS

F. APPLICATION OF SYSTEM THEORY TO EMPIRES

G. APPLICATION OF SYSTEMS THEORY TO WORLD PROBLEMS
Intriligator, Michael D.: Some Simple Models of Arms Races, Vol. 9. / 697
Davis, R.H., Carpenter, P.B. and Missler, C.W.: A Game for Studying the Problems of Arms Control, Vol. 8. / 703
A. INTRODUCTION

Section I dealt primarily with the analysis of interrelationships in the real world, using the general systems approach as a tool. Section II may be summarized as a set of articles dealing with the nature of systems—the analysis and synthesis of any system, leading to management of systems. Section III focuses on social organizations, stressing their dynamics of adaptation. Again, analysis, synthesis and management are the underlying themes. Section IV concentrates on some of the complexities of social organizations and factors which should be considered in the analysis and design of systems.

The section is divided into five parts in addition to the introduction. The two papers in part (B) review the methods available to the analyst and synthesizer, briefly review the history of these methods and provide an application in a particular sector. Part (C) deals with the usefulness, limitations and applicability of analytical methods while part (D) is concerned with approaches to the design of systems particularly the organizational aspects. Part (E) considers the various causes of stress in an individual and stress in the organization. Finally, part (F) deals with factors necessary for the health and stability of organizations.

In part (B) Sutherland states that the general system theorist recognizes only one epistemological\(^1\) dictate:

"it is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits; it is evidently equally foolish to accept probable reasoning from a mathematician and to demand from a rhetorician rigorous proofs" (emphasis added).

This dictate recognizes the two extreme or hard-line positions in inquiry into problem situations. The hard-line empiricist holds that the scientific method for the "hard" or physical sciences is totally transferable to any problem in the behavioural or social sciences. The hard-line deductivist or theoretician, on the other hand, holds that there are considerable differences between the subjects of the behavioural sciences and those of the physical sciences; this implies the need for a separate scientific methodology for the "soft" or behavioural sciences.

The articles in Part IV demonstrate that the truth lies somewhere in between, depending on the

\(^1\)The branch of philosophy dealing with theories of knowledge. Thus, GST recognizes only the Aristotelian-based theory.
particular problem situation. This in turn demonstrates that, while different general system theorists
lean to one end or the other of the continuum between the empirical and theoretical extremes, in ac­
cordance with their own training and experience, the field of general systems theory follows
Aristotle's epistemological dictate.

Sutherland claims to talk only of the analysis process. In fact, however, he demonstrates the
intertwining of analysis and synthesis. Synthesis or design of systems for particular purposes may be
stated as the output of the research system which examines a particular problem situation. Analysis
may be thought of as the input of the research system selected to produce the desired output. The
design output should include a regulator which controls the object system (the real world system
being looked at) via feedback linkages, etc. towards a predefined goal. This goal and its attendant
constraints may change if the environment and the object system are unstable (see Part II). Thus we
have an existing object system which is regulated or maintained in some manner. When this regula­
tion is deemed unsatisfactory, a research system should be developed to improve it or decline and
breakdown may result.

Sutherland talks of various information sets which occur in the research process:

1. **A Priori Information**: the stock of knowledge whose roots are found in theory, intuition
and prior empirical findings (i.e., experience), which the researcher initially deems useful to the prob­
lem situation. This knowledge is selected without incurring any current analytical expenditures.

2. **A Posteriori Information**: selected additions to or deletions from the a priori stock of
knowledge (modified as the research process continues), plus empirical information deliberately
sought for the particular problem at hand. This information, and its systematic assemblage with
the selected a priori knowledge, has current costs attached to it.

3. **Real Information**: an abstraction which represents our knowledge as to the feasible
limits of (1) and (2) in producing problem-relevant information. It represents a yardstick against
which the quality of the a priori and a posteriori information can be measured at any point in
the process. Of course, this quality measurement exercise also has attendant current costs.

Sutherland then goes on to deal with costs and benefits of additional research, to provide a
general problem-solving paradigm, and to categorize different problem types (from assumptions of
complete determinism to complete indeterminism) and their effects on the research system’s effi­
ciency and effectiveness. He concludes that all of the instruments in the scientific arsenal, from
theoretical to empirical, from “hard” sciences to “soft,” are appropriate at different times in the
research process. Their use depends on the particular problem-type.

Howland provides a partial history of General Systems Theory (GST) in a more “concrete”
health setting, discussing **analysis, design and management of medical facilities** through a “hard”
man-machine systems paradigm.

Starting with “scientific management” in the 1940s, he discusses briefly industrial engineer­
ing, human engineering, operations research and systems research.
Difficulties with each of these approaches led Howland to a set of concepts based on cybernetics, where the operational objective of his designed system is to provide patient care.

In part (C) Arrow deals with mathematical models at a more abstract level, discussing the scope, utility and limitations of mathematical models in the social sciences. His position is hard-line:

“any meaningful proposition can be expressed in a suitable mathematical form, and any generalizations about social behaviour can be formulated mathematically.”

In accordance with our knowledge of logic, his proposition is irrefutable. However, the philosophical question as to whether this knowledge of logic is complete is still open. Nevertheless, mathematical models have much to offer. It is because mathematical models are “contentless” (abstract, no empirical real world ties, no semantical or “meaning” linkages) that they may well be universally useful (to date, the only social science to use these models extensively is economics).

Their usefulness in adding clarity via logical deductions is undisputed, in that they formalize our way of thinking about real world (object) systems. Furthermore,

“it is by now a platitude of the scientific method that if theory without empirical evidence is unreliable, empirical inquiry without theoretical background is unfruitful.”

This leads us to a basic consideration of the simplest form of tests for our appropriate mathematical model for any given object system.

We need to:

1. empirically test the assumptions and results or conclusions of the model as well as
2. logically test the internal consistency of the model.

Contrary to the popular song by “Meatloaf,” two out of three is not sufficient; all three are necessary, and more is required to make these tests sufficient.

Arrow next discusses some classifications of models, wherein the most interesting may be his dichotomy of “individualistic” models (dealing with the behaviour of individuals) and “collective” models (dealing with the behaviour of groups). His own findings are at present generally accepted, that it is impossible to generate a social welfare function from individual utility functions. Again, however, the philosophical basis of his work may or may not hold up in time.

Another interesting aspect of Arrow’s discussion concerns the “principle of rationality.” Under this hypothesis:

1hypothesis = a proposed law, yet to be fully tested.
law = a principle held to be true by the scientific community but still tentative.
“the individual’s behaviour depends on his tastes— as expressed, say, by a utility index—and upon the obstacles, which are determined by exogenous factors, by the actions of others, and possibly by past actions of the individual and of others. We may say that the individual maximizes his utility, subject to the obstacles.”

Now the individual’s utility, representing his tastes, may be expressed either as his wants or his needs. The definition of either of these for a particular individual has proved to be a stumbling block.

The alternative position to the principle of rationality involves the development of ad hoc hypotheses, usually drawn from introspection or casual observation and dependent upon the researcher’s intuition and common sense. While this method allows one a step outside the formal boundary of the principle of rationality, it tends to lead to a profusion of contradictory statements. Again, some combination of rational and ad hoc procedures is probably appropriate as a methodology for the analysis, design and management of a particular object system.

Arrow’s development of the theory of games as mathematical models leads us to consider Ackoff’s paper.

"there is already evidence that, in the systems revolution modest results tend to be excessively generalized and that assumptions once stated— if stated at all— tend to be ignored in defining the realm to which results can be applied."

According to Ackoff, a person or group can be said to be in a problem situation under the following assumptions:

1. he has one or more unsatisfied wants or needs,
2. he has available alternative ways of pursuing the objective of satisfying the wants or needs, and is uncertain as to the relative effectiveness of the alternatives.

Game theory deals with this problem type, in that utilities or valued outcomes are involved and that alternative strategies are available.1 In addition, however, the tool of game theory requires several additional assumptions:

3. the set of possible strategies can be specified in advance,
4. a set of well-defined end-states exists,
5. a specified payoff is associated with each end-state.

Various criticisms have been leveled at game theory. One or more of the above conditions may not be satisfied. Mathematical solutions only exist for relatively simple game theoretic problem situations. Utilities are not transferable; this means that two or more players may formulate the same problem situation into two or more different payoff matrices. The payoffs may not be independent; that is, given the selection of a mixed strategy, the payoffs associated with any one

---

1These alternative strategies must be unequally effective to have a real choice. However, this information is not often available a priori.
strategy may depend on the frequency of choice of the other. This latter criticism has implications for the choice of a philosophy. A single philosophy is more predictable than a mixed one.

Ackoff further criticizes the implicit assumption of complete ignorance (on the part of the player) of the probabilities associated with the competitor's choices, by pointing out that in a real world problem situation the decision-maker is not given a game to play but must extract it out of the situation itself. Problem formulation and the selection of a solution are not separable aspects of problem-solving.

Ackoff, then, appreciates the precise formulation of "game-theoretic games" but heavily criticizes their unreality. On the other hand, he appreciates the partial reality of "operational games" such as military or business games but criticizes their lack of precise formulation and use.

Let us digress for a moment to summarize some of the major themes. A decision-maker or manager of a real-world system is faced with a problem-situation. He has certain a priori and potential information resources, either personally or through an agent. These resources include empirical and theoretical instruments which he or his agent can combine into a research system with the objective of analyzing, designing or synthesizing, and implementing as a modification of his real world system. He will so modify this object system if the costs and associated benefits are appropriate.

An initial comment by Ackoff is relevant to Rice's article (Part D).

"Until recently, scientists and engineers tended to treat systems as complexes whose output could be expressed as a single function of the outputs of the component parts. As a consequence, systems were designed from the inside out. Increasingly, researchers have come to deal with systems whose output cannot be expressed as a simple function of component outputs and it has become more productive to treat them holistically and to design them from the outside in."

In Section III, Bakke suggested seven services which an adequate concept of "social organization" can perform, of use to both researchers and practitioners. One of these services is to provide a systemic framework of the major parts of a social organization such that more elementary variables (selected aspects of elements of the object system) and more specific theories with respect to their relationships can be derived. In Part (D) of this Section Rice defines a "research model" (or "research system") as a strategy which helps in the selection and measurement of the system variables and relationships we wish to study. He develops in preliminary form a research system to deal with performance measurement of the contributions of organizations such as hospitals, schools and prisons to society. The output variables chosen are "care of patients," "protection" (of patients and hospitals), "social restoration" of patients, "training and education" of hospital personnel, scientific "research" into hospitals and the mental health of patients, and "administration" of the hospital systems in society.

His input variable comprise four classes of parameters: community type, measurable features
of patients entering the hospital, aspects of direct community participation in the hospital’s functioning, and other basic resources received by the hospital such as financing. His system variables, which can be used to describe the state of the hospital system at any time, include three classes of parameters: physical characteristics of the hospital system, staffing pattern, and policies and procedures regarding the hospital’s functioning as a going concern rather than its structure. Relationships among variables are dealt with using statistical correlation, where input and system measures (predictors) are statistically correlated with output measures (criterion variables).

The differences between Rice’s research system and those using an “operation research” (O.R.) methodology are stated as follows:

1. The scope of O.R. is typically a single organization, whereas that of Rice’s system is a class of organizations studied simultaneously.

2. O.R. typically first develops a model to represent system functioning, then collects empirical data to apply the model to a particular object system; Rice calls for the collection of data first, on which model construction is based.

Studer advances further into the field of collection of data first. He first postulates the difference between “ape” and “man.” Apes are seen to have no capacity to plan or design something “better;” man is seen to have total capacity to plan. A major human problem, then, is how to get from “IS” to “OUGHT”1 via planning or designing. We need then to hold on to “OUGHT” via control (regulation and maintenance, avoiding breakdown). He recommends an innovative, experimental approach to both individual and collective systems based on “directed behaviour changes” and leads us to (Skinner’s and) Mitroff and Sagasti’s world.

Mitroff and Sagasti deal conceptually, as opposed2 to technically, with “stimulus-response (S/R)” aspects of inputs and outputs of systems. They raise the question whether the philosophy of the physical sciences (scientific methodology in the “hard” sciences) can be transformed wholesale to the behavioural or social sciences, and even if it can, how?

Their major reference seems to be Churchman,3 their self-imposed task to conceptualize behavioural science experiments using philosophical analysis and systems theory to provide a new perspective. They provide a taxonomy of experimental problem (decision-making) situations where the inquiring system (IS) is both stimulus and response, concluding this is only one way of approaching S/R, with different systems appropriate to different situations.

The major theme of these writings is that their conclusion can be generalized further: GST is only one way of approaching problem situations, albeit a most promising one; within the GST

---

1 “IS” describes where we are; “OUGHT” prescribes where we want to be.

2 A major theme underlying GST is that dichotomies are only an analytical device representing continua along a dimension.

system of knowledge, either polar position or a point on the continuum between them may be appropriate to a particular problem situation.

In today's organizations, psychological stress is being recognized as a major health problem. In part (E) Horvath reviews some definitions of the concept of stress and some experimental research deemed relevant. He does not cover all research which is called an investigation of stress by its author. Furthermore, the research covered deals with stress in the individual rather than a large and complex group representing an organization. Arrow has (elsewhere) demonstrated the impossibility of aggregating individual utilities into a social welfare function. However, there may be some lessons in this individual stress research for a preliminary concept of organizational stress.

Looking at the "hardest" physical science, physics, Horvath finds general agreement about the concept of stress. "Stress" is the (input) force applied to an object (system), and "strain" is the resultant change (output) in the object system's dimensions.

Horvath then adopts and adapts Hans Selye's terminology due to his general acceptance among physiologists and partial acceptance among psychologists, both representing the "soft" sciences. The term "stressor" represents input stimuli eliciting stress, where "stress" refers to a state of the organism and "stress response" to an output response which characteristic of or a reflection of this state of stress (e.g., a fear or sex or hunger response). Thus, in outline at least, the "soft" and "hard" definitions are parallel.

After further considerations, Horvath derives an operational definition of stress deemed appropriate to psychology.

"Psychological stress is a state which occurs when an individual is subjected to conditions which disturb or threaten to disturb crucial psychological variables from within their normal limits."

An immediate problem here is that the definition is static: "normal limits" may expand or contract with time. This definition, however, does fit the various definitions of stress in psychology, according to Horvath, and he suggests that lessons may be learnt from physiology. Since these "soft" sciences study human organisms (i.e., a subset of those organisms which are alive), this conceptual research may also be appropriate to the study of organizations (assemblages of organisms which are alive).

Horvath examines the psychological literature and concludes that certain stress variables leading to certain organism behaviour are particularly sensitive to certain stressors as stimuli. His most generally descriptive stress variable is "functional limits." After measuring performance limits in some manner under neutral conditions, stressors may then be applied below (underload), at, or above (overload) the functional limits.

A second stress variable, which is useful in controlling the nuisance variable "motivation," is "level of attention." If an organism under controlled conditions has already been performing in a task requiring a high level of attention, he is less able to compensate by higher motivation, leading
to a higher level of attention when he is stressed.

Horvath’s third stress variable is “focus of attention,” another aspect of attention processes. Here, stress may reduce the focus and the ability to control the focus.

A fourth stress variable is “strength of habit” or “level of learning.” This variable reflects the notion that stronger habits or overlearned responses are less easily disrupted by stress than weaker habits or underlearned responses.

The fifth variable derived by Horvath is related to the fourth. “Reversal of habit or set” is a variable descriptive of tests having the common characteristic of flexibility needed to reverse a set or habit.

These, then, are the five most promising stress variables found by Horvath. Stressors may be tasks; or motivational arousers such as fear, frustration or anxiety; or distracting, irrelevant or novel stimuli; or disruptors of physiological homeostasis such as certain drugs or fatigue. These stressors and their related stress variables are emphasized by Horvath as preliminary findings only.

Gray asks the questions, “Does Johnny’s psychosis, or psychosomatic illness, or depression, serve a homeostatic function in the family system, the group system, or the network system? Should we use a systems perspective to focus our intervention on changing the family, group, or marital system, and if so, will the individual’s symptoms disappear?” Gray’s answer is a resounding “Yes,” as he claims the class of therapists—family, group, and transactional—will tell us. Herein lies his basic case for the application of GST in psychiatry, in particular to derive a general theory of personality. His thesis is epistemologically correct in a GST sense: changes of state from inputs within the system are certainly important, as are changes of state from inputs from the system’s environment. In particular, he makes the not unreasonable claim that intrapsychic inputs need to be studied as well as extrapsychic inputs, to explain an organism’s behaviour. Thus, his call is for a general system’s view of personality, the measurement of which is the most difficult in psychology.¹

One must wonder, for example, about what exists in an individual’s makeup to cause “fluid intelligence,” which is exhibited by performances characterized by adaptation to new situations, the “fluid” application of general ability, so to speak. On the other hand, what causes “crystallized intelligence,” another aspect of (or variable within) personality, where this form of intelligence is exhibited by cognitive performances in which “skilled judgment habits” have become fixed or crystallized.² And what, if anything, does this have to do with Newton’s experience of mixed inputs (an external input of an apple falling on his head, and an internal input of some quality, X, which we sometimes call intuition) causing him to draw the conclusion that bodies having mass exert some kind of pull towards other bodies?

²Ibid., pp. 682-3
Vickers deals, as have others, with the personality aspect termed "stress." He claims that the stability of a human system determines whether the application of a stressor will cause the system to break down or (merely) to function badly. If the system does break down, will it become disorganized (exhibiting random behaviour) or will it go through some introspective process of protective reorganization? The latter implies some movement, perhaps expansion, of Horwath's "normal limits," which may in turn be partially measured via his stress variable "functional limits." This in turn reminds one of Studer's postulate that humans, through learning about and applying planning, strive to reach and maintain the "man" state. In an organizational sense, Vickers, in Part II, discusses the function of management as:

1. trying to observe and work within the laws or principles involved in the self-regulation of the organization, and

2. attempting to intervene as a "helmsman" or "steersman" in order to influence the system's choice of action towards certain goals.

Part (F) examines this notion of planning for a future a little more closely, in the context of stable and unstable equilibrium. Smith quotes Tagore that "worries come only if you believe in a future. Believing in ghosts, you are freed from burden, all the worries enter the ghost's head." This fanciful analogy serves as an introduction to his plea for fusion of seemingly opposed forces, group dynamics and project management, where group dynamics is an interdisciplinary cluster of many behavioural scientists and project management is an interdisciplinary cluster of physical sciences. And so we are back to Sutherland's "square one" and Aristotle's dictum from over 2400 years ago. General Systems Theorists are supposed to recognize this dictum, which in essence implies a need to fuse the so-called "hard" and "soft" sciences if we expect to solve real world problems. However, these forces of "soft" and "hard" appear to want to dominate or sit at "the top" in science.

Let us consider another somewhat fanciful analogy, which may of course be an incorrect analogy, as may Smith's. Consider a hollow sphere and two marbles representing the "soft" and the "hard" sciences. Place the two marbles on the top of the sphere, which itself is fixed in position by some magnetic forces, say. The marbles, unless held by some complex magnetic forces, will fall off the sphere, exhibiting unstable equilibrium, and fission.

Now place the marbles within the sphere, in any position whatsoever. They will fall to the bottom of the sphere and "fuse," exhibiting stable equilibrium and homeostasis, which will be maintained unless external forces manage to overcome internal compensatory mechanisms and inputs from the system's environment representing order (see Young in Part III).

It may be helpful at this point to recall some notions as to whether an analogy is correct or incorrect. In Part III, the functions of a formal analogue were described as twofold: to teach and to extend theory. The latter is accomplished by what Mary Hesse calls "neutral analogies." It is up to the reader to decide whether the above analogy is "positive" or "negative," whether the "soft" and "hard" sciences can fuse into stability or will exhibit fission and instability.
Let us now use Bennis' article on organizational health to provide a link from analysis of complex social systems to synthesis or design or planning for complex social systems. Among other things, Bennis discusses two types of communication networks for problem-solving by a group of five persons, the "wheel" and the "circle." Consider his diagram, where o represents individual and relation.

![Diagram of "wheel" and "circle"](image)

Despite the far greater number of communication linkages (ten as opposed to four), Bennis states that experiments show more rapid acceptance of a new idea is possible with the circle than with the wheel. The man in the middle of the wheel is apt to discard an idea on the grounds that he is too busy or the idea is impractical. Furthermore, when the task is changed, the circle organization is better able to adapt to this change by developing a new code. In addition, and this brings us back to the notion of fusion versus fission of the "soft" and the "hard" sciences, a changing leadership in accordance with the unique leadership qualities of each member may be the best authority group strategy.

Nevertheless, as stated earlier, all hypotheses are "merely" tentative laws, and all scientific laws are "merely" tentative; we are told that the entire scientific community once regarded the earth as flat. Is it?

Whichever leadership strategy is adopted by the scientific community of general system theorists is yet to be seen.
B.

APPROACHES TO ANALYSIS
BEYOND SYSTEM ENGINEERING: THE GENERAL SYSTEM THEORY
POTENTIAL FOR SOCIAL SCIENCE SYSTEM ANALYSIS

John W. Sutherland

INTRODUCTION

There is much argument in the scientific community about the extent to which the social sciences' methodological practices reflect the conditions for objective, accurate inquiry. Even within the social sciences, there are two polar positions: the dictate from those who prefer empirical predictions that the proper instruments of science are the instruments of the natural sciences, and that the proper output of scientific analysis are nomothetic constructs; on the other side, the grand theory builders whose analytical preference is largely deductive and whose analytical results are generally idiographic or theoretical in nature. The position of the former is defended by trying to define social science subjects in terms of the phenomena dealt with by physics and the other natural sciences; the deductivist's position, on the other hand, reflects his opinion that there are considerable differences between the subjects of the social sciences and those traditionally dealt with by the "hard" disciplines.

Now, in the perspective of the general system theorist, neither of these polaristic positions have an exclusive option on the methodology of the social sciences. Rather, he recognizes only one epistemological dictate, that lent us by Aristotle over 2400 years ago:

"... it is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits. It is evidently equally foolish to accept probable reasoning from a mathematician and to demand from a rhetorician rigorous proofs."

In short, the properties of the phenomenon at hand should determine the parameters of the analytical approach—not any a prioristic preferences we happen to hold. To make this dictate somewhat more operational, we shall generate an array of four analytical ideal-types, approximative of the range of problems the social scientist may expect to encounter, and use these ideal-types to engineer congruence between problem properties and methodological platforms.

I. THE ANALYSIS PROCESS

The applied scientist is primarily a model builder. The models he builds are intended to capture the properties of real-world phenomena which we wish to describe, predict, or reconstruct. These models may, moreover, take several different forms: mathematical, statistical, metaphorical, physical (i.e., simulative), or some combination of these modalities. At any rate, the engine which drives models and provides their structure is information, and information is the output from an analysis exercise. In short, the scientist engages in research and analysis in order to generate the information necessary for the construction of accurate models of predictive or descriptive significance.

At any rate, we shall be concerned about information in three guises:

(a) There is, first of all, the stock of a priori information. This is the knowledge we possess (or think we possess) about a problem which exists prior to the inauguration of a formal, dedicated analysis process. That is, it is knowledge whose roots are to be found in experience, theory, intuition, or other non-empirical predications. Thus were we to set the analysis process in a time framework, the a priori stock of information would be that which we possess at time $t_0$. That is, it is information we have obtained without incurring any current analytical expenditures or without undertaking any current empirical observations.

(b) Secondly, there is the succession of a posteriori information stocks. These are the aggregations of information which have been acquired by iterations of analyses aimed at empirically capturing the properties of the problem at hand or, more generally, products of incremental expenditures of analytical resources. Thus, if the

1. Because of this lack of serious empiricializational, the social sciences have perpetuated an embarrassing schism, marked by acrimony and recrimination between what we will later refer to as positivists and idiographic model-builders. In his Social Theory and Social Structure (Free Press, 1968 edition, p. 139), Robert Merton cites the polar ends of the schism: "On the one hand we observe those sociologists who seek above all to generalize, to find their way as rapidly as possible to the formulation of sociological laws. Tending to assess the significance of sociological work in terms of scope rather than demonstrability of generalizations, they eschew the 'triviality' of detailed, small-scale observations and seek the grandeur of global summaries. At the other extreme stands a hardy band who do not hunt too closely to the implications of their research but who remain confident and assured that what they report is so . . . their reports of facts are verifiable and often verified, but they are somewhat at a loss to relate these facts to one another or even explain why these, rather than other, observations have been made."
a priori information stock exists at time \( t^j \) in the analysis process, the a posteriori stocks are those which exist at times \( t_1 \ldots t_n \), and each has a direct expenditure of analytical resources associated with its generation.

(c) Finally, there is a third information stock of interest to us: what we will call the real. This is an abstraction which both the a priori and a posteriori stocks must approximate, for it houses our expectations about the limits of information available about the entity under investigation. That is, it serves as a reference point against which we can measure the relative quality of the a priori and a posteriori information stocks at some point in the analysis process. Moreover, our concept of the limits of information available to us about the problem at hand, our concept of what the ideal information stock might look like, may have several different origins:

- it may result from past experience with similar problems;
- it may simply be an intuitionalistic formulation without empirical foundation;
- or it may be a product of deductive inference such that its properties are generated as the result of the application of some theoretical base to the current problem—a technique we shall discuss in detail in another section of this work.

At any rate, the real information stock represents our expectations about what can be achieved by the analysis process in its role as a producer of model-oriented information.

The analysis process, then, is the vehicle by which we attempt to move from a less favorable a priori informational state at time \( t^j \) to more favorable ones at times \( t_1 \ldots t_n \); and time \( t_n \) may, for the moment, be thought of as the point where we feel that the currently existing a posteriori information stock has adequately exhausted the postulated properties of the real information stock. Graphically, the process looks like this (Figure 1):

![Fig. 1](image)

This is what we shall refer to as the normative learning curve: it shows that increased expenditures of analytical resources (time; energy, computational cost, man-hours, etc.) are expected to result in a more accurate model in either of two ways:

- if our ambition for the model we are building is predictive (or if we are trying to causally reconstruct some historical event), then a reduction in error is realized when events predicted by the model and events which actually occur are identical—or to the extent that the variance between predicted and actual events declines;
- if the model we are building is intended to serve as a descriptive allegory, then error declines to the extent that we realize a morphological correlation between the real-world entity and the properties of the model.

But we will have considerable difficulty in dealing with the concept of predictive/descriptive error within the confines of the analysis process itself. Specifically, if we were to devise some sort of objective function for system analysts to operate under, it might read like this: Minimize the error component of a model! However, this is a flat and usually non-directive imperative. There are several things wrong with it. First, there are some problems which must be attacked which simply do not warrant much expenditure of time, energy, or effort in the attempt toward ultimate rectitude (i.e., those which, even if we are considerably wrong about them, will carry no significant loss for us). Thus we must insert the concept of economic marginality and suggest that the system analyst may better attempt to minimize error subject to the point where the expenditures for information are lower than the expected value of the information.

But this leads us directly to a seemingly impossible problem: the assessment of the value of information during the analysis process. That is, how may we begin to get some estimates of the economy of continuing the analysis process in an effort to concatenatively lower the error component associated with the model we are building? At what point does the acquisition of further increments of information become gratuitous or marginally unproductive?

This problem approaches impossibility, but not entirely. Initially, if the problem at hand is a precedent new one, such that we have had empirical association with it at some prior time, then the real information stock may be set to reflect this experience, and the analysis process would be continued until the a posteriori state at some time \( t_i \) was perceived to have exhausted the real. But, for the unprecedented problem, or one which is sufficiently dissimilar from previous experiences to make dubious a real state set as a product of inductive inference, we must take recourse in some deductive construct.

Initially, let's consider that almost any real world phenomenon we encounter will be admissible to modelling on four basic dimensions or levels of analysis:
(a) The state-variable level: here we try to identify the major structural or qualitative aspects of the entity at hand... the determinants of the problem.

(b) The parametric level: this involves the assignment of specific numerical (or precise qualitative) values to the state-variables.

(c) The relational level: this involves establishing the nature of the relationships among the state-variables (i.e., the array of interfaces and the directions of influence).

(d) The coefficient level: here we assign specific numerical or magnitudinal values to the state-variable interfaces.

Thus, as we set out to build a model to predict or allegorize some problem (phenomenon) of interest, the information we generate during the analysis process will be directed at forcing a congruence between the phenomenon and the model on these several dimensions. Again within the confines of the normative analysis process, the difference between the a priori and the a posteriori states on any of these levels of analysis becomes intelligible in terms of the following construct (Figure 2):

Now, this event-array which forms the base for the probability distributions may take on several different implications within the context of the model-building exercise. Here, however, we want to restrict it to one of several possibilities:

- the set of alternative values which a specific state variable may assume;
- the set of alternative coefficient values which a particular interface (i.e., set of state-variables) may assume; that is, a specific magnitude of interrelationship;
- the array of alternative sets of state-variables we might employ to establish the structural aspects of an entity—alternative sets of determinants proposed as fully exhausting the structural properties of the phenomenon we are treating;
- the array of alternative relationship conditions which a specific entity-interface (i.e., a specific set of purportedly interrelated state-variables) might evoke.

Then the central point of Figure 2 is this: the a posteriori event-probability distribution represents a more favorable informational condition than the a priori because (a) some events which were assigned positive probabilities of occurrence in the a priori phase have been eliminated from consideration in the a posteriori; (b) there is a more intense or localized density of probabilities for the remaining events associated with the a posteriori distribution than with the a priori. In short, the a posteriori distribution entails much less variance than the a priori, and therefore represents a situation of greater assurity about the event which will occur. In short, the expected value of an error in assignment from the real world entity to the model is lower in the a posteriori than the a priori case. By way of illustration, suppose that we faced at some point in time the problem of estimating the magnitude of the relationship between two variables (determinants) of a problem we are attempting to allegorize—the problem of estimating a coefficient. In the a priori phase, before the application of formal analysis and prior to the inauguration of empirical observations, let us say that we have only the vaguest notion of the value which might be assumed. More specifically, we know with assurity only that the value which the coefficient may assume may not be lower than x nor greater than y. Thus the a priori event-probability distribution we would establish would be similar to that of the diagram: x and y would be assigned lowest positive probabilities of occurrence and would, therefore, rest at the tails of the distribution. Now, in order to start formal analysis with the greatest possible variance allowed for in the a priori event-probability distribution, we might then generate a normal curve with a slight graduation of densities toward the value which sits midway between x and y. Once we begin empirical observations of the state variables in actual relationship, we would expect the a priori distribution (wide and flat) to be transformed gradually into one which converges on some value that is most recurrent in the empirical examinations. When the event-probability distribution spawned by successive observations begins to find a "limit" at some value, and when successive observations do not yield significantly different estimates, we can assume that we have pretty well exhausted the informational potential associated with this particular aspect of the entity we are modelling. In short, the value which we will then enter into our emerging model for the particular interface under investigation is that converged on by the successive a posteriori distributions. In the case of the diagram we have used, this turns out to be the central estimate—the mid-point between x and y. And we did this, not by coincidence, but to make a critical point: even if the most likely estimates yielded by the a priori and a posteriori probability distributions are the same, the a posteriori is still more favorable, for we are more certain about its prescriptions than we were with the a priori. So, was the expenditure of time, dollars, and effort required to produce the a posteriori concatenation wasted? Not from the standpoint of the...
JOHN W. SUTHERLAND

system scientist, because, for him, the value of information is directly related to its ability to cause just such a convergence: a reduction in the potential variance or error associated with an estimator to be entered into a model purporting to predict and/or describe a real world phenomenon.

This point may not be immediately clear, but it is critical in terms of the perspective it lends us. First, if a model is comprised of many such estimation processes, with each estimator entered carrying with it a specific index of accuracy (i.e., a probability of occurrence), then the aggregate index of expected error we would associate with the model as a whole would be some product of these error estimates associated with its components. Thus, by acting to minimize the expected error associated with each component of the model, we also act to decrease the variance or error we expect to be associated with the model as a whole (and this becomes especially critical when we recognize that errors in an integrated model tend to behave concatenatively, not simply additively).

We have now one more step to take before approaching the problem of estimating the value of information in the confines of the analysis process. It consists basically in defining the concept of the expected value of an error. This is a bivariate function, the product of the following factors: the probability of an error of a given magnitude occurring; the absolute cost (or loss) expected to be incurred should the error be realized.

Although there are other dimensions we might use under special circumstances, we try to estimate the cost or loss associated with errors we might make in terms of some standard unit, like dollars. Thus, were we building a model to predict demand levels for some product we are marketing, or trying to develop a model which would estimate the number of case workers we should hire and train to serve some casualty population, the relationship between errors of estimation in the entering of the model's components and the cost of those errors is direct: oversupply costs us money in two ways—the expenditures for product supplies (and/or personnel) which will not be used, and the opportunity loss, which must be calculated against profitable uses to which the expenditures could have been put were they not used to generate oversupplies.

The normative construct we would impose on these factors, then, is shown as Figure 3.

We can now suggest that, because expected loss is jointly a function of absolute cost of errors of certain magnitudes and the probability of incurring errors of those magnitudes, a reduction in the latter factor results in a direct reduction in the expected value of errors associated with a prediction or description problem. Or, more explicitly, the value of information generated via analysis process is imputed equal to the reduction in the expected cost of errors associated with its generation.

None of this says anything about the real or actual value of the information generated during an analysis process, i.e., how much was actually saved by conducting formal research, etc. The reason is a simple one; we cannot ever arrive at this value except after the problem has already ultimated such that the problem is now empirical rather than a target for analysis, per se. Thus, in a prediction exercise, our best estimators of what some true parameter will be, our best estimates of the nature of the relationship between two variables, are simply that: estimates. And, although we may have assigned them overwhelmingly significant probabilities of occurrence relative to all other alternatives we considered, our probability constructs do not engine the real world—the event which actually occurs may have little reference to that which even the most exhaustive pre-ultimation analysis process led us to expect would occur. Thus, when considering the value of information, we are forced to operate with expectations and use the abstract event-probability distributions to engine our decisions.

Simply, greater magnitudes of error carry with them greater absolute costs. Or, more specifically, as our best estimator for some event we are trying to predict is increasingly inaccurate, we must expect to incur successively greater losses if we act on the estimate. Thus it is more costly to oversupply a product by 1,000 units than by 100; similarly, it is more costly to have overestimated the demand for case workers by 10 than by 3, etc. As for the second factor, the probability of incurring an error of a given magnitude, we have already suggested that we want it to decrease with informational acquisitions. That is, information has its direct value in accomplishing the following (Figure 4).

We have now one more step to take before approaching the problem of estimating the value of information in the confines of the analysis process. It consists basically in defining the concept of the expected value of an error. This is a bivariate function, the product of the following factors: the probability of an error of a given magnitude occurring; the absolute cost (or loss) expected to be incurred should the error be realized.

Although there are other dimensions we might use under special circumstances, we try to estimate the cost or loss associated with errors we might make in terms of some standard unit, like dollars. Thus, were we building a model to predict demand levels for some product we are marketing, or trying to develop a model which would estimate the number of case workers we should hire and train to serve some casualty population, the relationship between errors of estimation in the entering of the model's components and the cost of those errors is direct: oversupply costs us money in two ways—the expenditures for product supplies (and/or personnel) which will not be used, and the opportunity loss, which must be calculated against profitable uses to which the expenditures could have been put were they not used to generate oversupplies.

The normative construct we would impose on these factors, then, is shown as Figure 3.

We can now suggest that, because expected loss is jointly a function of absolute cost of errors of certain magnitudes and the probability of incurring errors of those magnitudes, a reduction in the latter factor results in a direct reduction in the expected value of errors associated with a prediction or description problem. Or, more explicitly, the value of information generated via analysis process is imputed equal to the reduction in the expected cost of errors associated with its generation.

None of this says anything about the real or actual value of the information generated during an analysis process, i.e., how much was actually saved by conducting formal research, etc. The reason is a simple one; we cannot ever arrive at this value except after the problem has already ultimated such that the problem is now empirical rather than a target for analysis, per se. Thus, in a prediction exercise, our best estimators of what some true parameter will be, our best estimates of the nature of the relationship between two variables, are simply that: estimates. And, although we may have assigned them overwhelmingly significant probabilities of occurrence relative to all other alternatives we considered, our probability constructs do not engine the real world—the event which actually occurs may have little reference to that which even the most exhaustive pre-ultimation analysis process led us to expect would occur. Thus, when considering the value of information, we are forced to operate with expectations and use the abstract event-probability distributions to engine our decisions.

Simply, greater magnitudes of error carry with them greater absolute costs. Or, more specifically, as our best estimator for some event we are trying to predict is increasingly inaccurate, we must expect to incur successively greater losses if we act on the estimate. Thus it is more costly to oversupply a product by 1,000 units than by 100; similarly, it is more costly to have overestimated the demand for case workers by 10 than by 3, etc. As for the second factor, the probability of incurring an error of a given magnitude, we have already suggested that we want it to decrease with informational acquisitions. That is, information has its direct value in accomplishing the following (Figure 4).

We have now one more step to take before approaching the problem of estimating the value of information in the confines of the analysis process. It consists basically in defining the concept of the expected value of an error. This is a bivariate function, the product of the following factors: the probability of an error of a given magnitude occurring; the absolute cost (or loss) expected to be incurred should the error be realized.

Although there are other dimensions we might use under special circumstances, we try to estimate the cost or loss associated with errors we might make in terms of some standard unit, like dollars. Thus, were we building a model to predict demand levels for some product we are marketing, or trying to develop a model which would estimate the number of case workers we should hire and train to serve some casualty population, the relationship between errors of estimation in the entering of the model's components and the cost of those errors is direct: oversupply costs us money in two ways—the expenditures for product supplies (and/or personnel) which will not be used, and the opportunity loss, which must be calculated against profitable uses to which the expenditures could have been put were they not used to generate oversupplies.

The normative construct we would impose on these factors, then, is shown as Figure 3.

We can now suggest that, because expected loss is jointly a function of absolute cost of errors of certain magnitudes and the probability of incurring errors of those magnitudes, a reduction in the latter factor results in a direct reduction in the expected value of errors associated with a prediction or description problem. Or, more explicitly, the value of information generated via analysis process is imputed equal to the reduction in the expected cost of errors associated with its generation.

None of this says anything about the real or actual value of the information generated during an analysis process, i.e., how much was actually saved by conducting formal research, etc. The reason is a simple one; we cannot ever arrive at this value except after the problem has already ultimated such that the problem is now empirical rather than a target for analysis, per se. Thus, in a prediction exercise, our best estimators of what some true parameter will be, our best estimates of the nature of the relationship between two variables, are simply that: estimates. And, although we may have assigned them overwhelmingly significant probabilities of occurrence relative to all other alternatives we considered, our probability constructs do not engine the real world—the event which actually occurs may have little reference to that which even the most exhaustive pre-ultimation analysis process led us to expect would occur. Thus, when considering the value of information, we are forced to operate with expectations and use the abstract event-probability distributions to engine our decisions.
about when to continue or when to discontinue analysis and finalize the model at hand. In other words, we expend resources for analysis in the expectation that the marginal product of the information generated will be positive; that is, that the value of the reduction in the expected value of error will exceed the cost of obtaining the information engineering the reduction.

In summary, (a) The applied scientist exercises himself in the generation of models which have their "purpose" in the ambition to treat a real world phenomenon in terms of prediction, description, or causal reconstruction. (b) The models are predicated on information which may be either a priori or a posteriori in generation and which may be directed toward the completion of the model on any of four dimensions: state-variable, parameteric, relational, or coefficient. (c) The information generated, at positive expenditure of time, effort, and other analytical resources, finds its utility in the ability to engine more favorable a posteriori informational states from less favorable a priori ones, where the quality of the informational states is determined indirectly by the morphology of the event-probability distributions associated with them. (d) Hence, every model we build and every component of the model generated during the analysis process may be indexed with some probability of accuracy, where this refers to our confidence that the "event" we have elected to enter into the model will be identical with the real-world event as yet to occur (or, in the case of causal reconstruction, which occurred). (e) Finally, the imputed value of information is found in the reduction in the expected value of error associated with its acquisition; thus the transformation of the event-probability distributions engined by successive increments of information has a directly utilitarian significance for us, and the fundamental concepts of marginal economics may be called into play (however grossly they must be applied initially). So long as the marginal product of increments of information may be calculated as expectedly positive, analysis processes will be iterated and the finalization of the model delayed.

In general, then, when we consider the analysis process to be a discontinuous one and enter our as yet immature concept of the learning curve, we can suggest that the circumstances under which additional increments of information will be sought (i.e., the analysis process iterated) will be the following:

(a) There is a perceived significant variance between the existing stock of information (either a priori or a posteriori) and the postulated real stock of information.

(b) There is expected to be a positive marginal product associated with the acquisition of the next increment of information (that is, its acquisition is expected to be worth more than it will cost to obtain it).

(c) There is no requirement that the decision be made immediately, irrespective of the existing information state relative to the potential. In other words, the decision-horizon is still some distance away.

(d) If we are in the a posteriori state, a final condition is that there be some substantive and positive difference between the information stock at time $t_{j-1}$ and time $t_j$; that is, that we have a "history" of successive improvement in our informational acquisition via the analysis process. If successive observations on parameters or properties of interest have yielded no new information, we can assume that we have hit some kind of analytical ceiling, that the properties of the real state are changing more rapidly than we can encompass using the specific analysis instruments and strategies involved. In this case, we can either take decision action on the basis of what we have or explore the possibility of inaugurating a new research strategy or instrument array—a new analytical modality.

In the most elementary sense, then, we have the bases for what may be considered a simplified general system analysis model. It is "general" because it is, at this point, context-independent; it is simplified because we have incorporated none of the behavioral or affective factors which operate in real-world analytical situations. It is graphically reproduced in Figure 5. The story this logic diagram tells us is admittedly fabricative (normative), but useful for us both here and in subsequent discussions as a heuristic. The interpretation is abbreviated as follows:

(a) Initially, the decision maker may (usually illegitimately) perceive an effective identity between his a priori state at time $t_{j-1}$ and the postulated real state. When this is the case, he takes immediate decision action, drawing his premises from the a priori information stock. If this identity is not perceived, he inaugurates an analysis exercise.

(b) Here, an a posteriori information stock is produced which, at each iteration of the process, is compared against the real for identity. If identity is not perceived, the decision maker will look at the difference between the information stock at time $t_{j-1}$ relative to that which was obtained at time $t_j$ (associated with the previous iteration). So long as the relationship continues positive (and significant relative to historical expenditures per increments obtained), the analysis process will be continued.

2. For example, there are some individuals for whom formal, empirical analysis is gratuitous, despite the complexity of the problem situation. These are usually those equipped with a considerably inflated opinion of their a priori information, or individuals who are the product of strict dogmatic development (i.e., individuals convinced that the sun revolves around the earth because of exegetical inferences drawn from religious writings).

3. c.f., Frank George, Models of Thinking (Allen & Unwin, 1970), passim.
(c) However, should the learning curve turn negative or become diseconomic (such that the relationship between successive information stocks is unfavorable relative to costs), a decision action is made using whatever premises may be drawn from the existing information stock at that point.

These are the basic deflection points of the analysis process. But their practical implications for our work here will not become apparent until we have introduced another paradigm, one which takes a more detailed look at the procedures by which problem-solving models are constructed.

II. THE GENERAL PROBLEM-SOLVING PARADIGM

Once again we have to say something about the role of the applied scientist and the role of applied science in general. Specifically, the applied scientist, in addition to building models which attempt to predict and/or describe or causally reconstruct phenomena, finds his essential utility in the ability to devise solutions to problems. In this sense, the applied scientist builds not simply one model (the predictive/descriptive/reconstructive allegory) but two: the second is a model which allegorizes the solution in terms of its effect on the problem. In other words, he develops a model which defines the problem to be solved; then, given this definitional model, he develops a model of a solution. Thus the applied scientist differs from his Epicurean counterpart largely in the fact that, in addition to studying phenomena in their own right, he is also interested in somehow altering their properties or behavior.

In most cases, the problems which we encounter will take the form of a system, given the definitional parameters we outlined in Section I; also, in most cases, the solution proposed will also take the form of a system. The former is a natural phenomenon; the latter is a prescriptive fabrication of the scientist until it is reified via development and implementation. Thus, in the most basic sense, the applied science domain is characterized by a set of interfaces between problem systems and problem-solving systems. And, quite clearly, the contribution of the sciences to the world at large is directly dependent upon the success we enjoy in these interfaces.

So, the generalized problem-solving paradigm we will introduce has the following properties:

- It views the problem-solving process as the development of successively more precise and accurate allegories of both the problem to be solved and the system to be designed to solve it.

The design criterion for the system is congruence, which implies that it is both adequately effective and efficient in solving the problem (not, necessarily, optimal).
The criterion for rationality imposes the condition that, at all points, analysis be undertaken to the extent that the resultant information associated with analysis be expected to be marginally productive (in that the expected cost of dysfunction is reduced by an amount greater than the cost of obtaining the information, something which will be made clearer shortly).

Our paradigm suggests two major phases, within which are grouped eight tasks. We have assigned them a sequence which need not be followed in any particular instance for, as will be shown later, whereas the problem-solving process for the mechanistic problem situation is monolithic, that which we would associate with the pursuit of essentially gestalt-like, organic problems via complex systems is a reflexive, opportunistic process. Considerably abbreviated, the tasks of the paradigm are the following.

Tasks in the Development of an a priori Problem-Definition Model

The design of a successful problem-solving system depends on the rectitude of the problem definition which precedes the actual design process.

1. Identification of Problem Determinants. This first step asks that we identify all structural factors (major variables) likely to have a determining impact on the "state" of the problem during the time-frame with which we are concerned.

2. Development of Dynamic, Predictive Micro-Models. Each of the higher-order determinants identified in the first task will itself be a product of still lower-order factors, which determine the way the higher-order factor will be valued or will behave at any point in time. Thus, if we are to be able to predict some future problem state or fully define the structural processual properties of the problem at hand, we must have causal models allegorizing the way each of the major problem determinants is itself determined.

3. Development of an Integrated Macro-Model. In this final step of the problem definition phase, we are concerned with identifying the way the higher-order determinants identified in step 1 are related in "causing" the problem, much as in step 2 we were concerned with the way the determinants themselves were caused by the lower-order factors. This step involves synthesis, in that we are here trying to establish the way the major problem determinants behave in interrelationship with one another. When this step is completed, we should have arrived at a fully integrated, predictive model of the problem we are to solve, a model which will project future problem states, exhausting the structural/functional properties of the problem as fully as possible.


Tasks Leading to the Design of an Effective/ Efficient Problem-Solving Model

The performance requirements a problem-solving system must meet will be dictated by the a priori problem model just developed, subject to the condition that this largely deductive construct be empirically validated at all points prior to any firming of the system properties.

4. Identification of Sub-Problem/Sub-System Interfaces (or Modules) and Development of Treatment Schedules. Drawing on the network-type problem macro-model developed above, we isolate those problem components most in need of adjustment, modification, or elimination (either in their structural or processual properties), thereby establishing the performance requirements for a set of sub-systems to be dedicated to each of the sub-problems, thus responding to the "partitioning" criterion introduced earlier. Each of the various sub-problem/sub-system unions may involve one or more of the basic problem structural determinants identified in 1 above, depending on the nature of the interrelationships specified by the integrated, predictive problem model developed in 3. We must then decide which of the various problem system modules to implement first, etc., in order to maximize the information and developmental effects of our problem-solving exercise.

5. Development of a Subproblem-Solving Strategy for Each Subsystem. Drawing on the causal sequences allegorizing each of the major problem determinants developed in step 2, and considering the inter-determinant relationships predicted to be operative at the macro-problem level, we must devise sub-problem solving strategies for each of the various sub-problems specified in step 4, subject to the condition that the individual strategies be defined to avoid conflict, redundancy, or interference with the other strategies. Here we are concerned with introducing treatment integration in the system rather than structural integration, the former being a somewhat more simple analytical process than the latter, as will be shortly explained.

6. Assignment of Problem-Solving Instruments to Each Sub-System. The initial partitioning of the system allows us to treat each sub-problem/sub-system union as an essentially isolated entity at first (although this might seem a contradiction of established system design principles). At any rate, this task involves the selection of those instruments which will best see the sub-system's strategy translated into effective action. The sum of all such instrument arrays will, of course, be the sum of all resources (material, skilled personnel, equipment, and tactical algorithms) of the total system.

7. Empirical Validation Cycles. At this point, the problem constructs, the strategic platforms,
and the instrument arrays are predicated largely on
deductive inference (possibly some inductive infer­
ences as well) and all must now be subjected to
empirical field trials designed to determine their
validity in the face of real-world contingencies for
each of the sub-problem/sub-system modules. Thus
all a priori constructs will be examined for the
degree of divergence between predicted and actual
outcomes with significant divergences demanding
theoretical or inferential revisions, etc.

8. Development of an Aggregated Macro-
System Configuration Model. The functional (or
(treatment) integration of the system has been per­
formed in step 5, so here we are concerned in
viewing this partitioned system as an aggregate,
roughly equal to the sum of its parts. We can then
aim at effectively maximizing the efficiency of the
partitioned system through the elimination of any
structural or resource (instrumental) redundancies.
In the long run, it will be used to gradually elimi­
nate those sub-systems which have completed the
desired affect on their respective sub-problems,
reallocating or disposing of the resources assigned
them.

What we suggest here is graphically summa­
rized in Figure 6.

THE NORMATIVE PROBLEM-SOLVING PROCESS

Fig. 6

Within the context of this model, then, the
problem-solving process becomes an exercise in
allegory-building, one associated with the emergent
properties of the problem itself, gradually captur­
ing the properties of a problem-solving system
congruent with the developing problem definition.
Within the various sub-processes we find the four
levels of analysis, introduced in the previous sec­
tion, as follows:

<table>
<thead>
<tr>
<th>SUB-PROCESS</th>
<th>LEVEL OF EVENT</th>
</tr>
</thead>
</table>
| 1. Identificatio of major problem
determinants. | 1. State-Variable |
| 2. Development of causal sub-models. | 2. Relational |
| 3. Development of predictive macro­
model. | 3. Parametric/Coeficient |
| 4. Development of sub-system/
sub-problem interfaces. | 4. State-Variable |
| 5. Development of sub-problem
solving strategies. | 5. Relational |
| 6. Assignment of sub-problem
solving instruments. | 6. Relational |
| 7. Empirical validation exercises. | 7. Coefficient |
| 8. Development of macro-system
model. | 8. Parametric |

As we have suggested, the expediency with
which we will be able to pass through these mode^-
buiding procedures and the ultimate level of suc­
cess we can expect to achieve depends primarily on
the nature of the problem (i.e., the phenomenologi­
cal category to which it belongs). Not only will this
factor determine the morphology of the learning
curve we can legitimately expect to generate; it will
also (at least partially) determine the phenomeno­
logical category into which the problem-solving
system will fall.

III. THE ARRAY OF PROBLEMIC
IDEAL-TYPES

Recalling the Aristotelian dictate with which
we began, it is clear that the ease or difficulty
encountered in moving through the problem-solving
paradigm will depend primarily on the properties of
the phenomenon at hand. Those properties which
are of elemental concern for us may be abstracted
into four problemic ideal-type entities, as shown in
Figure 7.

Given these synthetic probability distributions
(Figure 7) it should be evident that as soon as the
system at hand begins to depart from the criteria
for the essentially deterministic entity, we shall
begin encountering difficulties in virtually all
aspects of the problem-solving process. Particu­
larly, as the problem begins to approach the
severely stochastic or indeterminate ideal-types,
the scientist will face the following inhibitory
factors.

Properties Inhibiting Problem Definition

The lack of identifiable, definable boundaries
for the problem (system) due to the constant inter­
change of forces and material with the environment.
This means that some problem determinants will be

416
I. DETERMINISTIC
Where, for any given set of starting-state conditions, there is one and only one event which may be assigned a significant probability of occurrence (i.e., as with the finite-state automata).

II. MODERATELY STOCHASTIC
Where, for any given set of starting-state conditions, a limited number of qualitatively similar events must be assigned significant probabilities of occurrence (as with the problem of trying to estimate next period sales levels for a well-precedented product).

III. SEVERELY STOCHASTIC
Where, for any given set of starting-state conditions, a large number of qualitatively different events must be assigned significantly high probabilities of occurrence (as in the area of conflict behavior or game-based analyses).

IV. INDETERMINATE
Where, for any given set of starting-state conditions, there is no event which can be assigned a significant probability of occurrence; thus the high probability that some outcome we have not been able to pre-specify will occur (as in extremely long-range forecasting problems).

The dynamic causal sequences leading to the problem's state(s) may be expected to be unallegorizable because of (a) components' capability for equifinal behavior (getting to a value-point by different paths); (b) reflectivity, by which every variable in the problem may legitimately be expected to exert determining force on any other (not hierarchical determinacy or otherwise 'ordered' behavior); (c) some of the forces determining the problem state will be transparent (we won't know they're there), especially those attitudinal and behavioral determinants which can be only reflectively observed (through their impact on observable parts).

The future state(s) of a problem are largely unpredictable because of structural/dynamic indeterminacies and because their components (human or social entities) have the capability for initiating strategic behavior in response to encroachments by the problem-solving system.


6. I have encountered in both Latin America and the Caribbean what might be called anabolic cultures, desiccated and transparent until they are inadvertently attacked by a development program. They then become very tangible defenders against social change. For example, the vestiges of medieval Catholic economic dogma, arguing against material accumulation, are generally overlooked by development officials in Northeast Brazil, yet they appear to inhibit local investment, as the associated implication of accrual of gains is a deep-seated religio-cultural offense.
Properties Inhibiting the Design of a Congruent (Effective/Efficient Problem-Solving System)

Because of the inability to accurately determine the future "states" the problem-system might evidence, there is a probability that any problem-solving strategy we elect to implement will prove dysfunctional; that is, we may have predicated our treatment on an event which will not in fact occur.

Because of the wide behavioral-reactive repertoire available to components of non-mechanical systems, evidenced especially in their facility for "strategic" behavior, any therapeutic or control instrument we elect to impose may prove dysfunctional because of (a) the lack of empirically validated correlations between instruments and effects within the context of organic entities; (b) the possibility that instruments employed in an organically structured entity (i.e., a reflexive-recursive system context) may interfere with each other or cause a priori unpredictable events which prove inhibitory to positive, purposive control or change.

There is the possibility that the problem-system we are trying to control or modify and the problem-solving system we have designed will be chronically out-of-phase, largely because state-changes may occur more rapidly in the system we are treating than can be adjusted for in the problem-solving system we have inaugurated.

In rather more general terms, Sorokin\(^7\) has walked the same ground, suggesting that:

...there is the fundamental difference between sociocultural and physiochemical—or even purely biological—phenomena. It consists in a profound difference between the componential structure of sociocultural phenomena on the one hand, and that of physiochemical and purely biological phenomena on the other. Any empirical sociocultural phenomenon consists of three components: (1) immaterial, spaceless and timeless meanings; (2) material (physiochemical and biological) vehicles that "materialize, externalize, or objectify" the meanings; and (3) human agents that bear, use, and operate the meanings with the help of material vehicles.

Thus, the entire analytical strategy of the scientist will have to change as the problem at hand begins to approach indeterminacy. Specifically, the scientist’s model-building exercise will be characterized by the following:

1. The array of problem determinants identified will probably be incomplete because social, economic and political problems cannot be isolated from the environment in which they occur. They are, rather, caused by external factors which may be spatially or temporally far removed and therefore not identifiable at problem-definition time. For example, an individual’s economic performance may be poor because of early diet deprivations, because of adverse social or cultural experiences, because of a religious belief which countermands material acquisition, etc. A system designed to treat poverty would, therefore, have to consider all these and many similarly analytically amorphous factors for which no models currently exist, except of the most speculative kind. Simply, all the structural factors comprising a problem are unlikely to be empirically accessible to the analyst.

2. The behavior of problem determinants can be accounted for only probabilistically, because we have virtually no models which treat largely stochastic or indeterminate both in their interface and processual conditions. We do not, for example, know the precise nature of the interrelationship between, say, heredity and academic potential, or between socio-cultural background and logico-mathematical capabilities of students. Therefore, any correlation we use is likely to have some positive probability of being wrong (either in magnitude or, more seriously, in direction). This reflects not only the inherent behavioral complexity of human beings themselves, but the analytical transparency—the empirical inaccessibility—of the forces driving human systems (valuational, affectual, or surrogate, the residual effect on accessible system components). Moreover, human components, unlike mechanical, can alter their behavior endogenously in response to localized changes, suggesting that an accurate dynamic allegory at one point in time may be highly inaccurate shortly afterward.

3. Therefore, the problem definition (or predictive problem model) we arrive at must be treated as probabilistic rather than deterministic. The major reason for this, aside from the two mentioned above, is that the determinants of a socio-economic or socio-political problem may interact with each other in unpredictable, unspecified ways. Whereas, for example, the combined effects of gravity, friction, and atmospheric density on a space vehicle of a certain configuration becomes calculable in terms of known, resolvable interaction models (e.g. force vectors), the interaction of diet, native intellect, cultural heritage, educational context is practically a matter or guesswork in the absence of predictive, interdisciplinary models. For example, does too good a diet make a student lazy and complacent, or does it give him energy to do good work? Within what limits might either hypothesis be true? Does high intellect discourage the detailed, rote drilling necessary to become facile in basic mathematics, or is it a prerequisite? To what extent might either be true? Our inability to subject such factors to unconfounded laboratory manipulation and measurement heightens this prediction problem in many cases.

---

\(^7\) Helmut Schoek has noted, for example, that "Certain stocks of germs are known to outwit the antibiotics researcher by selectively outbreeding his luck with resistant strains." Quoted from his *Scientism and Values*, (Van Nostrand, 1960), p. 136.

SOCIAL SCIENCE SYSTEM ANALYSIS

4. With a non-deterministic problem definition, the system performance criteria must be developed stochastically. Unless we can fully and precisely define the problem to be solved, we can only guess at the characteristics or properties a system designed to solve it should have. In other words, whatever strategy we decide on might, if the problem is non-deterministic, turn out to be wrong. For example, we might decide that one of the factors inhibiting the economic performance of minority citizens is their early dependence on their mothers (since the incidence of broken homes among minority groups is higher than that of majority races). A sub-problem then to be attacked is the introduction of a father’s influence (perhaps by making desertion subject to strict legal penalties or offering fathers who remain with their families some kind of subsidy). Therefore, our strategy might prove wrong and we will actually have exacerbated the very problem we were trying to solve.

5. Instrument selection must be treated as a stochastic exercise. That is, there is little in the way of empirically verified or statistically supported data to indicate what tactical instruments will be most effective and efficient in solving social, political, or economic problems. One of the major reasons is the fact that, in the human sector, a change introduced one place will reverberate far beyond the immediate problem area, raising the very real problem of differentiating between short-run or local and ultimate or distant effects. Or two different individuals can react to the same tactic instrument in two different ways (one man, for example, might find jail a rehabilitating experience and another might be led to seek revenge or commit suicide; one student might find the "open classroom" a stimulus to achievement, another might find it an excuse to do nothing)—in other words, instrumentically or indeterminately. In so many cases, when we try to measure correlations, so many other factors interfere that an unbiased estimate of effect becomes something of an analytical pipe-dream. Consequently, social instrument-selection is a very different thing than trying to decide what size valve will be best for a given flow of oil in a motor, or what size rocket engine is needed to develop a specified thrust during a given interval, or what level of inventory of a certain part should be maintained to minimize average downtime of a machine in a plant; all problems permitting eventually deterministic and optimal solutions within the current state of the analytical arts.

6. Developing a problem-solving system which will approach optimality in the social science domains will be a heuristic not a deterministic task. Being able to assume that a system design is an optimal one prior to actual implementation and field experimentation depends on the accuracy and precision with which each of the five previous tasks were performed; so the social, political, or economic problem-solver is in trouble right from the start. Unless the system is essentially a deterministic one, we cannot predict how it will perform with any degree of accuracy, and unless the problem is essentially a deterministic one, we cannot predict whether we are employing the right strategy or whether the instruments we’ve elected to use will actually be congruent. So we must proceed heuristically, implementing system parts, constantly testing their relative effectiveness and efficiency, modifying where indicated, and rapidly changing socio-economic and socio-political contexts, for yesterday’s optimal system design is sure to be tomorrow’s atavism.

These conditions simply reflect the fact that human-based systems are "open" rather than closed and that human beings as system components do not have behavioral repertoires that are causally, spatially, or processually constrained to the extent that we would expect from components of physical, natural (non-cognitive), or engineered systems. Indeed, if we were to map the structural and dynamic properties of most social, political, or economic systems, we would emerge with incredibly complex, reflexive, and equifinal networks, not with the neat hierarchical structures most system analysis tools presume.

In summary, the systemic ideal-type to which the problem belongs will have a direct and immediate determinacy on the efficiency and effectiveness of the problem-solving process. In the next section, we want to make this contention a great deal more appealing.

IV. THE CONCEPT OF THE LEARNING CURVE AND INSTRUMENTAL CONGRUENCE

Specifically, we want to offer a deductively-predicated proposition which will make clear the dependency of instruments of analysis on the properties of the problem at hand. First, however, let's review the definitional properties of our four phenomenological (systemic) ideal-types summarized in Table 1. With Table 1 redefinitions in

9. For a discussion of the immensity of hierarchical organizational types in the natural science domain, see Whyte, Wilson, and Wilson, eds., Hierarchical Structures (Elsevier, 1969). My major disappointment with attempts to make hierarchical organization the structural sine qua non of the social sciences is its inability to account for recursive reflexive behavior we note associated with socio-economic and socio-cultural entities, especially when the entire concept of higher-order entities, being impervious to determinacy from lower-order entities, is called into question by so much empirical evidence of the social sciences.
I. Deterministic

No significant relational or structural changes through time; state alterations are negligible.

II. Moderately Stochastic

Some significant changes in coefficient and parameter values, but invariant basic structural and relational properties.

III. Severely Stochastic

Significant relational and some structural (i.e., state-variable) changes through time, where relational changes are reasonably well contained and where structural changes are either periodic (replicative) or are drawn from a limited population of state-variables (determinant array).

IV. In-Determinate

Significant structural changes through time, such that state-variables or major determinate cannot be pre-assigned except partially and probabilistically; both causal and structural properties either empirically inaccessible or unassignable; state-changes independent of prior states.

<table>
<thead>
<tr>
<th>SYSTEM TYPE</th>
<th>HISTORICAL PERFORMANCE CHARACTERISTICS</th>
<th>SUBSTANTIVE EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Deterministic</td>
<td>No significant relational or structural changes through time; state alterations are negligible.</td>
<td>Finite-state automata; production functions for automated processes; properties of institutionalized (primitive) socio-economic systems.</td>
</tr>
<tr>
<td>II. Moderately Stochastic</td>
<td>Some significant changes in coefficient and parameter values, but invariant basic structural and relational properties.</td>
<td>Promotional elasticities associated with established market and products (and demand parameters); market shares within an oligopolistic industry; input-output ratios for bureaucratic organizations; demographic factors in given regions.</td>
</tr>
<tr>
<td>III. Severely Stochastic</td>
<td>Significant relational and some structural (i.e., state-variable) changes through time, where relational changes are reasonably well contained and where structural changes are either periodic (replicative) or are drawn from a limited population of state-variables (determinant array).</td>
<td>Stochastic-state machines: distributions of political offices; meteorological phenomena; athletic events; human reactions phenomena; military games; labor management confrontations; the problem of induced genetic mutations.</td>
</tr>
<tr>
<td>IV. In-Determinate</td>
<td>Significant structural changes through time, such that state-variables or major determinate cannot be pre-assigned except partially and probabilistically; both causal and structural properties either empirically inaccessible or unassignable; state-changes independent of prior states.</td>
<td>The fashion industry's market; artistic and creative enterprises; cosmological and teleological phenomena; opportunistic phenomena (i.e., guerrilla units); heuristic machines.</td>
</tr>
</tbody>
</table>

After the least number of observations or at the lowest expenditure of resources, the determinative system (the surrogate for the organic ideal-type), we notice the least decline per unit of expenditure and postulate that there will still remain a significantly positive probability of descriptive or predictive error were infinite resources expended or an infinite number of empirical observations made on the properties of interest. This means that one of the major assumptive tenets of learning theory (and mathematical learning models) can be legitimately expected to hold for the mechanical case, but not for the intermediate or organic. Specifically, that

$$\lim_{n \to \infty} E(X_n | Q_n) = X,$$

where $X_n$ is the most likely estimate for the true value $X$ of a parameter, generated by the $n$-th a posteriori probability distribution in a Bayesian type successive approximation process $Q_n$.

However, for non-mechanical subjects we cannot legitimately expect that the limiting value of our estimator of a parameter (or entity property) will approach the true value, even as $n$, here representing expenditure of time and/or analytical resources, approaches infinity. Rather, we must provide another construct for this type of system:

(a) \( \lim_{n \to \infty} P(\hat{X}_n | Q_n = X | F) << 1 \)

where $F$ is the true stochastic distribution for the parameter in question and $X$ is the most likely value of that parameter, given $F$, because

(b) \( \lim_{n \to \infty} P(Q_n = F) < 1 \)

In other words, we normatively expect our estimates of an organic entity's properties to be indexed with a probability of error, despite the extent of our analytical efforts (both for any parameter values themselves and for the successive a posteriori probability distributions purporting to reflect
the true stochastic distribution). Indeed, it is possible that some essentially organic entities might give rise to the ergodic-stochastic paradox of increasing information entropy, through study-time, because their potential for change exceeds the resolution power of our analytical instruments. Simply, we must assume that our predictive models will always be somewhat out of phase when working with subjects approximating the organic ideal-type.

To bring the concept of the learning curve into empirical perspective, we can replace the residual-probability-of-error factor with another variable: the first differences in stocks of information acquired during an analysis process. That is, one surrogate for the extent to which potential information about a subject is being exhausted is the difference between the stock of information which existed at time \( t \) and that which exists at time \( t+1 \), the interval being the time allocated to another iteration of the analysis process, i.e., the time for another empirical observation of the properties of interest. Let’s employ three general, familiar abstract ideal-types (the mechanical, man-machine, and human system) and set out learning curves for them (Figure 9). The curves describe the following function, where \( B \) is a stochastic variable:

\[
\lim_{n \to \infty} E (X_n - X_{n-1}) = B
\]

where \( X_n \) is the value of the \( n \)-th observation of the parameter \( X \) taken in a disciplined learning exercise.

Quite obviously, the concepts of the "limiting value for the learning curve" and the "residual probability of error" are closely related, for when the asymptotic value of \( B \) declines—as successive observations bring us less and less new quantitative or qualitative information—we can suggest

(a) \( \lim_{B \to 0} P(\hat{X}_n = X) = 1 \)

(b) \( \lim_{B \to \infty} P(\hat{X}_n \neq X) = 1 \).

Letting \( R \) stand for the residual probability of error associated with some level of \( B \), and setting 1, 2, and 3 as indices representing the mechanical, man-machine, and human system types respectively, we propose

(c) \( \lim_{n \to \infty} P(R_1 \mid B_1 < R_2 \mid B_2 < R_3 \mid B_3) = 1 \)

and (d) \( \lim_{n \to \infty} P(R_1 = B_1 = 0) = 1 \).

Simply stated, as the number of observations on a system increases toward infinity, the residual probability of error associated with the limiting value for the learning curve will be least for the mechanical entity, greatest for the organic-type entity, and intermediate for the man-machine type system. However, we need not consider only limiting values, for the inclusion of a decision horizon (Figure 9) which reflects the point in time where a decision must be made, irrespective of the current error level, indicates that, for any given level of resource expenditures or for any given number of observations, the residual probability of error and the first-differences between successive observations will be highest for the organic-type entity, our associated predictive models least accurate, etc.

Just as we can develop classes of phenomena as ideal-types (i.e., the three in Figure 9), we can also think of taking the entire arsenal of scientific instruments available to us and parsing it into meaningful categories. For those somewhat familiar with the components of this arsenal, we find that we can devise an instrumental trichotomy which parallels the following phenomenological trichotomy: (a) the set of deterministic instruments collected roughly under the heading of optimization techniques; (b) the set of successive approximation and stochastic programming instruments which serve to generate occurrence probabilities about a limited set of alternatives; (c) the set of heuristic instruments which are designed to gradually transform a priori chaos into some kind of actionable order via disciplined, controlled learning algorithms mixing deductive, inductive, and empirical processes as dictated.

However, as we earlier separated the stochastic category into two sub-categories (moderate and severe), we can also isolate two sub-categories of stochastic instruments: those designed to treat essentially moderately stochastic problems and those designed to handle problems falling into the...
severely stochastic range. The above table, then, presents a separation of the array of analytical instruments available to us into four significant categories, with the rationale for the system type/instrument type correlations included.

In summary, then, we can suggest that there will be some vector of analytical congruence which can be grossly defined between the four phenomenological (or systemic/problemic) ideal-types and the various instrument classes we defined, as in Table 3.

### Table 3

<table>
<thead>
<tr>
<th>SYSTEM TYPE</th>
<th>INSTRUMENT TYPE</th>
<th>Deterministic</th>
<th>Moderately Stochastic</th>
<th>Severely Stochastic</th>
<th>Indeterminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>Finite-State System</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferential/Incoercive</td>
<td>Stochasticity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorical/Convergent</td>
<td>Indeterminacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heuristic Meta-Hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 3, the main diagonal represents this vector of congruence, where the given analytical state is associated with that instrumental subset expected to deal most effectively and efficiently with it. We use the term "congruence" to suggest that this instrument class will have the highest logical probability of producing an adequate ultimate error level at the lowest associated expenditure of time and analytical resources. To the right of the main diagonal we strike an expected interface condition called "inefficiency." The use of this term reflects a probable misallocation of analytical resources. Here we are using instruments designed to deal with a more complex analytical state than that indicated, and these more powerful instruments will have a higher cost associated with them. Thus we might possibly employ stochastic tools to solve an essentially deterministic situation, but we will reach any terminal error level at a higher cost than had we used more economical, more expeditious deterministic or optimization instruments. To the left of the main diagonal falls a more serious situation, here called "ineffectiveness." The implication is that we are employing analytical instruments...
which are not powerful enough to resolve the properties of the problem at hand, suggesting that whatever eventual error level we do achieve will be unacceptable or considerably higher than necessary. For example, the economist, employing essentially deterministic "shock models" in an effort to describe the structural and dynamic properties of a regional economic system, trades off analytical expediency against ultimate rectitude. Any policy decisions made then on the basis of that model's output will inherit a significant probability of being dysfunctional, for they will have incorporated considerable oversimplifications of the real-world problem.

V. DEFENDING THE NORMATIVE LEARNING CURVES

Instrumental congruence is achieved when we employ an instrument which will yield the desired level of information at the lowest associated expenditure of analytical resources—or, more realistically perhaps, will yield a desired quantum of information at a reasonable expenditure level. As suggested earlier, we can gain a surrogate measure of the value of information in its effect on event probability-distributions associated with some phenomenon, and on the efficiency of the information gathering instrument by analysis of the successive first differences in information stocks—learning curve analysis.

In a natural obedience to these deductively-predicated propositions, we find the several sciences in possession of model bases which usually contain instruments drawn from all four categories. Using administrative problems as a focus, for example, we might find examples as shown in Table 4.

To return to the defense of the normative learning curves developed in the previous section, we have to take a closer look at the characteristics of these components of the aggregate model base and, in the process, look for reasons why analytical efforts aimed at deterministic phenomena result in learning curves which rapidly drop to an asymptote of zero-error, etc. We find the reasons in the morphological parallelism between the various analytical ideal-type categories and the instrumental sub-arsenals.

As we move from deterministic toward indeterminate phenomena, the empirically-validated component in the models declines in favor of judgmental or deductive factors, with a corresponding decrease in the morphological correlation between the entity under study and the allegory purporting to describe or predict it. Quite simply, then, we begin to displace fact with opinion and, in the process, incur ever greater levels of expected predictive or causal error. Hence the empirical predication of deterministic models is expected to yield (naturally) a lower variance between "predicted" and "real" events than a statistical inference-based model; the statistical-inference model is, in turn, expected to be more reliable in its predictive or causal allegories than the inductive inferences underlying the severely stochastic models; and, at the indeterminate extreme, the strong deductive (or intuitive) component argues that we expect significant divergence between predicted or allegory-driven events and real-world outcomes.

So, the model classes, like the system classes, differ on several critical dimensions summarized in Table 5. As the table indicates, the model or instrumental categories differ on the indicated dimensions as much as do the systemic/problemic ideal-types they are to serve. The precision of the information output generated by the models refers to the "narrowness" of numerical or qualitative intervals separating the various "events" generated (or the mathematical precision of the single event generated under the optimization category, i.e., the number of significant digits). The actionability of the output refers to its ultimate utility in so far as dictating "actions" are concerned. Clearly, for both the severely and moderately stochastic case, the ability of the information to direct subsequent action depends on the number of alternatives outlined and their degree of difference in qualitative terms; for the usual condition of scarce resources makes it impossible for us to pre-adapt to many, significantly different alternative
events assigned probabilities of occurrence (the inherent limitation of the utility of game-based or contingency-predicated analytical procedures, as with war games, etc.). Prescriptiveness is a sub-dimension of actionability and refers to the degree of detail actions dictated by the informational output assumed: the model’s ability to provide detailed, precise action references.

We shall shortly defend the several analytical bases we associate with the instrument categories but, for those familiar with epistemological predications, the correlations between the analytical bases and the nature of informational output are virtually tautological, as is the step from there to the contentions about the probability of predictive/prescriptive error associated with the various informational modalities. The assertions housed in Table 5 become somewhat clearer when we note that information per se is always a product of some model (i.e., some component of the available model-base) operating on or manipulating raw data (i.e., some elements drawn from the data base we have on hand pertinent to the system we are treating).

Ultimately, then, the information output from an analysis process (in the form of a purportedly predictive or explanatory-reconstructive allegory) will owe part of its substance to empirical data, part to logical or mathematical/statistical engines. To the extent that it relies heavily for its components on conceptual or intuitive devices, we suggest it to be primarily a product of deductive inference; on the other hand, when the majority of the informational content of the allegory is derived from empirical data (e.g., such that its coefficients are actually objectively measured), then we consider it basically a positivist construct, which, when generalized, is an exercise in inductive inference.

Inductively predicated allegories generally express probabilistic engination, such that an allegory may predict (or attempt to reconstruct) a phenomenon’s behavior under the assumption that it will behave according to certain empirically-generated generalizations with some significant probability. The data component is very strong, but the actual structure and substance of the allegory owes something to statistical, mathematical, or logical mediation, such that the morphology of the model or allegory need not be isomorphic with respect to the properties of the phenomenon being treated (but it must be a morphological extension or extrapolation). And, to a limited extent, some variables will be exogenous by way of entering certain contextual assumptions, but these too will be calculable, formally-derived extensions of empirically accessed phenomena. Thus a model imputing a certain behavior to a class of phenomena on the basis of successively more exhaustive analysis of individual members of that phenomenological class is an

<table>
<thead>
<tr>
<th>MODEL CLASS:</th>
<th>ANALYTICAL BASE:</th>
<th>NATURE OF INFORMATIONAL OUTPUT:</th>
<th>QUALITY OF OUTPUT:</th>
<th>PROBABILITY OF ERROR:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Optimization</strong></td>
<td>Positivism</td>
<td>Generates a single solution for any given set of predications within a closed-system context (i.e., as linear-programming models)...</td>
<td>High degree of precision... results depend directly on empirical data base with minimal judgmental intervention.</td>
<td>Extremely low (as there is no attempt to generalize from results).</td>
</tr>
<tr>
<td><strong>Extrapolative/Projective</strong></td>
<td>Inductive Inference</td>
<td>Generates a 'range' of possible solutions or a probabilistic estimate assigned some index of confidence...where alternative events are derived from data base.</td>
<td>Prescriptiveness of output depends on the 'width' of the range of event alternatives; precision on data base quality.</td>
<td>Depends on the 'quality' of the data base as a source of inferences...objectively determined.</td>
</tr>
<tr>
<td><strong>Contingent/Game-Based</strong></td>
<td>Deductive Inference</td>
<td>Generates an array of alternative events of different 'quality', where alternatives are usually independent of historical data base.</td>
<td>Alternatives are usually in the form of scenarios and not too precise; actionability depends on number of alternatives presented and their degree of 'difference'.</td>
<td>Usually quite high, as output is intended to provide events of unprecedented nature in the problem context at hand...but error is reduced as alternatives are empirically validated.</td>
</tr>
<tr>
<td><strong>Heuristic/Meta-Hypothetical</strong></td>
<td>A-prioristic or Intuitive -istic.</td>
<td>Generation of broad heuristics or learning-based paradigms to discipline initial informational pursuit.</td>
<td>These are prescriptive only in a methodological sense, not a substantive one.</td>
<td>The heuristic does not carry any pretentions to accuracy or realism, per se.</td>
</tr>
</tbody>
</table>
inductively predicated allegory, and its probable predictive or explanatory reliability is, to a large extent, calculable within statistically generated limits.

For the most part, deductively predicated allegories (those whose predictive or explanatory significance owes more to logical or intuitionalistic devices than to empirically predicated data base elements) are tolerated only when the inferences to be made are across wide tracks of time or space—for example, as with the admittedly logically-probable if not statistically-probable constructs of the archaeologists or the fabricative, futuristic scenarios of the long-range forecasters (cf., Kahn and Wiener, The Year 2000, Macmillan, 1967).

Thus the concepts of the model and data base, and the proportional reliance on each associated with some terminal predictive or descriptive allegory, serves as a prime dimension for the generation of an array of analytical modalities, which incorporates the following gross intervals:

1. **Positivistic Modality**—where no logical, mathematical, or statistical model mediates between the data base and the ultimate allegory, such that there is an almost perfect correlation between the components of the original data base and the components of the ultimate allegory. Alternatively, the positivistic modality is indicated by the fact that every element of the allegory is itself deterministic, being assigned no significant probability of departing from the "assigned" value (e.g., parametric or coefficient).

2. **Inductivist Modality**—where the reliance on the original data base is still extremely strong, but where some "model" has intervened such that the morphological correlation between data base and allegory is dampened. In general, inductivist allegories will be extensions or extrapolations of the elements of the data base, such that the parametric or coefficient values assigned the allegory are products of statistical inference, etc.

3. **Deductivist Modality**—where the reliance on the empirical data base is rather weak, such that there is little morphological correlation between the components of the ultimate allegory and the components of the original data base. In other words, under the deductivist modality, we introduce significant qualitative changes which mediate between the data base and the allegory.

4. **Heuristic Modality**—where there is no data base from which to work which has any relevance for the problem at hand, but only isolated and unintegrated scraps of historical-empirical evidence. Hence, under this modality, we deliberately fabricate possible futures and develop initially artificial (i.e., fictional or analytic) frameworks within which a subsequently disciplined learning exercise may take place; where the heuristics employed often owe their origin to idiographic or analogic constructs (i.e., abstract ideal-types).

Considering again our concept of congruence, we suggest the following Table 6.

<table>
<thead>
<tr>
<th>ANALYTICAL MODALITY</th>
<th>Deterministic</th>
<th>Moderately Stochastic</th>
<th>Severely Stochastic</th>
<th>Indeterminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positivistic</td>
<td>Determinism</td>
<td>Modality</td>
<td>Optimisation</td>
<td></td>
</tr>
<tr>
<td>Inductivist</td>
<td>Extrapo-</td>
<td>AWAHP</td>
<td>MOD(I1)</td>
<td></td>
</tr>
<tr>
<td>Deductivist</td>
<td>AREA OF</td>
<td>MOD(I2)</td>
<td>GEN-</td>
<td></td>
</tr>
<tr>
<td>Heuristic</td>
<td>Meta-Hypothesis</td>
<td>Learn-</td>
<td>BASE</td>
<td></td>
</tr>
</tbody>
</table>

The positivist, inductivist, deductivist, and heuristic modalities exist as broad analytical procedural paradigms associated with one or another of the problem or system classes (normatively). The intervals occupied by each are, of course, only grossly definable a priori, and the refinement of the intervals is the responsibility or legitimate ambition of the various social science disciplines per se. At any rate, we can suggest something about the relative "size" of the predictive error we might associate with each of the modalities on the basis of some arbitrary data/model trade-offs, as diagrammed in Figure 10.

As we begin, then, to displace fact with judgment and measured, empirically observed allegory components with deduced ones, we must expect to incur successively higher levels of analytical (e.g., predictive or explanatory) error. With this in mind, the contributions we can legitimately expect from inductive and deductive modalities reflects the following propositions:

(a) The probable accuracy of the results generated by inductive models increases (e.g., phenomenological population) as the proportion of the subject universe recorded in the associated data base increases.

(b) Deductive models, not depending on the existence of a data base, may be used in the absence of a high "quality" data base.

425
The proper proportional reliance on inferential models depends, then, largely on the quality of the data base, which depends jointly on:

1. The completeness of the data base which is reflected in the proportion of actual phenomenon properties which have been captured and recorded.
2. The accuracy of the data contained in the data base.
3. Finally, and most important, the relevance of the data contained in the data base (which depends, generally, on the extent to which any exogenous conditions present at the time of the data's recording still pertain).

However, the probable quality of the data base increases as the phenomenon under treatment approaches the concept of the essential mechanism or inherently deterministic entity; correlatively, the probable relevance (reliability) of inductive inferences or inductively-driven allegories for prediction or explanation will expectedly increase under this condition for, reiteratively, (1) the future performance of a subject system which approaches the concept of the mechanical ideal-type is likely to be a simple product of its past performance; (2) the unseen portions of a system which approaches the concept of the essential mechanism are most likely to be a calculable function of the seen, observed, or sampled portions of the system.

These, then, are the tentative foundations for our proposition that the essentially mechanistic phenomenon will allow greater reliance on data bases and, consequently, on inductively-predicated instruments, than the system approximating the organic ideal type.

It follows that there are two conditions which act to make the deductive modality a more favorable analytical alternative: (1) when the phenomenon at hand is inherently severely stochastic, such that no amount of historical information collected on observable or observed properties will adequately serve as a basis for structural and/or temporal predictions; (2) when the historical record associated with an entity is so incomplete or ill-structured that the future behavior of that entity cannot be effectively induced from known or observed properties. Hence we would expect that the informational predications of models we build will reflect the nature of the problem at hand. Particularly, for our two polar ideal-types (the inherently deterministic and inherently indeterminate entities), the attributions or sources of the total information quantum should be distributed roughly as shown in Figure 11. Extending this analysis somewhat, the heuristic modality becomes attractive when the problem at hand is so ill-structured that even the basic state-variables which should be treated are difficult to isolate or target in on; and, lacking this, we can't even afford the luxury of the deductive modality.

effectively indeterminate (as opposed to inherently indeterminate) problem, we expect to see something like Table 7.

In the sense of the Table 7 construct, we can see that the major ambition with respect to applied science system analysis is to gradually reduce a priori indeterminate problems to effective determinacy, passing through the various stages indicated, engined by the analytical modalities set out. And, when faced with problems falling into the severely stochastic and indeterminate analytical states, we can suggest several components of a congruent methodological strategy, as follows:

1. Social, political, economic and behavioral problems demand interdisciplinary attack, in that real-world problems seldom follow the neat demarcation lines one finds among academic faculties.

2. Such problems will have to be treated as systems rather than isolated, monolithic causal sequences. This suggests that the basic methodological vehicle must be a system approach of some kind.

3. Lack of empirically-validated, deterministic social science models and indisputable cause-effect correlations dictates that the analyst and administrator proceed probabilistically, presuming a positive probability of error at all points in the problem-solving process.

4. Finally, granting the prevalence of analytical and operational error, problem-solving systems cannot be taken right from the a priori drawing board into implementation. Rather, the construction of problem-solving systems in the indeterminate domain will have to be conducted as learning exercises, with controlled empirical trials of preceding elements (or modules) used to partially determine the properties built into following elements, etc.

Thus we find the heuristic approach arguing an end to academic parochialism and for an adoption of interdisciplinary platforms; arguing against simple statistical/mathematical models (e.g., the shock models of econometrics or the stimulus-response schemes of psychometricians) and for more elegant and complex systemic formulations; against deterministic models and for models which treat error explicitly; against the quest for quick "completion experiences" and for continuous involvement with emerging, real-world system/problem interfaces.

VI. RECONCILING THE METHODOLOGICAL DIFFERENCES

In summary, we have tried to show that all the instruments in the scientific arsenal have both a positive and necessary role to play within the confines of the applied social sciences, despite the fact that models or allegories predicated on one or another of the instrumental categories we have identified may differ considerably on the following dimensions:

- the proportion of the information set (e.g., the output of the analysis process) dependent on empirically generated data per se;
- the proportion of the information set dependent upon the operation of logical or mathematical statistical paradigms or verbal models;
- the extent to which the paradigms or models employed are themselves empirically validated; the extent to which they are axiomatic;
- the marginal productivity associated with the instruments employed, in terms of their informational output in relation to expenditures of analytical resources or time;
- the absolute effectiveness of the instrumentation used in terms of its ability to encompass and operate with problem properties of varying levels of complexity, irrespective of any marginal considerations (e.g., reduction in predictive error relative to cost).

Most critically, we have tried to deduce the congruent associations between instrument types and systemic ideal-types. The latter are the driving determinants of analysis, at least in the following terms:

(a) The probability of any a priori informational stock fully exhausting the postulated properties of the real is greatest for the essential mechanism (i.e., the Type-I system), least for the essential gestalt (Type-IV system).

(b) Similarly, the probability of any a posteriori informational state adequately exhausting the postulated properties of the real state dimin-

11. An "inherently" indeterminate system is one which is engined to be deliberately unpredictable and erratic in behavior, or which operates opportunistically and with respect to localized (non-universally accessible) criteria, i.e., the heuristic machine. An effectively indeterminate system, however, is one which is a priori so, but can be legitimately expected to move to a more favorable category when formal analysis has been inaugurated.
ishes as the phenomenon at hand departs from the properties associated with the Type-I entity, etc.

(c) The cost, in terms of time, resources or energy, associated with increasing the quality of the a posteriori informational state associated with some phenomenon under investigation is expected to be least for the Type-I entity, greatest for the Type-IV entity, etc. Correlatively, the cost of reducing the probability of predictive and/or causal error associated with the allegorization of some phenomenon, by any given increment, will vary inversely with the extent to which that phenomenon approaches the properties of the essential gestalt.

(d) Finally, the greater the approximation of the entity which we are trying to capture analytically to the essentially deterministic ideal-type, the greater the probability that components of the model predicated on extrapolative or projective instruments operating on an historical data base will be accurate, etc.

In the broader schema of social science system analysis, then, the parochial positions of the warring methodological platforms—the empiricist’s preference for nomothetic constructs, the deductivists preference for idiographic, rhetorical constructs—are hardly enlightened. Neither alone can move the applied scientist through the problemsolving paradigm on anything approaching an optimal learning curve. Rather, as we gradually transform the problem at hand from an ill-structured a priori one, the deductively-predicated practices necessary for the initiation of a complex analysis exercise may gradually be employed to provide us with an ultimately precise, empirically validated, and nomothetic a posteriori model or allegory, at least to the extent that the inherent tractability of the phenomenon under treatment permits.

At any rate, whatever the level of initial indeterminacy or the inherent complexity of the subject being studied, we need not depart from disciplined learning and investigatory criteria; after all, the heuristic approach is, in its way, as formal as the empirical and as appropriate to situations fraught with axiological enginering as the positivist is to essentially mechanical, constrained contexts.

The "truth" about socio-economic and socio-cultural issues, which we all presumably seek, is unlikely to be gained by strict attention to any one of the modalities; rather, it is most likely to occur in the emerging confluence between successively more specific deductive inferences and successively more general inductive inferences. Perhaps Lewin had this kind of concatenative process in mind when he set out what to him constituted a satisfactory theory. As Dorwin Cartwright reports with respect to Lewin’s position: 12

In order to develop a satisfactory system of concepts, the scientist has to be particularly careful about the way in which he develops his concepts. Before a system can be fully useful the concepts in it have to be defined in a way that (1) permits the treatment of both the "qualitative" and "quantitative" aspects of phenomena in a single system, (2) adequately represents the conditional-genetic (or causal) attributes of phenomena, (3) facilitates the measurement (or operational definition) of these attributes, and (4) allows both generalization to universal laws and concrete treatment of the individual case.

What we have done here can, of course, be only a hesitant and somewhat immature step toward the position Lewin ultimately asks us to assume and the stance that integrity will eventually demand from the social sciences.

---

The combination of rising costs of medical care and increasing demands for medical services poses many complex problems for the design and management of medical facilities. Because of these trends, trial-and-error solutions to hospital design and management problems are not feasible. More powerful and less subjective problem solving methods must be used. In this situation, hospital designers and managers are turning to problem solving methods which have been developed in military and industrial situations. These methods vary in comprehensiveness from methods engineering and time-study techniques to operations and systems research. They may all be conceptualized, however, as approaches to the basic problems of design, regulation, and control of man-machine systems.

Interest in the design and operation of man-machine systems has a long history which can be traced from the beginning of the industrial revolution to the vast man-machine complexes on which our military and economic security are based. The development of these systems has led to corresponding developments in methods for designing and operating them. These methods are summarized in Figure 1. Each represents an approach to the basic problem of designing and operating man-machine systems, and each has relevance for developing solutions to different aspects of hospital design and management problems.

**SCIENTIFIC MANAGEMENT**

Scientific management was developed to provide planning information for industrial managers. As Taylor (1947) pointed out: "The thing which we on management's side lacked more than anything else was exact knowledge as to how long it ought to take the workman to do his work [1]."

This information is a necessary prerequisite to production planning, purchasing, and organization of a labor force (Davidson, 1957) (2). Taylor's time study was a procedure for obtaining this information. Having decided what the workman ought to do, his wage payment plans provided incen-
In addition to concern with production problems, industrial engineers have long been interested in hospital problems. Barnes (1949) studied, operating room layout, and Fredericks (1951) investigated the task of the head nurse using link analysis techniques (7,8).

HUMAN ENGINEERING

Concurrent with the scientific management movement and the corresponding development of modern industrial engineering, psychologists were developing experimental procedures for the study of human performance. Wundt's investigation of the human response to various sensory inputs in the first psychological laboratory provided a conceptual basis for much present-day human engineering (9,10). Psychologists, faced with the problem of selecting pilots for the Army Air Force in World War II, found that the complexities of military equipment limited the user population (11). As a result, interest was focused on procedures for the design of military equipment so that it would be simpler to operate. This focus differed from that of the industrial engineer, interested primarily in lowering the cost of production.

An important human engineering problem is now being posed in hospital systems as a result of the development of devices for acquiring, transmitting, and displaying patient data: the performance of the human monitor (12,13). This interest stems from the possibility of remote observation of patients as a partial answer to critical personnel shortages. Selected aspects of patient behavior such as blood pressure, pulse rate, temperature, and respiration can be measured and presented on remote reading displays. Equipment for patient monitoring is being developed in conjunction with the military space programs (14). Blumberg (1961) has discussed the application of electronic equipment to hospital operations (15). Although savings may result from the automation of some hospital tasks, such as data processing, the automation of procedures which will require staff to monitor patients should be preceded by careful testing. The evidence shows that people are poor monitors, and that the probability of signal detection falls rapidly in the first thirty minutes of a monitoring situation (16).

OPERATIONS RESEARCH

Another major development in the design and operation of systems, operational research, began in Britain in the early years of World War II. Operations research provides solutions for many production-oriented problems, particularly those related to inventory control, resource allocation, and programming. The basis for problem solution is a mathematical "model" or representation of selected aspects of the "real world" system. A model representing the aspects which are of concern is manipulated according to rules of logic, and the consequences of policy decisions are predicted from model solutions. As long as the model represents the real world in all critical aspects, it is possible to predict the consequences of decisions from the model without having to actually manipulate the real world system. The object of manipulating the model is usually to determine an "optimum" solution to a resource allocation problem, where optimum is defined relative to some value system external to the model. When an optimum model solution has been determined, guidance for the solution of the real world problem is provided. The model solution is usually in the form of a maximization or minimization of a cost function. The form of the model and procedures for its solution have been described (17).

An optimum solution may be determined mathematically if formal statements of relationships between variables exist or can be developed. The relationships may also be determined numerically, by trial-and-error, or by simulation.

Although operations research has provided guidance for the solution of a wide range of military and industrial problems, it has limitations. The chief limitation is the difficulty of describing the complexity of the real world with existing formal models. Model solutions will be real world solutions if and only if correspondence exists between the model and real world. The required correspondence cannot be inferred from the existence of formal or mathematical relationships, as Churchman, et al., have pointed out (18). Model assumptions, although perfectly valid and necessary mathematically, may not be valid in the real world. The existence of the required similarity between real world and model must be demonstrated empirically.

SYSTEMS RESEARCH

Operations research is concerned primarily with the optimization of existing systems to cost criteria. Systems analysis, which is concerned primarily with the design of future systems, is a logical extension of operations research (19,20). Alternative designs are selected on a cost-effectiveness basis, where effectiveness is measured in terms of system performance, which depends in turn on allocation of resources.

Systems research procedures differ from operations research in a number of ways (21). The most important difference is that they are "descriptive" rather than "prescriptive" (22). They seek to develop behavioral abstractions of systems which can provide guidance for decision-making for a range of values rather than an optimum solution.
APPROACHES TO THE SYSTEMS PROBLEM

relative to any specified value system, i.e., they provide a behavioral description of the consequences of various policies of resource utilization, rather than the value of these courses of action. Specifically, they provide statements of relationships between system task and resource variables. These relationships are derived from empirical data and organized according to logical rules (23).

In our attempts to develop planning models of hospital systems, we have utilized the range of concepts from scientific management to systems research. As a result of difficulties encountered with each of these approaches to the problem, we have developed our own concepts, based on cybernetics (24). These concepts provide a basis for relating system design and operations, where the fundamental operational objective of the system is taken to be the provision of patient care.

Since patient care is provided in a very complex organizational environment, this environment has been abstracted as a hospital system model so that the effect of organizational decisions on patient care can be assessed. This model, and the steps leading to its development, will be presented in a subsequent paper.

REFERENCES

1. TAYLOR, F. W. Scientific Management. New York, Harper and Row, 1947. (Shop management, pp. 17-113; The principles of scientific management, pp. 9-144; Taylor's testimony before the special house committee, pp. 77-93.)
2. DAVIDSON, H. O. Functions and Bases of Time Standards. Columbus, Ohio, American Institute of Industrial Engineers, 1957.
8. FREDERICKS, P. M. The Development and Application of a System of Analyzing the Variable Task. Columbus, Ohio, Ohio State University, 1951. (Unpublished master's thesis)
18. Ibid., p. 592.
21. CHURCHMAN, op. cit.
C.

APPLICABILITY AND LIMITATIONS OF ANALYSIS
MATHEMATICAL MODELS IN THE SOCIAL SCIENCES*

Kenneth J. Arrow

I. THE USEFULNESS OF MATHEMATICAL REASONING

It is a commonplace remark among many social scientists that mathematics, however useful it may have proved in the physical sciences, can play no essential role in the development of the social sciences because the phenomena studied are somehow different—"human beings are not amenable to mathematical law." The social scientist who thinks a little more about the matter will perhaps add that mathematical analysis is quantitative, while his field calls for qualitative analysis. Doubtless he will concede that certain elementary facts of a numerical nature can be tabulated (e.g., distribution of income or population); and he will usually admit that for certain purposes, one might be permitted to add up a column of the table. Nevertheless, it is held that the judgment and intuition of the skilled investigator are fundamentally more useful in the social sciences than mathematical formulas based on quantitative observations.

To the mathematician or the individual trained in the spirit of modern mathematics, the views just presented seem to be based on nothing more profound than a misunderstanding. "Mathematics," said the American physicist Gibbs, "is a language." If this be true, any meaningful proposition can be expressed in a suitable mathematical form, and any generalizations about social behavior can be formulated mathematically. Mathematics, in this view, is distinguished from the other languages habitually used by the social scientist chiefly by its superior clarity and consistency. Furthermore, it is simply not true that mathematics is useful only in quantitative analysis. Doubtless many branches of mathematics—especially those most familiar to the average individual, such as algebra and the calculus—are quantitative in nature. But the whole field of mathematical or symbolic logic is purely qualitative. We can frame such questions as the following: Does the occurrence of one event always imply the occurrence of another? Is it possible that two events should both occur? The events here may be of a purely qualitative nature, such as the presence or absence of traits in a culture complex. It must further be observed that there may very well be a quantitative aspect to the study of even the most definitely qualitative phenomena; if we realize that we will rarely be able to assert universally valid laws about the relation between different traits, we will be willing to ask in what proportion of the observed phenomena will two traits be found together, i.e., what is the probability of their coexistence. The Mendelian theory of inheritance is the prototype of the transformation of qualitative into quantitative analysis via the probability calculus. It is, of course, also clear that quantitative phenomena enjoy equal claim with qualitative ones in the study of social forces. An understanding of a community doubtless requires knowledge of its religious and social beliefs, but it also involves knowledge of the distribution of income by size and by occupation, the total population and its distribution by occupation and social class, and the proportion of resources devoted to various social and individual needs.

Finally, the argument that only trained intuition can yield worth-while social analysis is rejected as meaningless by the mathematically trained. If the intuition of the investigator is reliable, it will yield the same judgments every time it is confronted with the same set of facts. But any such unique correspondence can always be represented by a mathematical relation of sufficiently complicated form. Hence, any intuitive knowledge can always be reduced to mathematical terms. Apart from this, there is the general presupposition that scientific knowledge should be interpersonally valid and transmittable and hence expressible in an objective, consistent language.

These arguments seem logically irrefutable, and yet, outside the realm of economics, very little use has been made of mathematical and symbolic methods. Even in economics, only a small minority of the theorists use anything more complicated than a very elementary form of calculus. How can we explain this failure on

*Reprinted from The Policy Sciences, Chapter VIII (Daniel Lerner and Harold D. Lasswell, editors), Stanford University Press, 1957.

1. The doctrine that mathematics is a superior language goes back to Leibnitz. The inclusion of all forms of logical reasoning in mathematics was begun by George Boole and continued by Charles S. Peirce, Gottfried W. Frege, and Bertrand Russell. See Russell's Introduction to Mathematical Philosophy (2nd ed., 1920). The best elementary presentation of the present state of mathematical logic is found in Alfred Tarski, Introduction to Logic (1941).

2. It is of interest to observe that the modern theory of statistical inference fundamental to any quantitative analysis in the social sciences owes its inception to the interest of Francis Galton and Karl Pearson in the problems of inheritance and evolution. A further discussion of the significance of probability laws in the social sciences is given later in this chapter.
the part of social scientists to accept in their practice the theoretically superior language of mathematics? The simple fact that they would like to shun such a difficult subject as mathematics can be regarded at best as only a partial explanation. There have always been enough mathematicians interested in the social sciences to have made the superiority of mathematics manifest if it were a clearly better tool.

There must be some flaw in the arguments advanced above for the use of mathematical methods in the social sciences; the most important one is concealed in the statement that every proposition "can" be expressed in mathematical form. The statement is doubtless true if we mean that there exists in some Platonic realm of being a mathematical expression of every given proposition, but it is not true if we mean that the mathematical expression in question can be given within the realms of mathematical theory now existing. Every mathematician realizes what a small part of all the potentially available mathematical knowledge is actually, grasped at the present time. The usual reaction of the "literary" social scientist when confronted with a mathematical system designed as a model of reality is to assert that it is "oversimplified," that it "does not represent all the complexities of reality." In effect, he is saying that the symbolic language in which the mathematical model is expressed is too poor to convey all the nuances of meaning which he can carry in his mind. What happens is that the very ambiguity and confusion of ordinary speech give rise to a richness of meaning which surpasses for the social scientist the limited resources of mathematics, in which each symbol has only one meaning. It is not surprising that there should be a difference between the social and the natural sciences in this regard. Language is itself a social phenomenon, and the multiple meanings of its symbols are very likely to be much better adapted to the conveying of social concepts than to those of the inanimate world. Furthermore, the empirical experience on which one's understanding of the social world is based consists to a large extent of symbolic expressions of other individuals; one can apprehend these expressions directly because one is himself part of the social world he observes. Such apprehension must inevitably take place on a largely unconscious level unamenable to mathematical expression (which is surely the acme of consciousness). It is precisely in the field of economics, where the individuals studied are engaged in relatively highly conscious calculating operations, that mathematical methods have been most successful.

It is true, then, that there are certain limitations of mathematical methods in the social sciences. Nevertheless, it must be insisted that the advantages are equally apparent and may frequently be worth a certain loss of realism. In the first place, clarity of thought is still a pearl of great price. In particular, the multiplicity of values of verbal symbols may be a great disadvantage when it comes to drawing the logical consequences of a proposition. Consider, for example, the following verbal arguments: (a) If prices are high, people will tend to buy less; when people buy less, manufacturers produce less, since they tend to produce only what they can sell; therefore, high prices are associated with low production. (b) If prices are high, manufacturers will produce more, since it is more profitable for them to do so; therefore, high prices are associated with high production. At a verbal level, both arguments are convincing; yet obviously they cannot both be valid. Let us try to isolate the postulates of the two arguments and express them symbolically. Let \( x_1 \) denote the amount which people will buy, \( x_2 \) the amount manufacturers will produce, and \( p \) the price. The first argument says that \( x_1 \) is a decreasing function of price, i.e., that \( x_1 = f(p) \) and that \( x_1 = x_2 \). The second argument says that \( x_2 \) is an increasing function of price, \( x_2 = g(p) \). Then we have three equations in three unknowns; thus, in general there is no inconsistency, all three relations simultaneously determining \( x_1, x_2, \) and \( p \). These relations express the behavior of consumers, the market, and producers respectively. So long as no change occurs in anyone's behavior pattern, the values of \( x_1, x_2, \) and \( p \) will remain constant. There will be no question whether

3. Consider the following apparently simple statement: Every even number is the sum of two odd prime numbers (a prime number is a number divisible only by itself and by 1). There is no known exception to the statement, yet it has never been proved in general. This is known as Goldbach’s problem and dates from the 18th century.
4. This point has been stressed by Frank H. Knight in "The Limitations of Scientific Method in Economics," in The Ethics of Completion and Other Essays (1935), pp. 105-47.
5. Credit for the first significant use of mathematics in economics, is due the great French economist Augustin Cournot, who published in 1838 his Recherches sur les principes mathématiques de la théorie des richesses. His work was largely neglected. It was the contribution of W. Stanley Jevons in The Theory of Political Economy (1871) and especially the contribution of Léon Walras in Éléments d'économie politique pure (1874), which brought the power and especially the contribution of Léon Walras in Éléments d'économie politique pure (1874), which brought the power of the mathematical methods to the attention of economists. The present status of mathematical economics is best summarized in Paul A. Samuelson, Foundations of Economic Analysis (1947); see also J. R. Hicks, La théorie mathématique de la valeur en régime de libre concurrence (1937), and Value and Capital (1939), particularly the Mathematical Appendix, for a more intensive survey of more limited areas than Samuelson's work. These works do not cover, however, a good deal of significant work published in the various economic journals, and particularly Econometrica.
supply varies directly or inversely with price, since neither moves at all. Suppose, however, that there were a shift in the tastes of consumers, so that \(f(p)\) changed to another function, say \(f_1(p)\). To obtain the values of \(x_1, x_2,\) and \(p\), we would solve the new system of simultaneous equations, 
\[
x_1 = f_1(p), \quad x_2 = x_2, \quad x_3 = g(p).
\]
Note now that the solutions \(x_1', x_2', p'\) to this equation system satisfy the conditions that \(x_2' = g(p')\). Hence, if \(p'\) is greater than \(p\), \(x_2'\) will be greater than \(x_2\). If therefore we have a sequence of observations in which consumers' tastes are varying but production conditions, as expressed by \(g(p)\), remain constant, prices and production will move together. On the other hand, if consumers' tastes are constant but production conditions are variable, high prices will be associated with low production. Thus, mathematical symbolism resolves the apparent contradiction between the two arguments and shows when each is valid.

In addition to the problem of clarity in logical deductions, there is another methodological question related to the formulation of theoretical models: the problem of inductive inference. It is by now a platitude of the scientific method that if theory without empirical evidence is unreliable, empirical inquiry without theoretical background is unfruitful. The theory of statistical inference, as it has been developed in recent years by Jerzy Neyman, Egon S. Pearson, and Abraham Wald on the basis of the earlier work of R. A. Fisher and Karl Pearson, is precisely the mathematical expression of the logic of induction. It has shown clearly how the optimum statistical methods depend critically on the theoretical model assumed.\(^6\) Now the derivation of the statistical methods appropriate for making inferences within a given theoretical model is usually a matter of considerable mathematical difficulty.\(^7\) To proceed with the derivation at all, the underlying theoretical presumptions must be set forth in explicit symbolic form. Hence, a second argument, besides that of greater clarity of thought, for the explicit formulation of theories in mathematical terms is the resultant opportunity to tap the great resources of modern theoretical statistics as an aid in empirical verification.

The observations made above as to the limitations of our present knowledge of mathematics are applicable here also. It is unfortunately true that it is very easy to formulate theoretical models in which the determination of the optimum statistical methods leads to mathematical problems which have not been solved; in other cases, the resultant problems can be solved in principle, but the computations needed to find the solution in any given case take an impractical amount of time.\(^8\) Here again we must revert to simplification. The customary procedure is to substitute a mathematically practicable theory, as similar as possible to the desired one, and use that as the basis for deriving statistical methods. For example, the assumption is frequently made in theoretical models in biology and economics that there exists some linear combination of the variables that is normally distributed. The assumptions of normality and linearity are introduced primarily because of the relative mathematical simplicity of deriving optimum methods of estimation and tests of significance.\(^9\)

In Section II, some alternative possibilities in the types of mathematical models which are being proposed are discussed in a classificatory way. In Section III, extended examination is given to one proposed model, the "game" theory of social behavior developed by John von Neumann and Oskar Morgenstern. Section IV will take up more briefly some other proposed mathematical theories of the social sciences. Finally, certain very important recent advances in the methodology of empirical work as guided by theoretical models will be taken up in Section V. While the literature of current economic thought is frequently referred to, the emphasis is on those parts which have wide applicability in the social sciences.

II. SOME CLASSIFICATIONS OF MODELS

1. Individualistic versus collective basis

In most mathematical and, generally, in most deductive studies in the social sciences, the starting point is the behavior of the individual. Each individual is conceived of as acting in a way determined partly by his psychology and his physical

6. The concept of "optimum" is, of course, always relative to a value system. As will be seen below, the statement in the text is not meant to imply that all the philosophical difficulties in the theory of induction have been completely resolved. We can, however, be reasonably sure that the optimum methods, however defined, for making a choice among alternative theories on the basis of given empirical evidence will depend critically upon the theoretical background of the inquiry, i.e., upon the range of theories which are believed to be sufficiently plausible that they are admitted to the set among which the choice is to be made.

7. The level of mathematics needed in modern statistical theory can be seen in such a work as Harald Cramér, Mathematical Methods of Statistics (1946).

8. An attempt, for example, to estimate statistically the economic laws governing a large modern country on the basis of a relatively simple model (say, a few hundred equations to be fitted) could easily occupy the best computing machines now available for the next five hundred to a thousand years.

9. Of course there are other justifications for the widespread use of the normal distribution, in particular the Central Limit Theorem, which asserts that under certain general conditions the sum of a large number of independent variates, each contributing a small part to the whole, will be normally distributed. See, for instance, Harald Cramér, Random Variables and Probability Distributions (1937).
surroundings and partly by the actions of others. If there are \( n \) individuals, we may denote the actions of individual \( i \) by \( A_i \), and the nonsocial determinants of his behavior by \( P_i \). Then the actions of the first individual may be described by a symbolic equation.

\[
A_1 = f(P_1, A_2, \ldots, A_n)
\]

There is one such equation for each individual. Together they constitute \( n \) equations in the \( n \) variables \( A_1, \ldots, A_n \). In general, these may then be solved to express the actions of all individuals in terms of the data \( P_1, \ldots, P_n \). Therefore, given the reaction of each individual to his total (social and other) environment, as expressed in relations of type (1), and given the nonsocial environmental factors, which we may term exogenous, we can determine the behavior of society in the sense that we can determine the behavior of any individual in society.

This individualistic viewpoint as we may term it, is explicit in the main tradition of economic thought and is completely accepted in the von Neumann-Morgenstern game theory. It seems also to be accepted by the other theorists whose work is discussed below, though George Zipf seems at times to be referring to the laws of behavior of society in some total sense. The individualistic viewpoint has been challenged recently by Rutledge Vining in the course of a methodological controversy with Tjalling C. Koopmans. Vining has stated, "I think that in a positive sense the aggregate has an existence over and above the existence of Koopmans' individual units and behavior characteristics that may not be deducible from the behavior of these component parts." Taken literally, this position seems indefensible. As Koopmans points out, a full characterization of each individual's behavior logically implies a knowledge of group behavior; there is nothing left out. The rejection of the organism approach to social problems has been fairly complete, and to my mind salutary, rejection of mysticism. But as usual in these problems, there is something to be said for at least the possibility of a collective basis for social theorizing, if not taken too literally. Much of our casual observation and much of our better statistical information relate to groups rather than individuals. We may therefore be led to generalize and form a theory whose subject is the total behavior of a group. So long as it is understood that such a theory is really a resultant of certain as yet unanalyzed laws of individual behavior, no harm will be done, and the greater convenience of empirical analysis on groups may be highly beneficial.

In fact, even in economics, the unit of the theory of production is not really the individual but the firm, which is an operating organization of individuals. Similarly, the unit of the theory of consumption is really the household, not the individual consumer. In empirical economics, the investigator is usually forced to use a collectively based model by the nature of the data. The Keynesian theory postulates that the consumption of a community is an increasing function of its total income. As an empirical equation, this is a statement about group behavior, not individual behavior. Similar remarks apply to all macro-economic models designed as a basis for empirical fitting, in which the variables that enter are obtained by aggregating the behavior of many individuals. They apply even to many models constructed for the purposes of theoretical analysis. The various representations of the Keynesian theory in mathematical form all involve functional relations among magnitudes which cannot be identified with the behavior of any individual.
The usual feeling among economists is that these macroeconomic models could be justified on the basis of an individualistic theory if suitable definitions of the aggregate magnitudes in terms of those pertaining to individuals were given. The problem of finding such definitions, generally known as the aggregation problem, has received a certain amount of attention in recent years. Because of the nature of the discussion, at once technical and unresolved, it is not summarized here. But one methodological principle emerges clearly: in order to have a useful theory of relations among aggregates, it is necessary that they be defined in a manner derived from the theory of individual behavior. In other words, even the definition of such magnitudes as national income cannot be undertaken without a previous theoretical understanding of the underlying individual phenomena. It also seems evident that the aggregated model must include among its variables some which characterize the distribution of various magnitudes among the individuals of the society; e.g., in the consumption function some measure of income inequality should be introduced as an additional variable. Doubtless, the same remarks apply to aggregation in other social realms.  

2. The principle of rationality

A postulate frequently encountered in theoretical economics and elsewhere in social theory is that the behavior of the individual or group can be described by saying that the individual or group is seeking to maximize some quantity. Thus, in the theory of the firm, the economist postulates that the individual seeks to choose that mode and scale of operation which will yield more profit than any other possible choice. In the theory of consumption, it is assumed that among all the combinations of commodities an individual can afford, he chooses that combination which maximizes his utility or satisfaction. Behavior of the type is frequently referred to as rational.

The basis for the assumption of rationality is the following seemingly quite general formulation of individual behavior in a social situation: Each individual at a given moment of time is free to choose among several possible courses of action; he decides among them on the basis of their consequences. The range of actions open to him and the consequences of these actions are determined by the contemporary actions of others, by the past actions of himself and of others, and by the exogenous factors. To put it briefly, we may say that the individual can choose at any instant among a limited range of consequences according to his tastes. It is natural to suppose that as the range varies, his choices from different ranges bear some sort of consistent relation to each other. In particular, we will suppose that he makes the same choice each time he is confronted with the same set of alternatives; further, if, when confronted with the two alternatives A and B, he chooses A, and when confronted with the two alternatives B and C, he chooses B, then it is reasonable to suppose that when confronted with a choice between A and C, he chooses A. Under these circumstances, the process of decision may be described as follows: we can imagine the individual as listing, once and for all, all conceivable consequences of his actions in order of his preference for them; then, on being informed of the choices actually available to him, he selects that one which stands highest on his list. The list represents his tastes, while the alternatives actually available to him are restricted by obstacles.

So far there has been no quantitative element at all in the formulation. In certain problems, of course, the desirability of the various courses of action is associated naturally with a quantitative variable; e.g., profits in the case of a firm deciding upon its productive operations, probability of winning in the case of a country deciding whether or not to go to war. Even if there is no natural quantitative variable associated with the problem, it is sometimes useful for analysis to introduce one artificially, by assigning numbers to each consequence of an individual's actions in such a way that, of two consequences, the higher number will be assigned to the preferred one. Such an assignment of numbers is known as a utility index. Let x stand for the consequences.
of some action, and let \( U_1(x) \) be the utility assigned to \( x \) under one assignment of the type described. Then \( U_1(x_1) > U_1(x_2) \) if and only if the individual prefers \( x_1 \) to \( x_2 \). There is, clearly, nothing unique about the assignment of a utility index, at least within the conditions laid down thus far. Let \( F(x) \) be any strictly increasing function of a real variable \( u \), i.e., if \( u_1 > u_2 \), then \( F(u_1) > F(u_2) \). Then, if we define \( U_2(x) \) as \( F(U_1(x)) \), it is clear that \( U_2(x) \) will also serve as a utility index.\(^{19}\)

Under the hypothesis of rationality, then, the individual’s behavior depends on his tastes—as expressed, say, by a utility index—and upon the obstacles, which are determined by exogenous factors, by the actions of others, and possibly by past actions of the individual and of others. We may say that the individual maximizes his utility, subject to the obstacles. We are thus led to an expression of form (1) above. If we postulate nothing more than that there exists some utility function in terms of which the individual’s behavior can be described in the above way, we have imposed some restrictions on the possible forms of (1), since the principle of rationality is not a pure tautology; i.e., it is logically possible that an individual may choose \( A \) over \( B \), \( B \) over \( C \), and yet choose \( C \) over \( A \).\(^{20}\) However, the degree of restrictiveness thus obtained is not great. It is customary to supplement the general principle of rationality in any particular instance by further assumptions as to the nature of the preferences involved. For example, in consumer’s demand theory, it is ordinarily understood that more of a commodity is preferred to less, all other things being equal.

A number of objections have been raised to the usefulness of the principle of rationality: (a) If the complicated nature of the range of choices possible in an actual social situation is even approximately taken into account in the theoretical model, the mathematical problems to be solved in the maximization of utility will become extremely complex, and it will be hard to derive results which have any simple meaning. This objection has been frequently raised, for example, against the Walrasian scheme of general equilibrium in economics. (b) There is no real reason to suppose that individual behavior does conform to the principle of rationality. This argument is partly related to the previous one; it is argued that if the rational choice is too difficult for the trained mathematician to find, it is certainly unreasonable to suppose that the untrained, reflecting, average individual will be able to locate it.\(^ {21}\) (c) The utility function itself, even if it plays the role assigned to it, is highly unstable over time; hence, for an understanding of social processes, more interest attaches to the determinants of the variation of tastes than to the line of causation from the utility index to the actual decision made.\(^ {22}\) (d) There is a fundamental ambiguity in the concept of rationality in a social situation. An individual will soon realize that his actions, in addition to their other consequences, will alter the obstacles faced by others, thereby affecting their actions and in turn altering the obstacles controlling his choices. Hence, his actions will be partly controlled by his realization of their repercussions on the actions of others. But the same statement is true of each other individual; thus, each will be concerned with the effect of his action on the others, and no determinate solution will be possible.\(^ {23}\)

There is no single sweeping principles which has been erected as a rival to that of rationality. To the extent that formal theoretical structures in the social sciences have not been based on the hypothesis of rational behavior, their postulates have been developed in a manner which we may term ad hoc. Such propositions are usually drawn from introspection or casual observation; sometimes they are of the nature of empirical regularities. They depend, of course, on the investigator’s intuition and common sense.

An example of this approach is Lewis F. Richardson’s approach to international relations.\(^ {24}\) What will make a government increase its armaments? Clearly, it will increase them more if the armaments of a potential enemy are greater;

19. It should be clearly understood that the variable \( x \) need not be numerical. The consequences of an action may be a power or prestige situation or a state of religious ecstasy, as well as a bundle of commodities for consumption or a sum of money. All that is required is that given two well-defined situations resultant from his actions, an individual should be able to say that he prefers one to the other.
20. For a discussion of the refutable, and therefore empirically meaningful, consequences of the principle of rationality in the field of consumers’ demand, see Paul A. Samuelson, Foundations of Economic Analysis (1947), chapt. v.
23. This fourth objection to the principle of rationality has been recognized in a general way in economics in the theories of oligopoly and bilateral monopoly. It was given a definitive statement by Oskar Morgenstern in his Wirtschaftsprognose (1928), p. 98.
24. Lewis F. Richardson, Generalized Foreign Politics (British Journal of Psychology, Monograph Supplement No. 23 (1939)).
but it will be deterred from increasing them by
the expense. Let \( x \) represent the armaments of
this country, \( y \) those of a potential enemy, and \( t \)
time; and let \( k \) and \( a \) be fixed coefficients. Then,
the above theory may be expressed in the equa-
tion
\[
dx/dt = ky - ax. \quad (2)
\]
A similar equation holds for the other country.
Taken together, they form a complete system of
differential equations which, with the initial condi-
tions and the values of the coefficients, define the
course of \( x \) and \( y \) over time. Richardson then
deals with the implications of these results and
those of more general models.

Because of the first and fourth difficulties
mentioned above, theoretical economics has gen-
erally had an admixture of ad hoc assumptions
which limit the scope of the principle of rational-
ity. The hypothesis of a perfect competition re-
moves the fourth difficulty because each individual
supposes his effect on other individuals is so
small that their actions are not influenced by him.
Under conditions of monopoly it is assumed that
while one individual can affect other individuals,
one of them is strong enough to affect him. The
monopolist's behavior incorporates his realization
of this situation, thereby again avoiding indeter-
minancy. In those cases where neither hypothesis
can be maintained (e.g., bargaining between a
labor union and a monopolistic employer, or the
behavior of an industry in which there are a few
large firms) still more complicated ad hoc as-
sumptions have been made about the way individual
firms took into account the anticipated reactions
to their actions in order eventually to reduce the
problem to a simple maximization.25

One somewhat digressive remark on the prin-
ciple of rationality may be in order: a rational
theory always has a dual interpretation. On the
one hand it may be taken as a description of
reality to the extent that individuals really are
consistent in the sense assumed. On the other
hand it may be taken rather as a normative
theory, which prescribes what individuals ought
do. Thus, theoretical economics has been
used to analyze what the optimum state of eco-
nomic welfare would be and how to attain it.26

This subject of "welfare economics" is not new,
being indeed as old as theoretical economics
itself, but clarification of its basic principles has
been a slow process. Even today there are a
number of profound unresolved difficulties, prin-
cipally revolving about the problem of comparing
the welfare of different individuals in arriving
at a concept of a social optimum. 27 Since we
are here concerned with the problems of a de-
scriptive social science, we shall not pursue this
matter any further.

Statistical inference may be viewed broadly
as the behavior of an individual under a certain
set of conditions; namely, he does not know com-
pletely the consequences of his actions. We may
therefore speak of a rational theory of statistical
inference. More precisely, statistical inference
may be described as follows: The true state of
the universe under investigation is known, a priori,
to belong to one of a class of states. A sample
of elements of the universe is drawn; the proba-
bility distribution of the sample depends on the
true state of the universe. The statistician must
then take some action (e.g., estimate some state
to be the true one, assert that the state lies in
some subclass of the one given a priori; or, in
industrial applications, accept or reject a lot of
goods), the consequences of which are a function
of the action and of the true state, being favorable
if the action is really appropriate to the true state
and unfavorable otherwise. For simplicity, let us
consider the case where the true state is known
to be one of a finite number, the possible sam-
ples to be drawn are finite in number (e.g., sup-
pose we draw a sample of three observations,
each of which is either "yes" or "no," as in a
questionnaire or in a quality inspection; then one
possible sample is "yes, yes, yes," another is "no,
yes, no," and so on; there are altogether eight
possible samples), and the number of actions the
statistician can take is finite. Let \( i \) stand for an
action, \( j \) for a true state, \( k \) a sample, \( p_{kj} \) for the
probability of observing sample \( k \) when the true
state is \( j \), and \( r_{ij} \) the loss to the statistician if
he takes action \( i \) when the true state is \( j \) \( (r_{ij}\)
is small or negative if \( i \) is appropriate to \( j \), large
otherwise). The statistician's problem is to choose
a decision function \( i(k) \), which tells him what ac-
tion to take for each possible sample \( k \). The
function \( i(k) \) is to be so chosen as to minimize in
some sense the probable loss.

This formulation includes all the classical
problems of statistics and more. For example,
in the above illustration the problem of estimation
is the case where an action consists of naming a
true state. The range of possible actions is then
the same as the range of possible states of nature.
Another case would be that of two possible actions,
one affirming and one denying that \( j = J \), where
\( J \) is a fixed possible true state. This is the
classical problem of testing a hypothesis.28

25. The most extended treatment of various possible assumptions of this type is found in Ragnar Frisch,
"Monopole, polypole: La Notion de la force dans l'équilibre économique," Nationalökonomisk Tidskrift (1933).
26. Recent systematic expositions of this subject will be found in Paul A. Samuelson, Foundations of Economic
Analysis (1947), chap. viii, and Melvin W. Reder, Studies in the Theory of Welfare Economics (1947), Parts I and
III.
28. This general formulation of the problem of statistical decision is due to A. Wald, "Foundations of a Gen-

439
We have here clearly a problem of behavior; the statistician must choose among the various possible functions \( i(k) \), and the consequences are given, though not with certainty, by the conditions of the problem, the \( p_{kj} \)'s and \( r_{ij} \)'s. The foundations of statistical inference, from the normative point of view, of proper scientific foundations of statistical inference, from the normative point of view, are then nothing but the application of a suitable principle of rationality to the problem just described. Indeed, it will be seen that Wald's important contributions to this field are closely related to an important development in social theory, the von Neumann-Morgenstern theory of games.

This relation between the theoretical analysis of behavior and the foundations of statistical inference can, of course, be applied in reverse. To the extent that actual behavior under conditions of uncertainty as to the consequences of any action is governed by the principle of rationality, the theory of statistical inference may also be interpreted as a descriptive theory. This point of view has especially been stressed in economics by Jacob Marschak.

III. THE THEORY OF GAMES

The von Neumann-Morgenstern theory of games is an attempt to provide a theory of social (primarily economic) interaction by analogy with ordinary games of strategy (such as chess or card games) as they would be played by thoroughly rational individuals. It goes well beyond any other systematic social theorizing in the complexity of its structure and the rigorous nature of its formal logic. Yet it is interesting to note that virtually no mathematics more difficult than algebra is employed although the chains of reasoning are frequently long and complicated.

1. Rational behavior in situations involving risk

As a preliminary to the theory of games, though not strictly part of it, von Neumann and Morgenstern formulate the principle of rationality applied to a situation in which the consequences of the different actions open to an individual are expressed as probability distributions rather than as certainties. A simple example is that of an individual choosing between two lottery tickets, one of which pays $1,000 with probability 0.01 and nothing otherwise, while the other pays $100 with probability 0.10 and nothing otherwise. More generally, the choice is among a number of "lottery tickets," each of which is a promise to pay amounts \( m_1, m_2, \ldots \) with probabilities \( p_1, p_2, \ldots \), the quantities \( m_1, m_2, \ldots \) and \( p_1, p_2, \ldots \) varying from ticket to ticket. Also, of course, in general the "payments" need not be sums of money but consequences of any sort which matter to the individual. A "lottery ticket" is a paradigm for any choice made by a human being in which the outcome of his action is not known with certainty but in which, on the basis of experience or intuition, he believes he knows the probabilities of the different possible outcomes.

In accordance with the principle of rationality, the individual may order all possible probability distributions of outcomes; he may then choose a utility index having the properties described in the second part of Section II of this chapter and act so as to maximize utility. The concept of rationality as applied to this situation has been formalized by von Neumann and Morgenstern in a series of axioms. The principal condition laid down, in addition to the general requirement that the probability distributions be ordered in accordance with the individual's preferences, may be expressed loosely as follows: If the individual prefers the outcome \( x_1 \) to the outcome \( x_2 \), then...
he will prefer the certainty of \( x_1 \) to an even chance of getting either \( x_1 \) or \( x_2 \). With the aid of this highly reasonable condition, and some other more technical ones, it is shown that among the many utility indexes which can be assigned to the possible probability distributions, there is one with the property that the utility attached to a probability distribution is the mathematical expectation of the utilities of the possible outcomes. I. e., if a given choice has possible outcomes \( x_1 \) \( \ldots \) \( x_n \) with probabilities \( p_1 \) \( \ldots \) \( p_n \), respectively, and if \( U(x) \) is the utility (in the index just described) of the outcome \( x \), then the utility attached to the given choice is \( \sum_{i=1}^{n} p_i U(x_i) \). (Note that this utility index is not unique; if \( U(x) \) is one such index and \( a \) is a positive number, than \( aU(x) + b \) is another such index.) Rational behavior in the choice of risky alternatives can thus be simplified to the statement that the individual seeks to maximize the expected value of his utility.

Milton Friedman and L. J. Savage have developed an interesting hypothesis about the shape of the von Neumann-Morgenstern utility curve for money which will explain some phenomena observed in the behavior of individuals in connection with gambling and insurance.

Since probability distributions are basic in the theory of games, the above construction enables the authors to speak of a numerical utility, having the convenient property that the utility attached to a chance event is precisely the mathematical expectation of the utilities of the various possible outcomes. However, they need further the idea of a transferable utility, i. e., some sort of measure the units of which mean the same, in some sense, to both players. In effect, they then restrict the realm of their theory greatly by assuming that the outcome of each game is expressible in some common units, which we may take to be money, and that the utility of a given sum of money, for each player, is simply a linear function of the sum, so that each players is seeking to maximize the expected amount of money he will receive. This assumption is to be understood as an intentional simplification to make the problem mathematically manageable. (Compare the remarks in the first section above.)

2. The general concept of a game

In general, a game will have a certain number of players, say \( n \). The game is composed of moves, which are of two types; personal, made by one of the players, and chance, in which one of several possible outcomes is selected by a chance device acting in accordance with certain probability laws. Thus, in a game of cards the distribution of the hands is made by a random device, while the players still have certain choices as to how to play their hands. The rules of the game prescribe the following; (a) whether the first move shall be personal or chance; if personal, who shall make it, and what range of choice he shall have; if chance, what the possible outcomes are and what the probability is of each; (b) after \( k \) moves have been made, whether the \((k + 1)\)st move shall be personal or chance; if personal, who shall make it, what range of choices he shall have, and what information he shall have about the outcomes of the previous moves; if chance, what the range of possible outcomes is and what probability is attached to each (note well that the rules may make each of these prescriptions dependent on the outcome of the preceding moves, it being required only that at each stage of the game the prescriptions in question must be unambiguously defined); (c) when the game shall stop and what amount shall be paid to each of the participants depending on the outcomes of the various moves. Thus, in chess, the rules specify that all moves are personal, with the two players taking alternate moves; that at each move each players knows the outcome of every previous move; and that in any given move, the range of permissible actions depends upon the present configuration of the pieces, which in turn depends upon the outcomes of all past moves. In card games, on the other hand, the dealing of the hands is a chance move, the outcome of which is not revealed to the other players.

It is easy to see that many economic and social situations can be described in these terms.

32. The explanation of choice among risky alternatives in terms of maximizing the expected value of utility dates back to Daniel Bernoulli, "Specimen theoriae novae de mensura sortis" (1738); German translation, Die Grundlage der modernen Wertlehre (1895). Bernoulli was interested in showing, in connection with the famous St. Petersburg paradox, that an individual might not take a gamble even if the expected money reward were infinite; his argument was that the individual was seeking to maximize the expected utility of money, which was given, he suggested, by the logarithm of the amount of money received. With the introduction of the marginal utility theory of value, it was natural to discuss choices among risky alternatives in these terms; see Alfred Marshall, Principles of Economics (8th ed., 1920), pp. 135 n., 843. However, Von Neumann and Morgenstern have been the first to show the assumptions about behavior which are implied in this theory.


34. See von Neumann and Morgenstern, op. cit., chap. II.
For example, the operations of a market may be described as a game, with the buyers and sellers as players, and bids, offers, and agreements to conclude transactions as moves. The initial distribution of goods and money may be taken as part of the rules of the game; the final distribution of goods and money describes the pay-off to the various players attendant upon the particular set of moves which they made. Similarly, conflicts between nations or between social and economic groups can be brought within the same rubric.

Now suppose, for the purpose of simplified description, that each individual does not play the game by waiting each time until his move comes and then deciding among the various alternatives presented to him, but rather that he prepares beforehand a strategy which specifies for each possible situation with which he may be confronted in the course of the play what choice he shall make. Indeed, something of the sort must be done by skillful players, since they realize that the effects of their choices at various moves are interrelated; a good move at one point may improve the range of choices available at another. A good choice of any one move may therefore involve consideration of the alternative possible developments of the game following each possible action. Suppose each individual were to write down his strategy and hand it to an umpire. It would then be unnecessary actually to play the game; the umpire would need only to follow the instructions of the appropriate individual's strategy at each personal move and make the indicated chance moves. Therefore, every game can be reduced to the following type: each player makes a single move, namely the choice of a strategy; then the pay-off to each player is a random variable whose probability distribution is determined by the strategies of the various players. Since, according to the assumption made at the end of Section I, the individual is interested only in the mathematical expectation of the pay-off we can consider the pay-off to be a single number, instead of a probability distribution. Thus every game, not matter how complicated, can be reduced to the following normal form: each player makes one move, knowing nothing about the moves of the others, from a range of choices fixed in advance; the pay-off to each player is then a function of the moves of all the players. This reduction permits a consideration of the basic nature of games undisturbed by the complications of particular rules. From now on, we will consider all games to be in normal form.

To simplify the analysis further, von Neumann and Morgenstern assume that the game will be terminated in a finite number of moves and that at each move there are only a finite number of possible choices. Then it can be seen that the number of possible strategies is finite. In the normal form, therefore, each individual has the choice among a finite number of alternatives for his move. It should be noted, though, that the finite number of strategies may be enormous; merely to enumerate all possible situations in chess is a tremendous task, and the number of possible ways of specifying responses to these situations stagger the imagination. If the restriction to a finite number of choices at each move is not made, then some extremely subtle mathematical difficulties arise. Some work has been done on this problem; it will be cited in footnotes where appropriate.

3. The Zero-sum two-person game

Consider now the special case of a game with two players in which the sum of the pay-offs to the two players is zero for all choices of strategies; e.g., if the outcome of a play of the game yields 5 to player a, then it necessarily constitutes a loss of 5 to player b. This is the model of a situation of pure conflict, since the gain of one player is precisely the loss of the other. Since there are only two players, there is nothing whatever to be gained by agreement between them. Assume that the game is in normal form; let player A have the choice of strategies 1, 2, 3, 4, 5, player B the choice of strategies 1, 2, 3, 4, 5. If A chooses strategy i and B strategy j, let the pay-off to player A be \( a_{ij} \); then the pay-off to player B is \(-a_{ij}\). Clearly, A wishes to maximize \( a_{ij} \), while B seeks to minimize the same magnitude. Suppose A is contemplating the use of a particular strategy i. It is natural for him to think of the worst that B could do to him; the latter will try to choose that j which will make \( a_{ij} \) as small as possible for the given strategy i. Hence, if player A chooses i, all he can be sure of is \( \min a_{ij} \) (read, "minimum with respect to j of \( a_{ij} \)). He can evaluate this magnitude for each value of i and choose that value of i which maximizes it; by this choice, he insures that his return is at least max \( a_{ij} \). Similarly, by suitable choice of j, player B can guarantee that player A will not receive more than min \( a_{ij} \). The pay-off matrix \( a_{ij} \) is such that,

35. The concept of the economic agent as choosing a strategy -- a plan for meeting contingencies -- rather than merely making specific decisions from time to time, was clearly expressed, in a somewhat different context, by Albert G. Hart in his "Risk, Uncertainty, and the Unprofitability of Compounding Probabilities," in Studies in Mathematical Economics and Econometrics, edited by Oscar Lange, Francis M. McIntyre, and T. O. Yntema (1942), pp. 110-18.


442
\[
\max_i \min_j a_{ij} = \min_j \max_i a_{ij}. \tag{3}
\]

If the common value of the two sides of (3) is \( v \), then A can, by choosing a suitable strategy \( i' \), guarantee himself at least \( v \), while B, by choosing a suitable strategy \( j' \), can guarantee that A will not get more than \( v \). It is clear that in this case, A will choose \( i' \), B will choose \( j' \), and the outcome of the game (termed its value) will be that B pays an amount \( v \) to A. (The quantity \( v \) could be negative, in which case the payment would be from A to B.) Equation (3) can be valid, as shown in the pay-off matrix given in Table I in which each player has the choice of two strategies.

**TABLE I**

<table>
<thead>
<tr>
<th>Strategies of Player A</th>
<th>Strategies of Player B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

If player A chose strategy 1 and player B chose strategy 2, A would receive 1 unit from B. If A chose strategy 2, he would run the risk of receiving nothing; if B chose strategy 1, he might have to pay 3. Hence, \( i' = 1, j' = 2 \), and \( v = 1 \) represent a stable solution; even if one player knew in advance the other one's strategy, he would have no incentive to change his own.

Unfortunately, equation (3) need not hold. Consider the pay-off matrix in Table II.

**TABLE II**

<table>
<thead>
<tr>
<th>Strategies of Player A</th>
<th>Strategies of Player B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

In this case, \( \max_i a_{ij} = 0 \), \( \min_j a_{ij} = 1 \). If a player B played so as to minimize \( \max_i a_{ij} \) he would choose strategy 1; but if player A knew this, he would choose strategy 1, rather than strategy 2, which he would choose if he were maximizing \( \min_j a_{ij} \).

If, then, B realized that A would reason in this way, he would choose strategy 2, and so forth, leading to a seemingly endless regress. Indeed, this is precisely a chief objection to the principle of rationality, as noted in the second part of Section II, above.

To arrive at a determinate stable solution, the following procedure is suggested: let player A have a random device which will choose strategy 1 with probability 0.4 and strategy 2 with probability 0.6. Then, if player B selects strategy 1, player A will have an expected return of \( 0.4(1) + 0.6(0) = 0.4 \); if B chooses strategy 2, the expected return is \( 0.4(-2) + 0.6(2) = 0.4 \). Hence, A can guarantee himself an expected return of 0.4. Similarly, B, by playing strategy 1 with probability 0.8 and strategy 2 with probability 0.2, can hold the expected pay-off to A down to 0.4. This solution, in terms of randomized or mixed strategies, has the stability found in the first case; even if one player found out the probabilities used by the other, he could not gain from this information.

In general, the players can be considered not as choosing a strategy but as choosing a probability for each strategy (possibly zero). Let player A play strategies \( 1, \ldots, m \) with probabilities \( x_1, \ldots, x_m \), respectively, while player B plays his strategies with probabilities \( y_1, \ldots, y_n \). Let \( \max \) mean the maximum with respect to permissible variations in \( x_1, \ldots, x_m \) and let \( \min \) have a corresponding meaning. The expected payoff to A if he chooses \( x_1, \ldots, x_m \) and B chooses \( y_1, \ldots, y_n \) is

\[
\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j.
\]

Then the analogue to (3) is

\[
\max_x \min_y \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j = \min_y \max_x \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j. \tag{4}
\]

If (4) holds, then we have a stable solution in terms of mixed strategies. It turns out, rather remarkably, that (4) always holds.\(^{37}\) The theory

of rational behavior in zero-sum two-person games can therefore be regarded as definitely solved, at least within the limitations of a transferable utility. However, the objection that the principle of rationality cannot be realistic because it imposes too great burdens on the ability of the individual is not refuted; the solution of even relatively simple social games is frequently very difficult.38

Before leaving the discussion of zero-sum two-person games, consider the special case of a game in which, before reduction to normal form, the rules provide that at every stage of the game each individual knows the outcome of all previous moves, i.e., a game of perfect information. Chess and backgammon are familiar examples. The more enduring features of many social conflicts are of this type; there is no possibility of concealing the existence of a major strike from the management, though certain preliminary tactical plans may be concealed for a while. It is shown that for such games relation (3) always holds. Mixed strategies become unnecessary for a determinate solution. This suggests that mixed strategies may be regarded essentially as a device for withholding information from the opponent, for bluffing.

4. The zero-sum n-person game39

Consider first the following simple game: There are three players; each is to write down the name of one of the others. If two players write down each other's name, they each get 1/2, and the third player pays 1. If no two players write down each other's name, then no one gets anything. The rational method of playing seems fairly obvious; two of the players will agree in advance to write down each other's name and then collect. In what sense is this really a determinate solution? In particular, it is to be noted that the solution is not unique; there are three possible pairs of players who can form coalitions. Von Neumann and Morgenstern argue as follows: Consider all possible distributions of payments to the players. These will include transfers not provided for by the rules of the game, since we permit one player to bribe another into following an acceptable course of action. Since the game is zero-sum, the sum of the payments must be zero; since each player can get -1 without any coalition, no distribution can be enforced which gives any player less than -1. Let \((a_1, a_2, a_3)\) denote a distribution which gives \(a_1\) to the first player, \(a_2\) to the second player, and \(a_3\) to the third player. Then, in symbols,

\[
\begin{align*}
  a_1 + a_2 + a_3 &= 0, \\
  a_i &
\end{align*}
\]

Any set of numbers satisfying (5) and (6) will be termed imputation. The particular imputations which seem intuitively to be the rational ones are, then, \((1/2, 1/2, -1)\), \((1/2, -1, 1/2)\), and \((-1, 1/2, 1/2)\). Let \(V\) be the set consisting of these imputations. Then the essential properties of \(V\) are (1) for every imputation not in \(V\), there are two players who would both prefer a particular imputation in \(V\) to the one not in \(V\); (2) for any two imputations in \(V\), there are not two players who would prefer one to the other. These are stability properties; no imputation not in \(V\) can be maintained, while there is no drive to change from one imputation in \(V\) to another, since in our simple game the behavior of pairs of players is decisive.

The decisive character of a pair of players who choose to agree consists in the following: any pair of players can, by agreement, get the amount 1 between them, while no imputation can, from (5) and (6), give them more than 1. Therefore, they are in a position to choose any imputation whatever, since no imputation gives them more than they can get by the rules of the game. These concepts all lead to natural generalization. Consider the following zero-sum n-person game: Let \(S\) be any set of players; let the rules specify a characteristic function \(v(S)\) which states the total amount any set \(S\) will get if they form a coalition. Because of the zero-sum character of the game, \(v(-S) = -v(S)\), where \(-S\) is the set of all players not in \(S\). In particular, let \((i)\) be the set consisting of the player \(i\) alone, so that \(v([i])\) is the amount player \(i\) will get (presumably negative) if he does not enter into a coalition with anyone. Then an imputation is a set of payments \((a_1, \ldots, a_n)\) to the respective players having the properties,

\[
\begin{align*}
  a_1 + \ldots + a_n &= 0, \\
  a_i &
\end{align*}
\]

A set of players \(S\) is said to be effective for an imputation \((a_1, \ldots, a_n)\) if \(\sum_{i \in S} a_i = v(S)\), where


38. See the detailed discussion needed for even a highly simplified form of poker in von Neumann and Morgenstern, op. cit., chap. iv, sec. 19.
39. Ibid., chaps. v and vi.
"i ∈ S" means "i belongs to S," i.e., if the imputation does not give the players in S more than they could get by forming a coalition. The imputation \((a_1, \ldots, a_n)\) is said to dominate the imputation \((b_1, \ldots, b_n)\) if there is a set \(S\) which is effective for \((a_1, \ldots, a_n)\) and such that \(a_i > b_i\) for every player \(i\) in \(S\); under these circumstances we would certainly expect the players in \(S\), by agreement, to change the imputation from \((b_1, \ldots, b_n)\) to \((a_1, \ldots, a_n)\).

A set of imputations \(V\) is said to be a solution if (1) every imputation not in \(V\) dominates another imputation in \(V\); (2) no imputation in \(V\) dominates another imputation in \(V\). A solution thus defines the outcome of the above game when the players behave rationally.

It might seem that the above discussion relates only to a very special type of game, in which coalitions enter through the explicit statement of the rules. It is argued, however, that every \(n\)-person zero-sum game can be expressed in this manner. Suppose a coalition \(S\) forms, which seeks to maximize the sum of the pay-offs to its members. The coalition will naturally suppose that the players not in \(S\) will form a countercoalition; in view of the zero-sum character of the game, the coalition \(-S\) will seek to minimize the sum of the pay-offs to members of \(S\). This situation is precisely a zero-sum two-person game, which has a determinate value, as has been seen. Therefore, we may define \(v(S)\) to be the value to the coalition \(S\) of the zero-sum two-person game just described. This can be done for each possible set of players \(S\), so that the game is reduced to the form discussed in the previous paragraph.

Games are classified as essential or inessential. An inessential game is one in which a coalition can get no more than the sum of the amounts that each of its members could get by playing separately, i.e., \(v(S) = \sum_{i \in S} v([i])\). In such a game, there is no incentive to form coalitions and hence no specifically social element.

In an essential game, on the other hand, such as the three-person game which introduced this discussion, there is a positive incentive to form coalitions. We have already observed that a solution may contain more than one imputation. This is a general phenomenon: it can be shown that in an essential game, a solution must contain at least two imputations. There seems to be some sort of fundamental indeterminacy in the purely rational theory of social behavior; the true significance of this indeterminacy must be regarded as obscure.

The definitions as framed do not exclude the possibility that there is more than one solution to a given game, and indeed this can happen, although there are also games with only one solution. In the three-person game just considered, let \(c\) be any number less than \(1/2\). Consider the class of all imputations in which \(a_3 = c\). It can easily be verified that this class is in fact a solution, within the meaning of the previous definitions. For each value of \(c\), there is a corresponding solution, so that the number of solutions is infinite. Unlike the first solution, these solutions do not have any symmetry. They may be termed discriminatory solution; the ruling coalition \((1, 2)\) agrees to give the excluded player a fixed amount \(c\), and then divide up the remainder between them. It is remarkable that discriminatory practices turn out to have certain stability properties even when the basic rules of the game are completely symmetric as between players. Even within the first solution found, there was a certain element of unfairness; each given imputation within the solution was unfair to one player, but there was an element of symmetry in the solution as a whole. So it can be seen that even a symmetrical game is not necessarily fair, in that it does not guarantee that each individual will be as well off as every other. Much more than equality of opportunity is needed to insure equality of outcome.

There is thus a hierarchy of successively more restrictive conditions on the state of society: the rules of the game, which define the class of possible imputations, the solution, and the imputation which actually prevails at any given instant of time. It is suggested that the first corresponds to the basic limitations of the society—physical, biological, and possibly psychological; the second to a socially stable standard of behavior which, however, does not uniquely determine the actual distribution of values among the members of the society. More than one standard of behavior may be compatible with the rules, but only those found by the theory will be stable enough to survive.

5. The general \(n\)-person game

The games thus far considered have lacked generality in one important respect. Society as a whole could not gain because of the zero-sum restriction. In the study of many social conflicts, this may not be an important limitation, but it does serve to eliminate all economic aspects. The crucial feature of economic behavior is precisely the possibility that by co-operation all individuals can be made better off.

The general \(n\)-person game is handled by a reduction to a zero-sum \((n + 1)\)-person game. Introduce a fictitious player, \(n + 1\), who can make no moves but who receives the negative of the
This enlarged game can be described in terms of the theory in the preceding section; \( v(S) \) again represents the amount that each coalition can guarantee for itself. The concept of domination goes through as before. However, not all the solutions of the extended game can be accepted. We may take it for granted that the genuine players will get as much for themselves collectively as they can, since this increases the amount of wealth to be divided among them. In effect, the genuine players will discriminate against the fictitious player. The only solutions that are therefore admissible are those in which every imputation gives the fictitious player the minimum possible.

In the case of a general two-person game, it turns out that there is just one solution, namely, the set of all imputations which maximize the sum of the returns to the two players. Suppose A and B are bargaining over a house which A owns but B values more highly. Then the house will be transferred to B, who will pay a price at least equal to the value A places on it but not more than the value of B. This conclusion extends readily to all cases of bilateral monopoly, even when there is not a unique indivisible object but a continuously divisible commodity: the amount of goods which changes hands is determinate, being that amount which maximizes the sum of the profits of the two bargainers, but the buyer will pay the seller at least enough so that he will not be worse off than if he sold nothing and not so much that the buyer will be worse off than if he bought nothing.41

The discussion of the two-person game yields a purely common-sense answer. Much less obvious results follow in the general three-person game, e.g., one seller and two buyers. In the case of a single indivisible object, again, it will be transferred to the individual who values it most highly, but the possible accompanying patterns of payments may be very complicated indeed. The game theory fully takes into account such possibilities as having one buyer bribe another to stay out of the market. The determination of the solutions to this game cannot be presented here in detail.42

6. Statistical inference as a game.

In the discussion of statistical inference as a problem in rational behavior (in the second part of Section II, above), it has been shown that the statistician must choose a function \( i(k) \) which determines the action \( i \) he will take for any possible sample \( k \). Suppose he chooses such a function, while the true state of nature is \( j \). If the sample \( k \) should come up, he will take action \( i(k) \) and therefore lose \( r_{i(k),j} \). But sample \( k \) will come up with probability \( p_{k,j} \). Therefore, if the statistician chooses a function \( i(k) \), his expected loss when the true state in \( j \) is given by

\[
\sum_k r_{i(k),j} = p_{k,j},
\]

which depends on the statistician's choice of \( i(k) \) and nature's "choice," \( j \). Wald has suggested interpreting statistical inference as a zero-sum two-person game, in which nature chooses a state \( j \), the statistician chooses a decision function \( i(k) \), and the pay-off to nature is given by \( (9) \). The optimal behavior of the statistician is to choose that decision function (possibly mixed) which minimizes the maximum expected loss.

IV. SOME OTHER MATHEMATICAL MODELS OF SOCIAL BEHAVIOR

1. Rashevsky's theory of human relations43

Nicolas Rashevsky of the University of Chicago, who is known primarily for his work in the application of mathematical methods to biology, has suggested that analogous methods could be applied to the study of human relations. His approach is of a generally ad hoc character, partly collective, partly individualistic. The bulk of his work is devoted to the interaction of social classes. A great variety of different hypotheses as to the nature of this interaction are proposed and formulated mathematically, and some consequences drawn. An example of his analysis will suffice to give the general flavor. It is assumed that there are two classes, the active and the passive. Active individuals do not imitate anyone, while passive individuals are influenced by active individuals in proportion to the number with whom they come into contact and also, to a lesser extent, by other passive individuals. Suppose that there are two possible activities, \( A \) and \( B \), that \( x \) and \( y \) of the active individuals engage in \( A \) and \( B \) respectively, and that at any instant of time \( x \) and \( y \) passive individuals engage in \( A \) and \( B \) respectively. The number of passive individuals engaged in \( A \) will tend to be increased by the

41. The earlier discussions of this problem in economic literature had given smaller ranges of indeterminancy; the contribution of game theory is to show that the earlier answers were artificial from the viewpoint of rational behavior (though not necessarily from that of actual behavior). However, Gerhard Tintner has already given the solution indicated above. See his "Note on the Problem of Bilateral Monopoly," Journal of Political Economy, XLVII (1939), 263-70.


number of active individuals engaged in A and, to a lesser extent, by the number of passive individuals engaged in A, and to be decreased by the number of active and passive individuals engaged in B. Symbolically,

\[ \frac{dx}{dt} = a_0 x_0 + ax - c_0 y_0 - ay, \quad (10) \]

with a similar equation for \( \frac{dy}{dt} \). These equations can be solved to reveal the variation in x and y over time as functions of the coefficients and to find the points of stable equilibrium to which the system will tend.\(^44\)

It is to be noted that the situation in question might be interpreted as a conflict between the active individuals engaged in A and those engaged in B for control over the passive individuals. It could easily therefore be brought within the general framework of game theory. It is noteworthy that Professor Rashevsky's model is broadly similar to models of conflict in other spheres, such as Richardson's work on international relations—compare (10) with (2) above—and Lanchester's theoretical treatment of military strategy.\(^45\)

There seem to be emerging two general analyses of social conflict situations employing mathematical methods: rationalistic game theory, and ad hoc dynamic analysis by means of systems of linear differential equations. There may be some complementarity in the two approaches; in particular, note that game theory, at least in its present state, deals only with equilibrium situations. If additional assumptions could be made as to how each player learns about the others from experience, the theory would be a dynamic one.

Rashevsky presents the above case as highly oversimplified and suggests more complicated variations of the same scheme. Some of these involve economic considerations in that the interaction of classes may be productive of wealth.\(^46\)

In the economic analyses, the ad hoc laws which relate, e.g., to certain aspects of the distribution of income, are not sufficient to define the situation uniquely; within the framework of these postulates, the individuals are assumed to act rationally in some appropriate sense, i.e., to maximize satisfaction or profits. One simple case is that of two classes of individuals, one of which is composed of organizers under whose supervision the members of the second class produce more than they would otherwise. It is assumed that the members of the first class decide on a fixed fraction of the total output to be divided among the members of the second class in proportion to the amount of work each does, while each member of the second class decides how much work to do on the basis of the given fraction. The first class then chooses the fraction in question so as to maximize their total profits. There are several ad hoc elements here; the fraction to be given to the second class need not be independent of the amount of work they do, it need not be divided among the members of that class in proportion to the amount of work each does, and each member of the first class may try to maximize his own profits, possibly by special bargains with some members of the second class, rather than maximize the total class profit. Some of these assumptions are similar to the assumption of perfect competition in classical economics.

The same techniques of analysis are used throughout Rashevsky's work, though there are many variant sets of assumptions made. Applications are made to the urban-rural distribution of population, the distribution of city sizes, international relations, historical change, and war. Rashevsky emphasizes the tentative and simplified character of his approach and suggests that its chief function is to make available a number of alternative models which might be useful in further theoretical and empirical work.

It might be remarked that the standards of mathematical rigor are high. The methods used are drawn from the calculus and the theory of ordinary linear differential equations, with a few tentative steps toward the use of integral equations.

2. Zipf's principle of least effort and Stewart's social physics

The work of George K. Zipf\(^47\) is an extraordinarily comprehensive effort to subsume the major part of human behavior, both individual and collective, under a single principle of least effort: Individuals and societies act so as to minimize the expected average rate of work. We shall here be concerned only with his discussion of social behavior in Part Two of his book. In terms of the classification in Section II of this chapter, the model is certainly rational; since it seems to be assumed that both individuals and society seek to minimize effort, the model has both collective and individualist elements.

It must be stated, however, that Dr. Zipf's work does not constitute a properly developed mathematical model. The fundamental postulates are nowhere stated explicitly; though mathematical symbols and formulas are sprinkled rather freely

44. Rashevsky, op. cit., chap. iii.
45. Lewis F. Richardson, Generalized Foreign Politics (British Journal of Psychology, Monograph Supplement 23 [1939]); and F. W. Lanchester, Aircraft in Warfare (1916), chaps. v and vi.
46. Rashevsky, op. cit., chaps. v, vi, xix, xx.
47. George K. Zipf, Human Behavior and the Principle of Least Effort (1949).
through a long work, the derivations involved are chiefly figures of speech and analogies, rather than true mathematical deduction; in some cases, they are simply wrong. Thus, as an attempt at a systematic social theory, Zipf’s work can only be regarded as a failure.

However, two empirical regularities do emerge which are highly suggestive and may prove promising for further research. It is held, with considerable empirical support, that if the cities of a nation are ranked in order of population, then the rank of a city is inversely proportional to its population; e.g., the largest city, with rank 1, has twice the population of the next largest, which has rank 2. This relation, it is suggested, is an equilibrium condition in that a nation in which it does not hold will either expand or break up until the relation does obtain. Some, though inconclusive, empirical evidence is adduced for this last conclusion, which is used to generalize about international relations. Similar relations hold for other types of frequency data, such as the number of stores of each kind.

An attempt is made to analyze a hierarchy of social classes with the aid of analogous rank-frequency relations, it being assumed that each class exploits the one beneath it and is exploited by the one above it. The strength of a class, and therefore its potential for rebellion, is measured by its total income, while the incentive to remain in statu quo is proportional to the income of an individual. To have equilibrium, the two magnitudes must be brought into a suitable relation. Dr. Zipf suggests that, if the classes are ranked upward from the bottom, the income of each individual should be proportional to his class rank while the number of individuals in any one class should be inversely proportional to the square of its rank. From this analysis, certain implications are drawn for the distribution of income. The postulates of this theory of social classes are again never clearly stated, beyond some generalizations that each individual would prefer exploiting others to working as a source of income. It would be interesting to define the problem more precisely and then compare the game-theoretical analysis with the suggestions of Dr. Zipf.

Another empirical regularity found is that the interaction between two cities is inversely proportional to the distance between them. This applies particularly to the flow of traffic or information between them. The study of this relation had been begun earlier by John Q. Stewart of Princeton University, an astronomer. Stewart has stressed the formal analogies of this relation to the law of gravitation. Let \( P_i \) be the population of place \( i \), and \( D_i \) be the distance of place \( i \) from a given place \( A \); then the demographic potential at \( A \) is defined to be \( \sum \frac{P_i}{D_i} \), the sum being taken over all populated places \( i \). Under the above law, the demographic potential should represent the total amount of transactions per unit population at \( A \) and therefore should correlate with other economic magnitudes. Some evidence has been found to support this assertion, but it can hardly be described as proved.

V. THE USE OF THEORETICAL MODELS IN INDUCTIVE INFERENCE

The familiar proposition in the methodology of science that empirical study without previous theoretical development is fruitless has recently been developed in detail with particular reference to the field of economic analysis by a group of individuals associated with the Cowles Commission for Research in Economics, at the University of Chicago. The general viewpoint was expressed at length by Trygve Haavelmo. The associated problems in mathematical statistics have been studied by Henry B. Mann and A. Wald, Tjalling C. Koopmans, Herman Rubin, and Roy B. Leipnik; and Theodore W. Anderson and H. Rubin. The most elementary exposition is that of J. Marschak; slightly more technical are those of Koopmans, Marschak, and Gershon.


49. There must be mentioned, if only in a footnote, certain other recent works relevant to the mathematical formulation of social theory. Stuart C. Dodd and Elliot D. Chappie have devoted much attention to the proper symbolism of social problems as a basis for the construction of models and for empirical research. See Dodd's Dimensions of Society (1942) and "A Systematics for Sociometry and for All Science," Sociometry, XI (1948), 113-30; and Chappie's Measuring Human Relations: An Introduction to the Study of the Interaction of Individuals (Genetic Psychology Monographs, Vol. XXII, No. 1 [1940]). Their work has been principally methodological rather than substantive. For other work of this type, see the journal Sociometry.

Norbert Wiener, in his Cybernetics (1948), has initiated the systematic study of communication as a feature common to physiology, modern servo-mechanisms, and human society; however, he has made little specific application to social study and, indeed, has expressed pessimism about the possibilities of any sort of social science.


Cooper. Though the work done is strongly influenced by the special problems of economics, there is unquestionably a great portion applicable to the social sciences in general.

1. Exact versus stochastic relations

In economics and in other social sciences, we may certainly expect that no exact relation will hold between the variables we measure for at least two reasons: (1) Not all the variables which are relevant are included in the analysis, and we always omit a host of unimportant factors which are too difficult to measure. (2) The variables we do observe are not measured precisely. In the statement of a relation, then, we must include not only the explicitly enumerated variables but an additional unmeasurable variable, known as a disturbance. Such a relation is said to be stochastic. Thus, we may say that, for an individual, consumption (c) is related to income (y) by a relation of the type c = f(y) + u, where u is a disturbance which stands for all the omitted variables which influence consumption. Merely to say that a relation holds with a disturbance in it is a tautology. For any function f(y), we can make the above relation true by defining u to be c - f(y). To give a relation empirical content, we must assert some regularities in the behavior of the disturbances.

We say that the disturbance in a given relation is a random variable with a probability distribution which is the same for each time the variables are observed. It is usually also assumed that the disturbances at different times are independent. Thus, in our consumption example, if the subscript t denotes time, the disturbances u_t = c_t - f(y_t) for different values of t can be regarded as forming a random sample from a fixed probability distribution. These concepts extend themselves naturally to all types of social laws. In general, then, we may say that the formulation of a generalization in social science is equivalent to an assertion about the probability distributions of certain disturbances.

2. Simultaneous relations and the concept of a structure

If we take the individual viewpoint as developed in the first part of Section II of this chapter, we find that the actions of all individuals are simultaneously determined by the equations of form (1). There is no unique line of causality; the actions of all individuals enter symmetrically. The situation is not altered if we assume that the equations (1) are stochastic. The variables P_i, on the other hand, play a different role. If they refer to exogenous factors, they will in general be determined by processes which are independent of the social context in which equations (1) are stochastic.

With respect to actions A_1, ..., A_n, the exogenous factors have a strictly unidirectional causal influence. The same is true, with some qualifications, if the variables P_1, ..., P_n contain past actions of individuals as well as truly exogenous factors, for at any given instant of time, the past has also a unidirectional causal influence on the present choice of actions. The causal variables P_i are referred to as predetermined; the mutually interdependent variables A_i are referred to as endogenous or jointly dependent. At any given instant of time, therefore, the endogenous variables are simultaneously determined by a set of stochastic equations,

\[ A_i = f_i(P_1, ..., P_n, A_{i-1}, A_n, u_i, ..., u_n). \]

Here again, equations (11) are tautologous unless the distribution of the disturbances u_i is specified. The equations and the distribution together are known as the structure of the social system. The predetermined character of the P_i's is expressed in the assertion that they are statistically independent of the u_i's.

Suppose the system (11) to be solved for the endogenous variables:

\[ A_i = g_i(P_1, ..., P_n, u_i, ..., u_n). \]

Equations (12) are known as the reduced form of the structure. The reduced form is itself a structure, and it might seem at first that a knowledge of it conveys as much information about the endogenous variables as does (11). This is not so if we realize that one or more of the equations (11) may change; for example, one of the equations may refer to a government policy, or it may refer to an industry whose behavior will be affected by technological change. If one of equations (11) changes, then all of equations (11) will be altered. The equations (11) have a higher degree of autonomy than (12), in the sense that each equation of (11) is invariant under a wider class of structural changes than the equations of (12).

If the laws under study were truly immutable, autonomy would be an irrelevant concept, and the
reduced form of a structure would be as useful a description of reality as any other form. The need for an autonomous structure is especially critical when the results of social analysis are to be used for policy purposes. Here, the question is that of choosing among several possible forms for the equation expressing government behavior. What we wish to do is to consider the structure for each possible form of the government equation and predict the behavior of the endogenous variables under each; that form is then chosen which yields the most satisfactory behavior. This procedure can be carried out, however, only if the equations other than the government equation are invariant with respect to changes in it. For prediction under changed structure, and in particular for policy decisions, it is therefore important that the structure be expressed in as autonomous a form as possible. The stress on models based on individual behavior arises from the argument that they are more apt to be autonomous than collective models.


Suppose now that we have a sample of observations on a system of endogenous and predetermined variables and we are faced with the problem of inferring the structure which generated them. This is a problem in statistical inference. Here, the states of nature are the different possible structures. As in the general formulation of the problem given earlier (see the second part of Sec. II and the sixth part of Sec. III, above), we assume it known that the possible structures belong to a certain class. That is certain statements can be made about the structure on a priori basis, so that we need not consider structures not compatible with those assertions. The class of admissible structures is known as the model. The determination of the optimum statistical methods for selecting one structure out of the model is then a technical problem which has been solved, at least in an approximate sense.60

There is a certain difficulty in the statistical analysis which does not seem to arise in the natural sciences. Take first a very simple nonstochastic problem. Assume that prices are determined by the equating of supply and demand, where each is a function of price. If, during the period of observation, neither supply nor demand shifted, the price and the quantity exchanged would not alter, and we would have only one point from which to infer the supply and demand curves. This is obviously impossible, since there are an infinite number of possible pairs of curves passing through the observed point. This same problem arises in stochastic structures; consider again the supply-and-demand problem:

\[ q = a_1 p + a_2 + u \]  
\[ q + b_1 p + b_2 + v \]  

where \( q \) is quantity exchanged, \( p \) is price, and \( u \) and \( v \) are disturbances with a given probability distribution. Equations (13) and (14) can be solved for \( q \) and \( p \) in terms of \( u \) and \( v \), and we can then find the distribution of \( q \) and \( p \) from that of \( u \) and \( v \). For example, suppose that \( a_1 = -1, a_2 = 2, b_1 = 1, b_2 = 1, \) and \( u \) and \( v \) are normally and independently distributed with means zero and variances 1. Then it can be shown that \( p \) and \( q \) are normally and independently distributed with means 1/2 and 3/2 respectively, and variances 1/2. Now suppose that \( a_1 = -1/2, a_2 = 7/4, b_1 = 2, b_2 = 1/2, \) and \( u \) and \( v \) are normally and independently distributed with means zero and variances 5/8 and 5/2, respectively. Then \( q \) and \( p \) have the same distribution as in the previous case. Since observations on a sample of \( p \) and \( q \) values can only yield information as to the distribution of \( p \) and \( q \), it is clear that no such sample, no matter how large, could enable a choice to be made between the two structures. A model containing both structures is said to be unidentified; in general, a model in which no two structures generate the same probability distribution of the endogenous variables is said to be identified. In order to make useful statistical inferences, the range of possible structures must be sufficiently restricted by a priori considerations so that the model is identified.

The conditions for identification have been discussed for models in which the structures are systems of linear equations. The concept of identification has also been generalized in various ways.61


4. Model-building and scientific tactics

The method of scientific investigation indicated in the preceding paragraphs calls then for intensive a priori thinking to formulate a model, followed by the selection of a best-fitting structure from that model by appropriate statistical techniques. It is the virtue of the Cowles Commission approach to have set forth this process clearly and to have resolved many of the statistical difficulties in the way of its fulfillment. It is clear that the crucial step is the choice of a model. If we can say very little on purely a priori grounds about the nature of the process under investigation, then the resulting model is unidentified, and further progress is stopped.

An alternative procedure employed by a number of economists is to start with a very wide and vaguely stated model and investigate empirical data which seem to be relevant. By examination of these data, more definite models will be suggested which will, in turn, provide the basis for further empirical research, and so forth. It might be asked how, if the original vague model is really unidentified, this procedure can lead to any results. Two related answers seem to be implied in the discussion: (1) The observations will select out of the original model a collection of structures compatible with the observations; among these, the "simplest" in some sense is selected. (2) There is really an identified model in the minds of the investigators, which, however, has not been expressed formally because of the limitations of mathematics or of language in general. Perhaps we may interpret the "simple" structures as those found in this unconsciously maintained model. The choice between the alternative scientific tactics indicated depends on the stage of formalization of the underlying theory. No dogmatism is possible; a certain amount of oversimplification is tolerable (and necessary) to gain the advantages of formalization and the use of optimum statistical methods, but there are limits; since the statistical methods are best only on the assumption that the model is correct, a serious error in formulating the model may invalidate all further empirical work based on it.

62. See, for example, the works named in footnote 11.
INTRODUCTION

In the last two decades we have witnessed the emergence of the "system" as a key concept in scientific research. Systems, of course, have been studied for centuries, but something new has been added. Until recently scientists and engineers tended to treat systems as complexes whose output could be expressed as a simple function of the outputs of the component parts. As a consequence, systems were designed from the inside out. Increasingly researchers have come to deal with systems whose output cannot be expressed as a simple function of component outputs and it has become more productive to treat them holistically and to design them from the outside in.

The tendency to study systems as an entity rather than as a conglomeration of parts is consistent with the tendency in contemporary science no longer to isolate phenomena in narrowly confined contexts, but rather to open interactions for examination and to examine larger and larger slices of nature.

Under the banner of systems research (and its many synonyms) we have also witnessed a convergence of many more specialized contemporary scientific developments. Where those interested in systems research gather we are likely to find representatives of every scientific discipline who have become expert in such diverse fields as communication theory, cybernetics, computers, decision theory, value theory, the theory of games, operational gaming, and organizational theory. These research pursuits and many others are being interwoven into a cooperative research effort involving an everwidening spectrum of scientific and engineering disciplines. We are participating in what is probably the most comprehensive effort to attain a synthesis of scientific knowledge yet made.

As in all boundary-breaking activities—in or out of science—enthusiasm for new vistas and the joy of playing in new areas tends to bring with it certain excesses. There is already evidence that, in the systems revolution, modest results tend to be excessively generalized and that assumptions once stated—if stated at all—tend to be ignored in defining the realm to which results can be applied. Furthermore, and perhaps more seriously, there is a tendency for more and more research time to be consumed in efforts to answer questions which have no operational significance. That is, a new type of metaphysics, one which is subtly cloaked in mathematics, is arising and drawing scientists into fruitless inquiries.

It is my contention that there is a fundamental misconception in game theory which has had a deleterious effect on both theoretical and experimental work done on individual and collective decision making. This misconception derives from the failure to distinguish what might be called an "exercise" and a "problem." It is manifested by the assumption that games are adequate models of at least some problem situations. It leads to a multiplicity of decision criteria under so-called "uncertainty" conditions and to an inability to determine which criterion is "best." This, in turn, has led many scientists to play games while suffering under the illusion that they are conducting theoretical or experimental inquiries.

These assertions require clarification and justification. It is to these tasks that this essay is directed.

GAMES, EXERCISES, AND PROBLEMS

A person (or group) can be said to be in a problem situation if the following conditions exist:

1. He has one or more unsatisfied desires; that is, he wants something that he does not have, and
2. He has available alternative ways of pursuing the objective(s) and these alternatives are not equally effective; that is, he has a real choice.

He can be said to have a problem if he is in a problem situation and is in doubt about the relative effectiveness of the alternatives; that is, he must have an uncertain state of mind.

It will be observed that in a game, as conceived in Game Theory, the first two conditions are satisfied. Utilities or valued outcomes are involved, as are unequally effective plays or strategies. To have a game, however, several additional conditions are required:

1. The possible plays can be specified in advance.
2. There is a set of well-defined end-states.

453
A specified pay-off is associated with each end-state.

It has frequently been pointed out that Game Theory cannot be applied in many real situations because either (1) one or more of the three conditions just specified cannot be satisfied, or (2) the mathematical solutions which are available cover only relatively simple competitive situations. There have also been a number of more subtle and more fundamental criticisms leveled at the use of game-theoretic games as models of decision making. One of these observes that even if each of a pair of opponents can assign a measurement of the utility (to them) of each outcome, they may not impute the utilities to each other which are the same as their self-assigned utilities. That is, two opponents may formulate the same problem situation into two different payoff matrices. A second criticism is based on the observation that in many problem situations in which a mixed strategy can be selected the payoffs associated with any one strategy may depend on the frequency of choice of the other; that is, the payoffs may not be independent as is assumed in Game Theory.

The inadequacy of Game Theory to which I want to address this discussion, however, is different from these. It is, however, at least as fundamental as these, but, in general, has been overlooked. It is found in an examination of the assumption of complete ignorance of the probabilities of the competitor's choices.

The importance of this assumption has already been widely observed. For example, referring to the criteria of choice employed in Game Theory, Luce and Raiffa (1957) remark:

A common criticism of such criteria as the maximin utility, minimax regret, Hurwicz $\alpha$, and that based on the principle of insufficient reason is that they are rationalized on some notion of complete ignorance. In practice, however, the decision maker usually has some partial information concerning the true state. No matter how vague it is, he may not wish to endorse any characterization of complete ignorance, and so the heart is cut out of criteria based on this notion. (p. 299)

The assumption of complete ignorance can, perhaps, be defended at least as a limiting case if one accepts the game, as formulated in Game Theory, as an adequate model of a problem situation. I shall argue that such a formulation is not an adequate model of a problem situation and that what must be added removes whatever justification the assumption of complete ignorance might have.

In a game the problem is already formulated for the decision maker and hence he is deprived of the information which is required to formulate a real problem in this form. In a real problem-solving situation the decision maker is not given a game to play, he must extract it out of the situation itself. A pre-formulated problem is a contrived exercise in which the decision maker is presented with incomplete information. The procedure by which he does or should solve this exercise is not necessarily the same as what he does or should do in a real problem situation that can be formulated as a game.

This can be seen by reference to a hypothetical application (there is practically no other kind) of game theory to an advertising campaign which was made by Shubik (1955, p. 47):

Two firms, A and B, each have a million dollars to spend on advertising their products in a certain market area. They can use the media of radio, television, newspapers, magazines, and billboards. For simplicity, we will group these five alternatives into radio, television, and printed media. The marketing research sections of each firm work out the expected effect of any contingency. We will discuss the decision-making at firm A only. A payoff matrix of $4 \times 4$ is drawn up. This contains information on the $16$ contingencies that might arise if either firm spent all its money advertising solely by means of radio, television, or printed media, or decided to save the million dollars and not advertise at all. Each entry in the payoff matrix represents the amount of extra revenue above cost estimated under these circumstances (in millions of dollars).

First we observe that there are many possible advertising media which are not included in this game. For example, the firms could advertise by use of skywriting, point-of-sale displays, travelling exhibitions, and so on. In practice these could be excluded for only one reason: their choice by competition is extremely unlikely. To do so, then, presupposes some knowledge of the probabilities of competitive choices. Now one might reply that this is doing injustice to an example, that the alternatives might be formulated so as to be exclusive and exhaustive. For example, one might formulate the alternatives as:

(a) visual but not auditory
(b) auditory but not visual
(c) visual and auditory
(d) neither
Now it will be observed that there are different types of advertising that can occur in any of these categories. For example, "visual but not auditory" includes newspapers, magazines, billboards, skywriting, and so on. To associate a payoff with such an alternative, then, presuppose some knowledge of the probabilities of choice associated with these sub-alternatives. No matter how fine the classification, this problem remains.

Even in Shubik's original classification many sub-alternatives are possible in each category. For example, "radio" may mean spot announcements or sponsored programs, day or night programs, and many other variations. To associate a payoff with "radio," then, also assumes some knowledge of the distribution of probabilities associated with these sub-alternatives.

Secondly, it is assumed that Firm A knows the amount of resources Firm B has available. On practical grounds, it is difficult for me to conceive of a situation in which this could happen and in which A would have no knowledge of B's probability of choices. In practice A would have knowledge of B's previous advertising choices and could use this information to estimate probabilities of choice. If B were a completely new competitor with no past history, information would be obtained about the personnel involved in B, and from their past histories inferences would be drawn about possible future actions and their likelihood.

Now one might argue that the advertising example is a poor one and that other competitive situations are not susceptible to these observations. Then let us consider the type of situation which, in my opinion, most closely approximates a game.

In the development of a new weapon and policies for its use we do not know who the enemy will be, where the weapon will be used, when, and under what circumstances. Here we seem to be in a state of complete ignorance. In military projects of this type in which I have participated and with which I am familiar there are two fairly standard practices. First, a countermeasures group is established which tries to develop instruments and tactics to combat the new development. The enemy is assumed to be likely to make a similar study. As a consequence, on the basis of the counter-measure studies estimates are made as to what the enemy is likely to do. Secondly, the weapon and its use is usually tried in "map problems" which are simulated military engagements. These are designed involving the most likely enemies, the most likely locations, times, and so on. For example, Britain or Canada is seldom, if ever, used as the hypothetical enemy in such problems. Once a likely enemy is designated extensive studies of their present and future capabilities are used to estimate the likelihood that they will employ certain specified tactics.

Even if one agrees with all this he might still argue that there are situations in which complete ignorance exists, in situations adequately modelled by what might be called "statistical games." One familiar formulation of such a game is as follows:

You have a bowl which is filled with black and white balls and a method of random selection of these balls. On n draws you have selected m black balls and (n-m) white balls. What is the probability that the next ball drawn will be white (or black)?

This exercise fails to model reality for exactly the same reasons as the competitive game does: it fails to provide the decision maker with the information required to formulate the problem. Specifically, it does not tell him how it was determined that there are only black and white balls or even only balls in the bowl, and how it is known that the draws are made at random. Consideration of how these assertions could be made in practice reveals that information would be required which could be used to formulate a reasonable assumption concerning the distribution of colors of the balls in the bowl.

A number of my colleagues have objected to the argument I have just given with the observation that in many "real life" situations certain information cannot be gotten directly but can only be inferred. My argument, however, is not based on the insistance of access to direct observation involving no inferences. To the contrary, in other places—for example, (Churchman and Ackoff, 1950)—I have maintained that all data are the result of inferential processes. My argument is based on the observation that in such statistical games as I have described the information used to infer the characteristics of the exercise must be made available to him who must solve it before the game can become an adequate model of a problem situation.

There is a serious type of statistical problem, a real one, in which the statistician is conceived of as playing a game with nature under conditions of complete ignorance. In testing a hypothesis, say, of the form: \( H_0 : X = a \), there are two types of errors which can be made.

Type I: rejecting the hypothesis when it is true
Type II: accepting the hypothesis when it is false.

Whereas the first type of error has a unique value—the significance level at which the test is run—the second can only be represented as a
function. This function expresses the probability of accepting \( X = a \) for possible true values of \( X \), \( a + \Delta a \), where \( \Delta a \) may take on negative as well as positive values. For most statistical tests it is possible to derive a "power function" which allows us to determine these probabilities which are usually shown graphically in what are called "Operating Characteristic (OC) Curves." In order to evaluate a statistical testing procedure we would ideally want to determine the expected cost of the type II error. But in order to determine this expected cost it is necessary to know the probability associated with each possible true value of \( X \); that is, the probability that \( X \) is actually equal to \( a + \Delta a \).

This situation is frequently formulated as a game between the statistician and nature in which nature chooses the true value of \( \Delta a \) and in which the statistician has no knowledge of the probability of nature's choices. Having formulated the problem in this way the same decision criteria employed in Game Theory—ones assuming complete ignorance—are then applied to selecting a testing procedure. But again, in any real problem situation involving such a test, the information used to formulate the hypothesis and to establish its relevance to the problem at hand provides some basis for some expectations concerning the true value of the parameter in question. For example, in testing a hypothesis concerning the length of, say, a table, the fact that the table is in a room whose size is known at least approximately, that we can walk around it, and so on, enables us to know the most likely range of possible true values of the length. Even crude information to this effect can be used by the introduction of weighting functions to evaluate a testing procedure. To take an extreme instance, we may assume that the observations on the table's length are normally distributed. Under this assumption we would derive Operating Characteristic Curves which would provide probabilities for possible negative lengths of the table as well as positive ones. We certainly would assume that the probability of a negative value being true is equal to zero.

It is becoming common practice to present a subject with a game-theoretically formulated problem situation in a laboratory for the declared purpose of determining how the individual or group makes decisions. Does a person who has part of his mind tied behind him go about solving problems in the same way as he does without this restriction? That we do not normally think so is apparent from the fact that most teachers of pure and applied science do not believe that exercises done in class are a complete substitute for solving real problems in the field. To be sure, they are not unrelated, but those of us in teaching appreciate the magnitude of the gap between them. We do not know enough at present to make any verified assertions about the differences in decision making in an exercise and problem situation. If we are to continue to use exercises as models of problems, inquiry into the nature of this difference should have a high priority on our research agenda.

To summarize this section briefly, it appears to be a mistake to conceive of problem formulation and the selection of a solution as separable aspects of problem solving. A decision reached and the way it is reached is as much affected by how the problem is formulated as by the criteria employed by the decision maker.

**OPERATIONAL GAMING**

Up to this point we have considered theoretical and experimental work on decision making which involves games that can be formulated in game-theoretic form. Another type of game (operational games) is coming into increasing use, one which is much too complicated (relative to our present state of knowledge) to be given any precise formulation. Military games and business games are generally of this type. The development of such games has become as popular a pastime for many researchers as playing them has for managers of industrial and military operations.

In most instances such games are developed without any clear notion as to how they can or should be used. Rationalizations for the effort are occasionally offered, but in general they have been developed either for the fun of it or for the nonscientific "return" they make possible. Claims have been made for them as useful devices in training, personnel selection, and demonstration. As yet, however, there has been no controlled evaluation of their performance in these regards. It is not with these uses that we are concerned here but with their use as research tools to determine how decisions are made by individuals or groups.

Put another way, our interest here is in the use of these operational games to gather information and develop knowledge about the structure and functioning of organized groups in decision-making situations.

The degree of realism associated with such games has struck experimenters and players alike. Both realize that these games are not completely realistic but few are concerned with

1. See (Thomas and Deemer, 1957).
2. Thomas and Deemer have also provided an excellent evaluation of their use to determine what the best decision is.
The operational game is also conceived of as some kind of model of a specific (or class of) real problem situations. The game has certain properties which are supposed to correspond with the properties of real problem situations. In the game these properties of reality may be represented by the same property with a transformation of scale. For example, the time dimension may be collapsed or the amounts of real money involved may be decreased. A model which has the same properties as that which it represented may be called an "iconic" model. In the operational game real properties may be represented by other properties. For example, instead of making a product the players may make words. A model which consists of substitute properties is an "analogue." Finally, properties may be represented symbolically as in a mathematical model. It is clear that most operational games are mixtures of iconic models and analogues. It is because of their iconic properties that they seem so realistic.

A model is always a simplification of reality and it is only because it is that it is useful in science. It is important, however, to know the nature and significance of the simplification. It is important because only by use of such knowledge can our ability to infer from the model to reality be established.

In our current practices concerning the use of operational games there is, I believe, a tendency to confuse the use of an analogue model and arguing by analogy. These are not the same thing; we can only draw very weak inferences by analogy, which is what is being done. The inferences being drawn, however, are being given a degree of respect they would deserve only if they had been drawn from manipulation of an adequate analogue or other type of model.

In an analogue the revelant properties of the situation modeled are each represented and they are related to each by the designer of the analogue in a way which represents the structure of the real situation. The behavior of the "real" system is inferred from the behavior of the model. In an analogy one situation which has some properties in common with the situation of interest is used without any effort to show either that all the revelant properties are the same or that these properties have the same structural relationship. Most operational games are not constructed as models of the situation of interest; in fact, in many cases the situation of interest is not clearly designated. They are constructed as analogies not as analogue models.

In the construction of most management games a selection of properties of reality is made with no attempt to determine the importance of, or structural relationship between the aspects of reality which are excluded. That is, the same principles of selection and relating of properties normally employed in model construction are not employed in the development of the game. As a consequence, once the game is constructed it may have only a subset of the revelant properties of reality and these may not be properly related. Can we argue that behavior exhibited under the game conditions is similar to that exhibited under real situations? If we do so argue, we argue by analogy, not by an inferential process involving knowledge or assumptions concerning the significance of the omitted variables and their relationship.

The inadequacies of arguments from analogy are too well known to discuss here. What is needed is the same care in the construction of a game as is usually exercised in the construction of an analytic model. The principal difference between the game and an analytic model lies in the use of decision makers rather than representations of them. This difference should not affect the care with which the game is constructed. The principal advantage of the game lies in situations in which the properties of the decision process cannot be modeled and hence the decision making process itself is included in the model.

Consider, for example, two games, from one of which we would not hesitate to draw inferences, from the other of which the reflective experimenter would hesitate to draw inferences.

In one industrial problem in which the Operations Research Group at Case was involved it was necessary to find the order in which items requiring production should be processed over an assembly line. The setup costs associated with each product depended on which item preceded it over the assembly line. The problem was to minimize the sum of the setup costs. The problem could be represented by a matrix in which the cost of making each product after each other product is shown. The matrix was not symmetrical since the cost of setting up for product A after product B was not necessarily the same as for setting up for product B after product A. It was recognized that this was an n-dimensional traveling salesman problem, where n is the number of products to be produced. No analytic solution to such a model was available. Study of the problem revealed several decision rules which appeared to yield lower costs than one would expect to obtain by using intuition and experience to sequence the production. The rules, however, did not cover every
possible situation. Judgment by the decision maker was still required. The researchers replanned the production of the last three years using the proposed decision rules and compared the resulting costs with those actually incurred. A substantial reduction was obtained. The question remained, however, as to whether such improvements could be obtained by the people who actually made the decisions. A game was set up involving the rescheduling of production over that three year period. The people who actually had scheduled production over that period were taught the decision rules and asked to reschedule production over the period. They did so and the results showed the same improvement obtained by the researchers. On the strength of these results the rules were adopted and subsequently showed a continuing improvement over previous methods.

Few would argue with the inference that the improvement in performance obtained by use of the decision rules in the game were a legitimate basis for forecasting an improvement if they were used in reality. The confidence one has in such an inference derives from the adequacy of the model of the problem situation; its correspondence to reality.

On the other hand most of us would hesitate to draw inferences from the performance of a manager in most management games, such as that developed by the American Management Association (Bellman, et al., 1957) to his performance in a real situation. The problem here is that we do not know precisely what the game is a model of. The experimenter is put in the position of searching for a reality that corresponds to the game, or of looking for a problem that fits a solution. To be sure the situation which appears to be modelled in the AMA game is more complex than the product-sequencing situation described above. This may explain its deficiencies but it does not affect its inadequacy as an inferential base to anything but the game itself.

Most of us are excited about the possibilities of the use of operational games for the study of decision making in complex situations. We cannot, however, allow our enthusiasm to suppress our judgment and the realization that as much care must be taken in the construction of a game as in the construction of any other type of model. That models of very complex situations can be constructed has been demonstrated with increasing frequency. One recent and exciting construction of a model of a firm's decision-making process is reported by Cyert et al (1959). From a model such as this an effective game for use in studying organizational decision making could be constructed.

The critical attitude toward games taken here is not to be interpreted as an attempt to rid science of them. To the contrary, I believe that the use of games offers great possibilities in the study of the individual and group decision process. Further, I believe with Rapoport (1959) that the concepts involved in gaming have opened important new vistas, particularly in the behavioral sciences. But these enthusiasms should not close our eyes to the deficiencies in current theoretical and experimental work with games.

REFERENCES


D.

APPROACHES TO ORGANIZATIONAL DESIGN
In recent years, the administrators and personnel in some of our larger social organizations (such as schools, hospitals, and prisons), have been appealing to social scientists for help concerning the following problems:

1. How can these organizations' contributions to society be specified? More especially, that is, how can such an organization effectively evaluate its performance, when its accomplishments seem intangible when compared with industrial or military organizations?

2. How can such organizations determine what kinds of policies, procedures, facilities, and personnel to employ?

The purposes of this paper are to show that advances in the area of the first problem have implications for a concerted attack on the second, and to describe a research strategy being developed to study these problems in public psychiatric hospitals. It is believed that the model described in this paper might prove useful in studying the function of many large social organizations.

The research project, Medical Audit Plan for Psychiatric Hospitals has been organized to develop a methodology which can be used by the staffs of hospitals in evaluation of performance and in the assessment of policies and resources. The term "medical audit" refers to procedures for obtaining information about hospitals often used as a basis for decision on whether or not a hospital meets specific standards for accreditation. Medical audit data has thus been used both to set up standards and to detect deviations from these standards.

Application of typical medical audit to a psychiatric hospital meets with some difficulty in that medical audits have no specified external criteria for evaluating the performance of an entire hospital. The validity of accreditation standards therefore remains untested. For example, one might feel that a given ratio of physicians per patient is optimal, but unless it can be empirically demonstrated that this ratio is related in some measure to hospital performance we cannot be certain that the ratio actually represents optimum function. The use of this ratio as a standard for evaluation of hospitals is consequently seen to be of dubious value.

What seems to be needed here is a research model which involves: (a) the development of criterion measures for psychiatric hospital performance, (b) the development of measures for hospital resources, programs, and policies, and (c) the determination of possible relationships between these two sets of measures. Data pointing to the existence of such relationships might then be used to guide the hospital administrator in deciding what policies or procedures to adopt in order to achieve a certain kind of hospital performance.

THE RESEARCH MODEL

A "research model," as we use the term, is simply a strategy which helps in the selection and measurement of the system variables and relationships we wish to study. (The term is not necessarily synonymous with the term "explanatory model.") It has proven convenient to structure our research model in terms of General Systems Concepts.

One such strategy may be called the "classical model," and its application to the above problems in a single large social organization might proceed in the following steps:

a) Certain variables of the system would be designated as dependent variables whose measures would be designed to reflect various aspects of an organization's performance or behavior.

b) Specific independent variables would be selected and appropriate measures devised.

c) The independent variables would be systematically manipulated or allowed to vary with time, and their effects on the performance variables noted.

There would, however, be several difficulties in this approach. One, imposing variation on the system's independent variables would be impractical except in the cases of very limited problems. Direct manipulation by an investigator might disturb the system's routine functioning, while variation over time, for many variables, would take an inordinately long period of observation. Another, the results derived from a limited study of one organization may be of such a nature that to generalize the findings to other similar organizations
would be unfeasible. Solutions to some restricted problems of one organization would not necessarily benefit administrators of other similar institutions having their own unique problems and settings.

What we would want, therefore, is a broad-scale approach which would provide data to aid administrative decision-making for all social organizations of a given type. The research model we have been developing, which might be called a "multiple systems" model, has the following features:

a) The initial step in our thinking about the problem involved the conceptualization of a class of social organizations. One might attempt to sort organizations into classes by comparing sizes or internal structures, but our approach was limited to generalized concepts of the goals of organizational functioning. The term "organizations of a given type" then means organizations sharing common purposes or goals. In other words, if we consider some of the behavior of social organizations to be "goal-seeking," then public psychiatric hospitals differ from other organizations in that their goals are different. The first step involves the specifications of these common goals along with the selection of variables whose measurement would reflect the performance of any such organization with respect to these goals. These variables are designated as output variables.

b) Having conceptualized the organization as a system with output, the next step concerned the specification of input variables and system variables. The distinction between these two classes of variables has not been rigorously defined in our thinking because it is difficult to specify the boundaries of the system in regard to input variables. However, it should be pointed out that we are not making use of a black box model since many aspects of the "inner" structure and functioning of the system would seem to be amenable to observation and measurement.

c) Variation in the three classes of variables is observed by measuring these variables in many organizations of the given type over the same time interval. All variables are represented by a single measure considered a sample of that variable's values during the given time interval. Then, using each organization as an experimental unit, input and system measures will be related to output measures by means of statistical correlation methods. In addition, input measures might be related to system measures in order to study the effects of input variation upon changes in the system variables.

This approach, which is an extension of the field-study or correlational model widely used in psychology and sociology, will now be described with more detail in its proposed application to the study of public psychiatric hospitals.

The first step in our strategy was selecting those aspects of a psychiatric hospital's behavior to serve as output variables. One means of selecting output variables for social organization study is to ask, "For what purposes was this system created and for what purposes is it being maintained?" These organizational purposes, or goals, usually represent the conceptualization of organization members, and of the community as to what contributions the system is and should be making to society. We wanted a list of goals that would be an abstraction in that it would not represent exactly the goal-structures of any single psychiatric hospital, but would be a generalization from all such hospitals.

Discussion with mental hospital personnel and a survey of the literature resulted in the following list of public psychiatric hospital goals:

1. Care of patients is defined as encompassing those hospital activities which involve the "daily maintenance" of patients, i.e., keeping them alive, physically healthy, and as comfortable as possible while in the hospital.
2. Protection includes preventing the occurrence of harm or injury to patients, hospital staff, hospital property, or to the surrounding community.
3. Social restoration includes those hospital activities aimed principally at releasing patients into the community so that they are able to remain there with an improved degree of adjustment.
4. Training and education involves the training of professional people and hospital employees as well as the dissemination of information to the community.
5. Research encompasses the scientific study of hospital and mental health problems.

The final goal, for want of a better name, has been termed "administration." It includes those hospital activities aimed at controlling and coordinating the operations of various sub-units of the system, as well as the relationship of the hospital to other organizations of society.

It should be noted that these concepts are neither mutually exclusive nor do they constitute the only list that could be formulated. We wanted a list that represented a parsimonious abstraction of hospital purposes while simultaneous comprehensive.

The comprehensiveness of our list resulted from a nation-wide survey of over a thousand psychiatrists, psychologists, psychiatric nurses, psychiatric social workers, and board members of the National Association of Mental Health. These people were asked to supply their own lists of mental hospital goals. We are able to define our list so that well over 90 per cent of the items received in the survey could be classified reliably under one or another of the goals mentioned above.
It should further be noted that the six goals are not themselves variables, but rather classes of variables. These groups constitute the set of hospital behaviors designating the output variables of our design. The survey aided us in deciding which variables, or hospital behaviors, to include in each goal category.

Having proceeded to the point where we had conceptualized classes of output variables, it was then necessary to examine these classes in order to determine which of the variables in each category seemed amenable to operational definition and thereby to measurement. We examined those hospital behaviors subsumed in order to see which of them appeared promising in terms of feasible measurement procedures. The measures of the variables thus selected were to represent estimates of a hospital's performance with respect to care, protection, etc.

In addition to measurability, we have also been guided in our selection of output variables by the notion that we wanted our measures to represent characteristics of "products" of the system's functioning. This criterion is admittedly vague, but it may be illuminating to note that many of our output measures are expressed in terms of things that "happened" to groups of patients after they had been exposed to the hospital environment. In a sense, for some of our output measures, we have been thinking of patients as being "processed" by the hospital.

At this point, it may be enlightening to briefly describe our measures of hospital performance with respect to one of the goals, social restoration. Consideration of hospital behavior subsumed under this goal resulted in our focusing on the "returning patients to the community" concept.

Basically, we have three variables here. The first of these is measured by the probability that a patient, admitted to hospital A in time interval $T_1$, will be released to the community during a subsequent interval $T_2$, following a hospital decision that he is "ready" for release. This is called the release rate of hospital A and it measures, you might say, the social restoration performance of hospital A insofar as quantity of output is concerned.

The second measure is given by the probability that a patient admitted to hospital A in $T_1$, and released from A in interval $T_2$, will return to institutional status (in A or other psychiatric hospital, prison, home for aged, etc.) during subsequent interval $T_3$. This is the return rate of Hospital A and, continuing the analogy, represents a first approximation for the measurement of the quality of Hospital A's social restoration output.

The third measure represents the probability that a patient who was admitted to Hospital A in $T_1$, released in $T_2$, and remained in the community throughout $T_3$, will show a prehospital-posthospital improvement in adjusting to the community's expectations. If one considers the helping of patients to improve or improve their status as socially-acceptable community residents to be the essence of social restoration, then this measure is seen to be critical for representing a hospital's Social Restoration performance.

Procedures for collecting the necessary information for these measures have been worked out in some detail, but I will not attempt to describe them in this paper. The reader who is interested in a detailed account of these measures and of the data-collection procedures underlying them is referred to another publication by the author.

We have no guarantee, of course, that our output measures will prove workable, for we have no generally accepted and reliable measures. We are working in new territory. All that can be done at this stage is to devise measures which appear relevant to the general goal definitions, free of contradictions and opportunity for bias and distortion, sensitive enough to detect inter-hospital differences, and which involve information, comparable from hospital to hospital, we can reasonably hope to gather.

The Input and System Variables

Turning to a discussion of the input and system variables, we are faced again with the task of selecting which variables to include in the design. There is no solid notion of which input and system variables have relevance for any given output variable simply because there has been very little research in this area. Selection has been guided by choosing those variables which seem to embody issues and problems of relevance to hospital administrators and mental health personnel. A few have been suggested in consulting the literature in the fields of administrative science, social psychology, and psychiatry.

With respect to the study of hospital output in the area of social restoration the following types of input and system variables are included in our design:

a) Input variables. By input is usually meant "those parameters of the system's environment whose variability affects the system's functioning." This definition includes the notion of something entering the system or impinging upon its boundaries. Although it is difficult to decide, in the case of social organizations, what is system and what is environment, we have achieved at least a preliminary specification of the input variables.

One class of input parameters involves specification of the type of community in which the

hospital is located. This includes such things as rural vs. urban location, as well as various characteristics of the larger mental health organization (state mental health department) to which the hospital is subordinate. This class of parameters may be looked upon as providing some real constraints to the variation in the remaining classes of input variables.

A second, and very basic, class of input variables involves certain measurable features in the patient population entering the hospital throughout a given time interval. These variables include the size of this population and measures of central tendency with respect to age, socio-economic level, education, and pre-hospital level of social adjustment.

A third class of input parameters comprises the various aspects of direct community participation in the hospital's functioning, such as amount of community volunteer work and amount of professional consultation afforded to the hospital during a given interval of time.

A final class of input variables encompasses basic resources which the hospital receives during a given time period, such as financial allocations, food, and medical supplies, etc.

b) System variables. Herein are included variables which we feel characterize the "internal" structure and functioning of the system. The system variables also fall into various categories.

The first class of system variables contains physical characteristics of the hospital such as measures of size and capacity, measures of treatment, maintenance, and recreation facilities.

The second class of system parameters includes variables reflecting the hospital's staffing pattern, such as size of staff, formal and informal patterns of authority, and attitudes of the staff toward mental illness and mental patients.

A third and final class of system variables might be termed "policies and procedures." The parameters in this class may be said to reflect the hospital's internal functioning rather than its structure. Included here are a great many variables such as characteristics of inter-departmental contact and communication and measures of the "operating characteristics" of the various hospital services and departments. In short, this class contains variables which reflect "what goes on" inside the system.

Having gained a general idea of what the input and system variables were to be, the next task is devising measures for these variables. Here again, one finds quite often there are no "ready-made" measures whose usefulness as operational definitions have been demonstrated.

Some of the variables present little difficulty because their measurements result more or less directly from the level of simple observation, e.g., the total number of buildings, or the total number of physicians employed full-time by the hospital, or the sex of the director of nursing. These are measures of what we will call "low-level" variables.

Constructs and measures of a higher order may be defined by combining "low-level" constructs. For example, one could consider the variable "physical size of hospital" whose measure could be defined in terms of the number of buildings in conjunction with the area of the grounds and the average amount of floor space per building. Or, one might define the construct "degree of dignity with which patients are treated" as a great number of variables of lower order such as, amount of hospital censorship of patient mail, degree of privacy of patient's bathing facilities, and amount of time allotted for patients to eat in hospital dining halls. It becomes obvious that there are many possibilities for constructing higher order variables and each of their measures could be constructed from a great many different combinations of lower order variables.

We have been striving to specify some of these variables to have as high an order as possible and yet be meaningful in the light of our knowledge of hospital operations. One reason we are doing this is to reduce the number of variables considered. This a priori construction of parameters has its pitfalls, however. If we have measures x, y, and z, and combine these to derive higher order measure A, which we then relate to a given output measure, we may fail to demonstrate a relationship simply because we specified an "incorrect functional relationship between A and variables x, y, and z, or because variable w was omitted from the construction of A. All is not lost, however, because we will, after data collection, have measurements for w, x, y, and z, and we are free to try other functional relationships which might be suggested by the data. It is anticipated that there will be a considerable amount of empirical construction in higher order variables.

It is possible, of course, to construct higher order variables by using the empirical method only. A factor or principal components analysis of the inter-correlations of all input or system variables might seem to be indicated here so that "clusters" of variables could be elicited from the data and used as the basis for higher order parameters. This would represent an attempt to discover the basic dimensions underlying the measures.

At the present time, however, it is felt that the rational, a priori approach should definitely accompany the empirical in the construction of higher order input and system measures. Our aim is to provide the staff of a mental hospital with some tools for manipulating very real and palpable features of the hospital so as to enhance a specific output. This aim has guided us in the construction of these measures. Should a higher order construct appear to be of little predictive value, one of its constituent lower level measures may
still be quite useful in being highly correlated with a given output.

The collection of data relevant to the input and system measures will consist largely of structured interviews and systematic observations on visits to public psychiatric hospitals throughout the country. At present, we anticipate data collection relevant to several hundred such variables. This "broad" approach is pragmatically necessary if entire hospitals are to be used as the experimental units. It is much more efficient to gather a "broad" approach is pragmatically necessary if entire hospitals are to be used as the experimental units. It is much more efficient to gather a great deal of information at one time from a hospital than to return again and again for additional data.

Input-Output Relationships

Our research design is a correlational one, in which input and system measures (predictors) are to be statistically correlated with output measures (criterion measures), both sets of measures resulting from the observation of many hospitals. For each hospital we will have several hundred measures of input and system variables and an anticipated two or three dozen measures of output. Our task will then consist of trying to detect relationships between the two sets of variables, input and system variables on the one hand and output variables on the other. Many problems confront us here, some of which have already been suggested, and I can do no more than to outline our thinking as to how we should proceed.

First, some of the input and system measures may fail to reflect any inter-hospital variation. These variables will then be dropped from further consideration because, insofar as our measuring techniques could detect, they would not be variables at all. This points up a difficulty in our "multiple systems" approach. The lack of variation displayed by an input or system variable could result from the "single-value" nature of our measure. It would not be legitimate to infer, in this case, that such a variable has no relationship to a given output. We can only report that the variable's lack of inter-hospital variance precludes any inferences about its relationships to output and that its further consideration should be shelved, but not discarded altogether.

Our next step will be working with each class of output variables, one at a time. Each input and system variable will be related to each of the output measures in the given class. As a first approximation, this would be accomplished by the use of product-moment correlation. (Should a higher order variable, of a priori construction, fail to be correlated with any output, its lower level constituents would then be tried.)

For any given input-output or system-output pair we would then begin to "partial out" other variables one at a time in order to gain a truer picture of the relationship. This "partiallying-out" process will be limited by pragmatic considerations, and guided by the realization that psychiatric hospital people have some specific questions and problems about which we desire to supply some information. Should the use of multivariate statistical methods, such as principal components analysis, point up some underlying dimensions of the input of system measures, these too might be exploited by relating them to output measures.

I might remark that input-output and system-output relationships are not the only ones that might be considered. Relationships between output variables might be of vital importance to administrators if high performance with respect to one goal, say protection, tended to be accompanied by low (or high) performance in another goal area, say social restoration. Should such relationships be found it could prove helpful to the management of hospitals in deciding which objectives should receive emphasis in the planning of long-range policies.

Another aspect of our model is that it lends itself to the determination of input-system relationships. It would be of particular interest, not only to a hospital administrator, but also the social scientist, to learn how changes in input affect the structure and functioning of the system.

Our research strategy will not enable us to derive a determinate system of variables; causal relationships will not be deduced. For this, the "classical" strategy, involving more restricted sets of variables, would be necessary. Data from our research, pointing up the possibility of useful relationships, would be a guide in such efforts.

We realize, of course, that we are dealing with extremely complex systems of variables. However, it is felt that this "broad-scale" research strategy is a necessary first step in establishing an empirical base for further investigation in the area of research in organization theory and in psychopathology. Even if all of the data is not put to immediate use, it is felt that this broad approach will provide mental hospital personnel with some new ways of looking at the operations of their hospitals. Our preliminary work in several state mental hospitals has already resulted in a stimulating interchange of ideas between our staff and the personnel of these institutions.

A Summary of the Model

The research model described in this paper provides an approach to the study of a given class of social organizations, especially where it is
desirable to evaluate the effectiveness of the organization's functioning and thereby determining which aspects of the organization's environment, structure, and "inner functioning" are related to the effectiveness of the organization's performance.

The first step in such an approach requires defining a class of social organizations in terms of a generalized set of organizational goals. These organizations are viewed as systems having input and output.

Output is defined by total system performance with respect to the achievement of organization goals, and indices thereby devised to measure this performance. Input and system variables are defined and measures are designed to reflect these variables. The input variables represent those features of the environment which are seen as impinging upon the system, or serving as constraints to system behavior. The system parameters represent various structural and operating characteristics of the organization.

We then wish to relate the input and system parameters to the output variables by observing the covariation of their measures. Because it may be difficult or impractical to impose variation on the system's variables in any direct fashion our research strategy involves the measurement of these variables in a great many of the organizations, comprising the given class, within the same time interval. Relations between the measures will be ascertained by means of statistical correlation methods.

Input-output and system-output relationships furnish information which can be used by organization members as a basis for policy decisions aimed at achieving a given level of output. Input-system relationships can be used to understand how organizational structure and functioning are affected by changes in environmental conditions. This "multiple-systems" model can be used as a basis for studying and evaluating the effectiveness of many types of social organizations and agencies such as schools, prisons, libraries, and hospitals.

It may be instructive, in closing, to note how the above research strategy differs from typical Operations Research procedures whose aims are very similar.

The first difference can be seen by comparing the scope of the two approaches. Operations Research methodology focuses typically on the functioning of a single organization, the aim being to provide information relating to the optimum performance of the system with regard to the achievement of objectives which are more or less unique to that organization. Our approach, on the other hand, involves the simultaneous study of many organizations of a given class, and measures performance in terms of the achievement of generalized goals which typify the entire class of organizations.

Secondly, Operations Research usually involves the development of an a priori model (that is, the model is developed prior to data-collection) to represent system functioning. This model, and its solutions, are then tested by employing empirical data. Our approach does not involve this type of model-construction, but calls for the collection of data which can then be used as the basis for constructing models of system functioning.
Rene Dubos reports that we have at last discovered the missing link between ape and man—it is us. This observation provides a kind of insight regarding the difficulties we face in planning human settings. If homo sapiens were less highly developed he probably wouldn't have the capacity to plan; if he were more highly developed (a completely rational organism) there would be little reason to plan. In either case man would be in harmony with his environment and his evolutional course. Our failures in planning the course of human events are due in great part to our misunderstanding of man's capacity to participate in and execute such plans. In other words, our plans have not as a rule been appropriately linked to the realities of human functioning. We are well aware that things are not going as we would individually and collectively wish. Our interest in change is thus predicated on an assumption that there are critical problems, and the hope that things could be better. The question which remains is how we should go about changing the present state of affairs in the future.

A human problem can be properly defined as the disparity between things as they are and things as they ought to be. When there is a perceived disparity between human intentions and accomplishments, we are faced with a design problem. Planning and design involves the development of strategies for organizing human settings to overcome extant social, biological, and environmental deficiencies. The design of a human system, then, is concerned with things as they ought to be and with development of the means for moving the system to this new state. Intervention and change—the prediction and control of human affairs—is the implicit objective of planning and design. Such an enterprise cannot, needless to say, be approached casually, but requires the most careful formulation possible. In this paper I will explore, at rather fundamental levels, the methodological future, the control of the planner's dilemma, i.e., the real solution to the planner's dilemma, i.e., the "rational" plan versus the "irrational" interests within the community. Need for the advocacy

In most if not all contemporary cultures, the concept of large-scale planning and design is understood by only a few, and planning by the few for the many is the common tradition. Most members experience dysfunctions and attempt to respond to these as they arise.

A planning culture can be described as one wherein collective planning is an integral aspect of the individual member's thought and behavioral repertoire—an enculturated orientation toward existence. Few cultures have existed without planning, but it is just possible that no planning cultures have emerged. In a planning culture, where collective planning strategies are supported, participated in, and taken for granted by the membership, emphasis is upon directed response to the future. Objectives are identified and resources are mobilized to attain these objectives. A planning culture would seem to emerge when (1) dysfunctions are widespread, (2) technological (and other) resources are available, and (3) the processes of enculturation (education) are available on a large scale. The United States is hopefully becoming a planning culture. The unplanned imposition of technology, the squandering of resources, and the linkage of these to questionable ideologies have produced a level of social and environmental pollution which has become a model of what can go wrong without planning. Certainly, the sophistication of our planning is inadequate considering the complexity of our problems. We in the United States, "blessed" as we are with an abundance of means and a highly developed technology, have everything to learn about planning—organizing ourselves to exist in harmony with our environment and other cultures.

Even in a planning culture, particularly a technologically developed one, there emerges the need for specialists, those with particular skills and specialized knowledge of relevant phenomena. In becoming a planning culture, where guessing is replaced by a controlled approach to the future, there emerge questions about who plans what for whom. Planning to date has occurred in spite of rather than because of widespread support, but in a culture which understands and subscribes to the need for it, planning becomes a powerful force. In United States metropolitan centers, advocacy planning has emerged as a dampening force, a safeguard against the implementation of plans in which the interests of certain populations are not fully represented. As it exists, advocacy is not a real solution to the planner's dilemma, i.e., the "rational" plan versus the "irrational" interests within the community. Need for the advocacy

*This article is based on a paper presented at the 2nd International Conference on the Problems of Modernization in Asia and the Pacific, held at the East-West Center, University of Hawaii, Honolulu, August 1970, the proceedings of which will be published in a volume entitled Modernization in Asia: A Futuristic Perspective. This version is published with the permission of the Center.
movement represents a failure of the planning and design disciplines to date, but it establishes the fundamental principle that formulation of plans for the future must systematically involve the people affected. In a planning culture, this is neither excessively difficult nor productive of bad plans.

The mention of planning (not to mention human systems design) arouses concern regarding freedom. No human wants to be under the control of another, and planning implies control. The following comments could easily be misunderstood in this regard. This analysis is, however, predicated on the belief that any approach which does not seek to maximize the freedom of affected individuals to pursue their own goals is not worthy of support. This proposition, one which we all ostensibly subscribe to, is not so simple. Freedom from what and for what? One man's freedom obviously ends where another's begins, and most problems come about as individuals and collections of them pursue inimical goals. People seek autonomy to pursue their individual goals, but these goals are integrally linked to those of others. Like it or not, all humans are essentially interdependent; our biological and extrabiological survival is conditioned by this reality. The difficulty is that the environment is so arranged that some are in a position to, and are rewarded for, exploiting this interdependence. The task of the human systems designer is to arrange the environment so as to maximize autonomy in the context of collective goals, and minimize the probabilities of exploitation of the many for the benefit of the few. In short, planning and design is not antithetical to human freedom; it is absolutely essential to insure it.

The problems of planning and designing systems for increasingly large and complex settings are, of course, monumental. Traditional problem spaces are by necessity being collapsed and re-defined in complicated ways, requiring more complicated approaches to problem solving. The necessity for synthesizing discipline and subdiscipline objectives is manifest. However, we as yet have no conceptual framework, no metastructure of operationalizing a truly transdisciplinary approach to human problems. Specialization of one sort or another has been essential to the development of scientific knowledge. One consequence of this is conflicting epistemologies, conflicting research and problem-solving paradigms, disparate levels of addressing problems, and closed systems of analysis. These make the required interdisciplinary synthesis extremely difficult.

This analysis attempts to outline an a-disciplinary framework for addressing human problems, for identifying relevant resources, and areas of research which may help to solve them. The approach is not without bias toward what constitutes a useful view of reality and scientific relevance. It isn't assumed for a moment that there is universal agreement on these matters, and the recommendations may be judged by some social scientists as unfeasible, considering our present knowledge base—or just downright objectionable on other grounds (e.g., excessively mechanistic). However, our performance to date in organizing human affairs is, to say the very least, inadequate. When facing a precarious present and a frightfully uncertain future, some rather bold suggestions seem in order.

The following formulations are extrapolated for the most part from earlier work dealing with physical design issues, but with certain adjustments the concepts appear to apply more generally to a discussion of "software" for social change. This discussion involves neither predictions of nor prescriptions for the future. Rather it attempts to outline a means of getting to any number of futures from the present.

An Approach to the Design and Maintenance of Human Systems

One cannot long discuss the analysis or design of human settings without extensive reference to the environment. Perhaps this is a good place to enter the discussion. The term "environment" is a construct employed for conceptual convenience to study the effects of one system upon another. The distinction between man and environment is but one such case. There is not one human environment, but many; it can be partitioned into infinite sub-environments depending upon the objectives of the analysis in question. It is entirely reasonable to isolate and speak of the economic, political, or natural environments, or to speak more generically of a problem situation as composed of internal, i.e., the system under analysis, and external, i.e., the system impinging upon the system under analysis, environments (see below).

In approaching the planning and design activity, a traditional dilemma has emerged in distinguishing between "natural" and "a-natural" phenomena. That which is natural is generally considered desirable; that which is a-natural or artificial is generally considered undesirable. But most humans exist in man-made environments. One would be hard-pressed to identify a physical environment completely devoid of influence or alteration by man, and who can identify a social structure which is more "natural" than another. The search for immutable natural laws has been the hallmark of science. In adopting paradigms from the most mature and successful physical sciences, aspects of the social and behavioral sciences have also sought to discover "natural" human states. On the other hand, we have the undeniable fact that environments which facilitate human ontogeny are subject to a degree of "arbitrariness" associated with the man-made. This is true unless one wishes to take the position that anything man does is "natural"—a tautology of questionable utility, in which case we indeed have no solvable problems.
Since most mammals are not presently living in the environment which selected their principal genetic or behavioral features, attempts to return human systems to their "natural" states would seem a rather fruitless undertaking. We may envy the social structure of the ant which has stabilized in a mere 50 million years, but analogous "natural" processes in humans, if indeed such ever did exist, have been altered to the point where we can no longer seek such mythical states. Whether to intervene is no longer a live issue. If we do nothing there are risks; if we do something wrong there are comparable risks. Human systems, including the environments which support them, must be designed: we must intervene more, not less. This is a formidable proposition, but we really have little choice in the matter. The first order of business is to survive biologically, but we must go beyond survival in the biological sense; and it is this sense of survival and well-being which presents the more interesting and complex class of problems.

This matter of the natural vs. the a-natural has been belabored (to excess, perhaps) in order to point up an extremely useful and quite necessary orientation in approaching the design and management of future human systems. In his recent book Sciences of the Artificial (1969), Herbert Simon attempts to resolve the dilemma of natural vs. a-natural phenomena. I will utilize some of his arguments, together with some of my own, in order to document an approach to the subject at hand.

The realization of artificial systems involves synthesis or design; the concern is with "how things ought to be in order to attain goals and to function" (Simon 1969). The design of artificial systems, then, involves normative as well as descriptive issues. Simon questions whether we should maintain the traditional exclusion of normative (as opposed to descriptions of how things are) when we move from natural to artificial phenomena—from analysis to synthesis. He identifies the following characteristics of artificial systems: (1) artificial things are synthesized by man (and usually prespecified); (2) artificial things may imitate appearances in natural things while lacking in one or many respects the reality of the latter; (3) artificial things can be characterized in terms of functions, goals, adaptation; (4) artificial things are discussed in terms of imperatives as well as descriptives.

The artificial systems design problem involves the following functional elements (see Figures 1 and 2): (1) the external environment, (2) the internal environment, (3) the goals of the system to be realized, and (4) the transducers relating external and internal environment. The external environment is that within which the system under analysis is to function. It operates according to certain natural laws and impinges upon the system in question in systematic ways. The internal or designed environment (Studer 1967) is that which is to be organized in order to attain the particular goals the system is intended to attain. Both the internal and external environments operate within the same natural laws. For example, realization of a bridge or man-made shelter requires the design of a structural system. This structural system (an aspect of the internal environment) is subject to the same gravitational forces as trees or other physical elements in the external environment, and must be analyzed with the same tools. Likewise, the same biological or psychophysical laws apply equally to persons included in a system to be designed as to those not included.

The conditions for goal attainment in artificial systems are determined by the external environment. If properly designed, the internal environment will be adapted to the external in that...
its behavior will be largely determined by the behavior of the latter, i.e., "the behavior takes the shape of the task environment" (Simon 1969). The interface between internal and external environments can be mediated by transducers which relate these two systems in desired ways, i.e., to meet the goals. It should be carefully noted that, once an artificial system is realized, it becomes a part of and alters the external system which conditioned its design. This is a consideration which leads me to take mild exception to this particular aspect of Simon's formulation. It appears that an internal environment might, depending on the scale involved, exert an influence on the external in a somewhat less subordinate relationship than that depicted by Simon. In any event,

Description of an artifice in terms of its organization and functioning—its interface between inner and outer environments—is a major objective of invention and design activity (Simon 1969).

Assuming that I have not misrepresented Simon's analysis in order to support one of my own, I am encouraged to go on now to assert that human systems must be developed within precisely these same principles. We are and have been for some time faced with the organization of artificial human systems. It is essential that we more fully recognize the artificial nature of our contemporary existence. We simply cannot assume that natural forces will self-correct social, biological and environmental dysfunctions without the demise of large numbers of people.

Whether engaged in analysis or synthesis, a fundamental concern is selection of appropriate units of analysis. These elements direct empirical observations and organization of the data. Similarly, the problem-solving task involves particular units (independent and dependent variables). Moreover, the problem-solver's selected unit of analysis reflects a philosophy of science concerning the phenomena of problematic concern (T. Kuhn 1962).

As a unit of analysis for human problems, "need" has received popular recognition. However, this unit lacks certain characteristics which are necessary to either the analysis or synthesis of human events. "Need" can only be inferred from or operationally defined in terms of units of behavior. Behavior¹ is a more likely candidate for problem formulation, an index of both biological and social need. In short, human problems are behavioral problems requiring behavioral solutions. If we formulate problems in these dimensions, human systems design becomes a more reasonable proposition. Culture is, after all, simply a shorthand way of talking about behavior (Hernstein 1966).

The objective of human systems design is to realize an alternative state of (behavioral) affairs, to organize energy and matter (living and non-living, human and non-human) within the internal environment bounded by the problem situation under analysis. In directing attention to the goals of the population and maintaining a commitment to behavior as the class of independent variables, we are in a position to identify an approach, i.e., a behavior-contingent paradigm, for the design and maintenance of human systems.

The above series of arguments, it is claimed, condition a program of human systems design. These can be summarized as follows: Human systems planning and design is necessary and conditioned by the fact that humans do not, by virtue of their phylogeny and/or ontogeny, behave with unfailing "rationality"—have not demonstrated the capacity to organize and regulate their affairs without intervention and self-conscious planning. The human environment has evolved via extensive "unselfconscious" human influence in such a way that widespread disparities between human intentions and accomplishments are the inevitable products. The human environment is for the most part artificial, not shaped entirely by "natural" processes but, for better or worse, configured by man. We often observe human groups "muddling through," deluded by the belief that their dysfunctions are the products of inevitable "natural" processes. Aspects of the behavioral and social sciences reinforce this view. The intensity of change, interaction, and the attendant complexities occurring in most cultures demand a more systematic approach to the design of human systems, including the social, biological, and environmental domains. The emergence of planning cultures is both necessary and inevitable. Human systems of interest to this analysis are those classified as artificial, i.e., designed in response to the goals of population under analysis. Human behavior, an empirically accessible change in organismic state, is identified as the relevant unit of analysis since human problems are essentially behavioral problems (see Figure 3).

Some interesting methodological arguments follow from the above tenets, and these can be most effectively developed via a (rather tedious) discussion of the processes of decision-making. So essential is a systematic analytic perspective in these matters that there seems little need to argue for its implementation in approaching complex design problems. This perspective underlies the comments which follow. That is, a problem-solving task can be seen to involve the following kinds of sequential operations shown in Figure 4: (1) model (what ought to be), (2) simulate (what would happen

1. The generic term "behavior," without further qualification, has been criticized as too vague to deliver the kind of substance—the richness and complexity—required in describing human systems of interest. The definition of behavior as an empirically accessible change in organismic state will suffice for purposes of this discussion.
In this and the following two sections I will attempt to outline a framework for the design and management of human systems in three phases, increasing in complexity as new arguments are accommodated. The first series of documents deals with the problem solving structure which emerges given the above assumptions. That is, human system design strategies (1) are directed toward the realization of goal-oriented systems, (2) must be organized around a systems analytic perspective, i.e., a model-simulate-implement-test paradigm, (3) are behavior-contingent, i.e., such systems must be analyzed, realized, and evaluated in terms of empirically accessible units of behavior.

MODELLING desirable human settings begins with goals. As mentioned above, artificial systems are designed in response to the way things ought to be rather than the way they are. The normative task of identifying goals for human settings is among the most frustrating and complicated aspects of planning and design. The issues are too complex to fully explore here, but can be briefly mentioned.

The identification of goals for human settings cannot be accomplished empirically. We cannot observe things the way they ought to be, only the way they are. Goals or intentions are ostensibly...
embodied in the population, the situation under analysis. Clearly, there are individual and collective intentions which are not being accommodated; otherwise there would be no problem(s) to solve. Two kinds of technical issues are involved in goal-setting. The first is that of externalizing and communicating goals in order to act upon them. The second is that of analyzing these to detect iminical aspects, gross violations of constraints and their projected implication.

One of the principle failings of most political systems (at both micro and macro levels), as we all know, is communication—the inability of the decision-making apparatus to receive, analyze, and act upon the intentions of a population. The perceived goals of the members and subgroups within a human setting should be, but usually are not, available to the planning and design instrumentality via open, minimally dampened channels. Beyond communication, however, is the need to probe the attitudes, "feelings," and internalized models (Miller, Galanter and Pribram, 1960) which direct the behavior of those involved. Once this information is available, it must be analyzed and ordered. A distinction must be made, for example, between individual and collective goals (the latter are usually not simply an aggregate of the former), and resolution of inevitable conflicts between (1) individual and collective, (2) inter-individual, (3) intra-individual goals.

There must be a hierarchical ordering of immediate, intermediate, and long-range goals. Some goals are obviously impossible for a particular population to attain, and these must be identified and revised. A geriatric population would employ certain constraints which would be dissimilar to another. However, there are many ways to attain goals, and one should be cautious not to lightly discount certain possibilities. The objective of human systems design is, after all, to overcome extant limitations to goal attainment.

Finally, goals must be delineated in dimensions which admit functional extensions. Unless goals can be operationally defined in empirical units of behavior, they are pretty nearly vacuous for design purposes; i.e., they must implicitly contain information and directives for design decisions.

These dimensions of goal setting are obvious and well known. The purpose in reviewing them here is to emphasize that here is an aspect of planning which must become much more refined, and it is one requiring a great deal more research attention. Because goal setting is an area of normative concern, it has usually been viewed as beyond systematic consideration. Tools exist to address this issue and these should be further developed—more innovatively and integratively utilized. As noted above, advocacy planning has emerged as a means of making the recipients' desires felt in decision-making. It is operationally rather primitive, however. There is need to invent and/or apply the advancing techniques of communication and information management, for the entire population concerned should be somehow involved in goal-setting (and other) aspects of the planning and design process. Other tools which might be mentioned include (1) attitudinal surveys, (2) participant gaming techniques for goal setting, (3) simulation (of goal consequences to assess conflicts). It is clear that goals and their determination have always been an implicit aspect of human processes; it must become a more explicit aspect before human systems design can reach the required level of sophistication.

Goal setting is an initial step in addressing human problems, but goals cannot be directly transferred into solutions. Rather, they must be transformed into an explicit problem statement. Human behavior has been identified as the relevant class of variables. The goals of a population must be operationally defined in empirically accessible units of behavior before they can be acted upon. It is necessary to develop explicit descriptions of what a human system is intended to accomplish before alternative conditions (the environment) can be arranged to support them. In operations research jargon, an objective function is needed to direct the design operations. The development of a behavioral model is itself a design problem. The "art" of behavioral design is another of the important areas which has not really received the attention it deserves. The descriptive or analytic mode which characterizes research in the human sciences precludes great involvement in the problems of behavioral design. Modelling a requisite, goal-oriented behavioral network with appropriate precision is, admittedly, complicated. We have no coherent methods, no effective algorithms for approaching this phenomenon. Developed resources, such as organization theory, network theory, information theory, simulation of human groups, and several other areas have relevance here. But these resources are disparate, and must be integrated and augmented to realize a more coherent approach to this modelling task. Hopefully, emerging interest in design of the artificial (e.g., H. Simon 1969) will stimulate appropriate research in this area.

Because human systems design is behavior-contingent, the behavioral model, once developed, becomes the generating force, the class of independent variables for subsequent operations. The second stage of modelling involves development of characteristics of the supporting environmental structure required to realize the requisite state of behavioral affairs. What is required is a model of an environmental-behavioral ensemble. The behavioral component describes what is needed in terms of human outcomes; the environmental component describes how these outcomes can be accommodated. This model involves a prediction that a given environmental structure will produce
HUMAN SYSTEMS DESIGN AND THE MANAGEMENT OF CHANGE

a particular behavioral outcome. Clearly, this kind of functional model requires resources (i.e., a complete science of behavior), which do not yet exist; perhaps they never will. This problem concerning incomplete knowledge on which to base predictions will be discussed below. It can be noted here that the complexities of modelling predictable environment-behavior structures are such that this kind of operation is necessarily crude. Perhaps we should leave it at that for the moment.

SIMULATION becomes essential in one form or another in dealing with complex processes for which there are no available algorithms. There are no systematic means for transforming environment-behavior models into reality. The techniques of simulation are not highly developed where the vicissitudes of human behavior are concerned, but such techniques provide an important resource for realizing a modelled human system in the real world. Certainly, it is an alternative to guessing.

But there are additional difficulties which impinge on the process at this point. In order to produce an accurate simulation of an operating human system other conditions must be accounted for. The environment-behavior model is essentially a representation of the internal system. In order to pretest its performance in the real world, the external environment (the product of a particular spatial-temporal context) must be accounted for. The result is a more complex (simulation) model which represents not only the previous elements (the internal environment) but elements of the external environment as well. The output of an appropriate simulation is clearly the resultant behavior structure produced by the environment represented in the model.

Because of (1) the incompleteness of knowledge required to develop an environmental model, (2) the lack of algorithms to transform this model into a real system, and (3) the inherent complexity of interactions of (internal and external) variables, it is highly unlikely that the results of an initial simulation will produce the requisite human states. The process must, therefore, be seen as iterative (i.e., simulations must be reformulated and tested via behavioral outcomes) until there is sufficiently high probability that the conceptualized ensemble will perform properly in terms of the originally proposed behavioral outcomes (goals). The importance of simulation as an integral aspect of problem solving is that it increases the likelihood and demonstrates that an ensemble will perform as predicted. Human systems, even highly successful ones, are extremely fragile. Pretesting a solution via simulation does not, of course, insure success, but these tools are essential and must be greatly improved to reduce the risk of implementing untested models.

It appears that several modes of simulation are required to pretest and refine models of human systems including, (1) computer—both digital and analog, numerical, and graphic—, (2) human, (3) hybrid, and (4) physical. The future is closely tied to developments in simulation. As these tools are refined in the future and (they definitely will be), fidelity between simulated and real phenomena will be increased to the point where highly complex systems can be pretested and implemented with minimal risk of failure.

IMPLEMENTATION of a designed human system, realization of an alternative to that which exists, presents great difficulties in most situations where democratic processes are in effect. If the above operations have been properly executed, implementation of a system should come as a matter of course—that is, if the goal structure has been properly developed, and if the population under analysis plays an interactive role in the decisions. The greatest challenge to implementation, then, is the nature of rapport between social change agents and the recipients. Important as the interpersonal dynamics of social change are, they will not be further pursued here. Assuming that a human systems design has been implemented in the real world, we come to perhaps the most neglected requirement.

TESTING or evaluation of implemented systems unfortunately has not been an integral aspect of planning and design. This weakness is particularly obvious in the domain of physical systems, and attempts to systematically evaluate other human services (e.g., health systems) are relatively new and still somewhat unsuccessful. Unless problems are formulated in dimensions which are meaningful in human terms and empirically accessible, there is little basis for asserting that a human problem has actually been solved. The above decision structure provides the required basis for a test; i.e., modelled behavior can be compared with the actual behavior produced—a systematic comparison between what was intended and what was accomplished.

The most difficult aspect of testing implemented systems involves problems of monitoring behavioral outcomes on which to base such a test. Here, then, is another challenge for the future (i.e., the means of monitoring human systems to identify and analyze dysfunctions in order to make

3. A bridge designer can, of course, produce a solution with the unerring prediction that the bridge will perform in a given environment. The resources of the more mature physical sciences have been utilized in developing algorithms for this task. The engineer cannot, however, predict the human response to his bridge. On this matter, conventional wisdom is invoked.

4. It becomes very clear that representative actors from the population under analysis participate in the simulation (as well as other aspects) since the possibilities for simulating their functions are highly limited.
adjustments before these become serious. The area of social indicators may have promise, but I suspect that more direct measures will be developed (e.g., more precise remote monitoring via technological innovation). Trade-offs between the need for individual privacy and the need to detect dissonances before they reach critical levels are matters requiring extensive debate and resolution. One prospect of systematically testing implemented human system designs is that such a test may be negative in certain respects. This suggests the need for continuing iterations and adjustments until there is consonance between modelled and resultant behavioral structures. Considering extant technology (hardware and software), the need for this kind of iteration—adjustment of systems in response to post-testing—presents another challenging problem in realizing a better future. This particular point will be elaborated as the implications of this behavior-contingent approach are further explored.

Even in accepting a satisficing (Simon 1957) rather than an optimizing principle, we must operate with predictive models based on incomplete understanding of human processes (functional relations), and with problem situations which exhibit enormous complexity. The risk of failure is ever present. A model-simulate-implement-test problem-solving format with appropriate, iterative feedback links provides a powerful method of problem solving with partial knowledge. While the desirability and logic of this approach to planning and design are fairly self-evident, counterexamples in present planning procedures are the rule.

Problems of the Human Dynamic

The arguments for a behavior-contingent approach to designing artificial human systems require refinement to account for certain characteristics ignored in the above formulations. Research in general systems theory tells us that a fundamental distinction must be made between closed and open systems. Closed systems are isolated from their environment; open systems are not. The Second Law of Thermodynamics holds for closed systems, i.e., "entropy (roughly a measure of disorder) of a closed system will always increase toward a maximum, attained in equilibrium" (A. Rapoport 1968). In closed systems, equilibrium can be assigned by virtue of the behavior of the system, i.e., homogeneity of the closed environment. All living systems are open, and the concept of equilibrium has no intrinsic meaning. In constructing an open physical system "purposefulness," or a precept of goal state, is a part of the design along with appropriate feedback links, e.g., an automatic steering device. In natural living systems, internal and external events afford self-correcting mechanisms which tend to "stabilize the system," but not in the equilibrium sense of a closed physical system. Human systems are, of course, open systems, and design of a support environment for them (e.g., economic, political, physical) must accommodate this characteristic.

While exposing some frightful complexities, the previous analysis was still simplistic (as are many design strategies) in treating human systems as closed. The implication was that a steady-state environment-behavior ensemble would, once designed and implemented, remain stable. We all know that this is not so. Like all living systems, human systems (individually and collectively) receive information and energy from the external environment which changes the state of that system (the internal environment). This is clearly what makes predictions so difficult and insures a pervasive uncertainty in the design and maintenance of human structures. Human systems, i.e., man-environment ensembles, are subject to constant change as a result of exogenous events. These changes in state require adjustment in the internal environment along several dimensions and time scales. In the context of previous formulations, the sources of variability can be readily identified. First of all, goals are subject to change in an open living system if that system happens to be human. When these goals change, new behavioral and, thus, environmental states are required. To the extent that goals change without the required adjustment in affected aspects, dissonance results.

A second source of variability is in the impinging external environment, i.e., its constraints on the internal environment. These are in constant flux, and when the impinging variables reach significant levels of change—when the interface between internal and external environment is significantly altered—the internal supporting structures, (e.g., economic, political) require adjustment. If there is no such adjustment, a state of bad fit or dissonance occurs between external impingements and the behavioral goals.

A third and perhaps most profound source of variability in an implemented setting involves changes in the human participants. Humans are both individually and collectively state-changing, open systems, and their response probability in the presence of a particular environment will vary as a result of deprivation states, adaptation, gross physiological changes, and perhaps most important, learning. This last item, the dynamics of learning, turns out to be the central concern of this approach to designing and realizing human systems. It suggests that the environments required for most human enterprises of interest are considerably more dynamic than those presently available. This is a point requiring further development, but first let us see where this analysis has brought us.

The objective of more systematically organizing the human environment to meet the goals of its inhabitants is certainly justified, and the decision-making structure outlined is, I submit, quite viable in this regard. However, the implementation
of such a decision-making format is subject to three kinds of limitations. Our ability to model predictable environment-behavior ensembles is limited by the state of knowledge in human processes. We have no algorithms with which to realize predictable environment-behavior models even if these were available. Finally, the dynamics of man-environment systems (i.e., goal structures, the external environment, and the human participants) can produce dissonance in the ensemble requiring systematic adjustments. Clearly we are dealing with a problem-solving domain which embodies great uncertainty. Is there a way to deal with this unruly and fundamental problem? I think there is.

Programmed Change in Human Systems

We can get to the future only from the present. It helps little to visualize or even to systematically design future states unless we can describe the means for getting there. "Getting there" inevitably involves a change in human behavioral potential. To design a new culture or subculture is one thing, to realize it is another. In the two previous sections, I discussed an approach to designing behavior-contingent artificial human systems, and then elaborated these arguments to account for the intrinsic demands of living, as opposed to non-living processes.

An indispensable aspect of planning is prognostication. But "technical forecasts... should not be confused with prophecies in the general sense" (M. Brown 1969), for planning predictions involve a self-realizing element. Even so, such predictions are highly complicated, and the kinds of predictive environment-behavior models possible must be classified as conspicuously crude without a comprehensive, functional science of human behavior from which to draw resources (Studer 1969). Simulation has been identified as a likely tool in overcoming some of this difficulty, but even so the enterprise remains highly equivocal. In order to relieve some of this uncertainty, let us look more closely at the nature of the design task.

This entire argument is based on the observation that human problems produce the need for alternative behavioral structures. Specifically, the problem of human systems design is this: a requisite behavioral system has been specified in response to the goals of the human population under analysis; this population is not emitting this system of behaviors—otherwise there would be no discernible problem to solve; this population has, in all likelihood, a diverse behavioral history. The problem is to specify and realize an environment which will produce, with the highest probability, the goal-related behaviors. What this clearly describes is a learning problem, i.e., the acquisition of or modification toward a new set of behaviors.

Behavior is caused by genetic endowment (phylogeny), history of interaction with the environment (ontogeny), and the extant environment. In most problem situations, only the third class of variables is available to effect behavior change. (This may or may not be true in the future; e.g., genetic engineering.)

As everyone knows, the behavioral sciences encompass a variety of conflicting explanatory theories and an array of clinical formats for bringing about behavioral change. It is not in the interest of this analysis to enter into a comparative examination of these. We can only look to the future for productive unification of competing "schools" of behavioral explanation. It is, after all, a very young science.

There is a branch of psychological research which appears in its applied form to have great relevance for human systems design, i.e., directed behavior change. An operant analysis of behavior has been demonstrated as highly effective for an understanding of the conditions under which behavioral topographies can be seen to systematically change form. The experimental analysis of behavior as developed over several decades in the laboratory (e.g., Skinner 1938, 1957, 1969; W. Honig 1966) and to a more limited degree in applied settings (e.g., Ulrich, et al. 1966, Bandura 1966), explores the methods by which behavior can be brought under the control of specific elements of the environment. It involves the spatial-temporal relationship between behavior and its consequences. A particular temporal-spatial arrangement of consequences in the environment will affect the probabilities of specific behavioral topographies' recurrence. That is, if these consequences are made contingent upon specific behaviors, these behaviors come under their control. This is a complicated way of saying that people need "reasons" for behaving in one or another way.

Environmental consequences can be either favorable (e.g., money, social praise) or aversive (e.g., fines, punishment, repression). The new technology of teaching (Skinner 1968), for example, is based on the principle of systematically relating desired behavior to positive consequences. A great deal of our general behavior is regulated, however, by negative consequences because aversive control brings more immediate results and is more economical to maintain (on the part of the controller)—but at a human price (on the part of the controllee). The aversive control of behavior has many side effects which are biologically, socially, and technically undesirable. Human settings which are regulated by the extensive use of negative controls are in constant jeopardy of extensive dysfunction. This is common sense; but it has also been verified in rigorous laboratory settings. It is a principle which the designer of human systems must understand quite clearly if we are to get to the future better off than we were in the past. A full explication of the principles of systematic behavior modification is quite complicated. These are but the barest elements, but they are sufficient to make
some general points regarding the management of change.

The emerging applied discipline of contingency management represents a much broader application of operant principles. Contingency management has been utilized primarily in institutional and educational settings with notable success. An expanded utilization of these techniques in some form is predicted for the future. The prospect of utilizing the techniques of operant conditioning traditionally raises ethical issues (the "who controls the controller" response). These cannot be fully addressed here except to note that a social change agent, i.e., contingency manager, can be seen to develop a "contract" with an individual or group to produce a behavioral situation which meets their goals. There appears to be nothing sinister in such a proposition. This is what educational settings are all about, is it not?

To return now to the original point of concern, the objective of human systems design is to move extant behaviors to another state, and maintain it until a new goal structure is specified. Social change is behavioral change (J. Kunkel 1969), and it is well known that the acquisition or modification of complex behaviors is not accomplished in an "all or none" fashion. Research in human learning confirms that transformation from one complex state to another involves a series of intermediate states, i.e., behavioral shaping (Goldiamond 1966). Plans for and planners of human settings have not generally accommodated this learning dynamic. Little wonder that finite-state environments (most clearly illustrated, for example, in architectural systems) rarely consummate the specified goals of the recipients.

This analysis began with the identification of (normative) goals. Beyond the need to externalize and analyze the goals of the existing population, the management of social (behavioral) change suggests the need for additional data and elaboration of certain arguments. The process of implementing a program of behavioral change can be generally described as (1) empirically analyzing behavioral (and environmental) baseline states, (2) defining alternative behavioral states, (3) managing contingencies in the environment which will bring about and maintain the alternative states. This process is, as noted, a dynamic one which embodies the principle of successive approximation.

These realities of behavior modification, growing out of the laboratory and applied settings, suggest some elaboration of the previous formulations concerning planning and design processes. To begin with, the potential for change must be found within the population under analysis. An existing human setting must be empirically analyzed to assess baseline states both in the behavioral and environmental structures, as well as the functional relations between them, i.e., an assessment of the contingencies in the extant setting which are maintaining the observed behaviors. Another important area requiring attention is that of modeling the environmental-behavioral states required to meet the defined goals.

The model describing the requisite environment-behavior system must describe not only the organization of a single terminal state (for a given goal structure) but also a series of interim states which incrementally modify these structures toward that state.

In other words, the environment-behavior interface is dynamic; functional relations must be described as taking n states along a continuum. What we have here is a dynamic (to accommodate the process of arriving at a requisite state) within a dynamic (to accommodate changes in the external environment). Figure 5 summarizes the arguments for a behavior-contingent approach.

Conclusions: A Controlled Approach to an Uncertain Future

Even in its generality this has been a rather laborious approach to dealing with the future. A straightforward forecast or description of a desirable future would have perhaps been much simpler. But one must take seriously Karl Popper's (1962) warning that we should not commit ourselves to a particular plan for the future. Quite aside from our inability to predict the events which will alter such plans, overcommitment to the ideologies which inevitably accompany them have led to human disasters which we do not want to repeat. I favor, along with many others, a flexible but controlled approach to the future, where any number of plans are realizable. Uncertainty is, and probably will remain, a fundamental condition of man, but uncertainty obviously does not preclude action. It is perhaps true that there is no progress—only process—in human affairs, and there are probably no solutions to human problems, per se. Our ability to respond incisively to human dysfunctions is, within the bounds of present knowledge, critically limited; more important, human problems change before solutions to them can be realized. In approaching the future, our design objective should respond more directly to the variable nature of open, living systems. We must address the problems of unpredictable states and general maintenance of the environment-behavior interfaces on a continuing basis. Artificial human systems must be viewed literally as experiments in which relevant variables are systematically managed to move the system toward consonance with the external environment and its endogenous goals. A planning committed to adjustment and change is one in which the process of living becomes a process of (systematic) experimentation to upgrade human existence. The hardware and software involved in realizing experimental settings is complex and no doubt expensive. But the consequences of human stress, obsolete ideologies, and the untenable squandering
of human and material resources is also expensive. An innovative, experimental approach to individual and collective existence has more to offer.

The position that human systems design is behavior-contingent has led to a possible approach to the management of change, a vehicle for moving from the present to an uncertain future in a controlled manner. There has been little discussion of the characteristics of behavior-contingent human systems themselves. The term "human sys-
tem" has been used quite often in this discussion. A system has been defined as "a whole which is a whole by virtue of the interdependence of its parts" (Rapoport 1968). It could be observed that some human systems are more "virtuous" than others in this regard. A behavior-contingent perspective, and all that is implied by it, has much to commend it in terms of its organizing potential. The greatest difficulty in realizing viable human settings, it is argued, grows out of misdefining problems. Traditional approaches to problem solving are inundated with preconceptions regarding the various components and arrangements of these in human settings. For example, the concept of "city"—the various elements connoted, and relations among them—seems unnecessarily and inappropriately constrained by extant configurations and preconceptions. The various subsystems which make up the human environment (e.g., transportation, housing, medical, educational) are considered discrete, requiring specialists to deal with them. Indeed vast, complex bureaucracies have evolved which insure that these subsystems are neither redefined nor properly interrelated. The economy of effort and sheer cost of maintaining such systems is clearly untenable, as every urbanologist knows, and the problems of management seem insurmountable as the fragmentation and deterioration continue.

Entirely new human systems topographies are required. A behavior-contingent perspective, I claim, offers a direction for more viable taxonomies and a new sense of system in the environment based on the interdependence of human functioning. A behavioral analysis—an understanding of the contingencies which define interactions with the environment and with each other—embodies the substance for eventual reorganization. Basic human service systems, such as education, the administration of justice, medical services, recreation, and transportation, each require serious reexamination in order to fully realize not only new structures but also the potential for reorganizing these into new subsystems, contingent on human behavioral goals. The new teaching technologies are aggregated until they reach the level of crisis. A behavioral analysis—an understanding of the contingencies which define interactions with the environment and with each other—embodies the substance for eventual reorganization. Basic human service systems, such as education, the administration of justice, medical services, recreation, and transportation, each require serious reexamination in order to fully realize not only new structures but also the potential for reorganizing these into new subsystems, contingent on human behavioral goals. The new teaching technologies are aggregated until they reach the level of crisis. An alternative approach would be one in which the environment is organized so as to make the consequences of pollution behavior immediately available to the polluter. That is, we need to disaggregate these consequences and reorganize them into finer-grained networks which bring them into more effective spatial-temporal proximity to the related behaviors.

This principle of feedback from the environment through the organization of appropriate contingencies is an important one. The challenge to biological and extra biological survival comes down to our ability to organize man-made environments which exhibit the subtle order found in natural living systems. These systems are composed of complex networks of interdependencies which can be characterized as self-regulating. Not only have these natural systems come to terms with population control (the first priority of our own design efforts), but they also have evolved elegant social control mechanisms (e.g., McBride, et al., 1963). Due to the misapplication of technology, many of these self-regulating mechanisms are absent in contemporary human settings; they must be designed into future efforts. The behavior-contingent approach to human systems design provides the substance for this objective.
REFERENCES


A SYSTEMS APPROACH FOR THE ANALYSIS AND DESIGN OF STIMULI, RESPONSES, AND EXPERIMENTS

Ian I. Mitroff and Francisco Sagasti

1. INTRODUCTION

In many respects, the most troublesome problems of any science center around its most basic terms and fundamental concepts, not around its more sophisticated concerns. Indeed, to the extent that everything either follows from or is based on a discipline’s basic terms and fundamental concepts, problems at a higher level can always be traced back to problems at a more fundamental level. Psychology in particular and the behavioral sciences in general are certainly no exception. One of the most irritating and troublesome problems facing the behavioral sciences is that adequate definitions for the terms "stimulus" and "response" still do not exist; their fields of definition continue to remain a continual battleground. There is confusion, disagreement, and even outright antagonism over the definitions of these basic terms.

In an excellent paper, "The Concept of The Stimulus in Psychology," Gibson (1960) surveyed the various senses in which the term stimulus was used in psychology. He made the highly pertinent observation that while we have made continual efforts to found a systematic theory of behavior, we have made no such efforts to found a systematic theory of the stimulus. He also noted that while every psychologist makes use of the concept of a stimulus, few have bothered to define it with precision. Further, of the definitions we do have, many, if not most of them, flatly contradict one another. Garner (1970), on the other hand, while paying tribute to Gibson’s paper, takes a somewhat different point of view. While he acknowledges the confusion and multiplicities of meanings surrounding the term, Garner nevertheless argues that the problem is not necessarily to restrict the definition of the term stimulus to a single, unambiguous, and generally accepted meaning but rather to find out under which conditions, which definition or concept is best suited for which kinds of experiments or problems in Psychology.

Basic as the problem of defining a discipline’s fundamental terms is, it is indicative (as all problems of definition are) of even more generic difficulties. At the heart of the difficulties in providing adequate definitions for the terms "stimulus" and "response" is the fact that the behavioral sciences are sorely in need of their own philosophy of science, i.e., a philosophy of science that is especially suited to their particular phenomena and methodology (Campbell 1969, Mischel 1969, Toulmen 1969, Turner 1967). Wolman (1971) has put the problem squarely in front of us by posing the issue rhetorically, i.e., in an article entitled, "Does Psychology Need Its Own Philosophy of Science?" A simultaneous reading of Gibson (1960) and Wolman (1971) leads one to the conclusion that many of our difficulties with the terms stimulus (S) and response (R) follow almost directly from the unreflective and almost wholesale transfer of philosophies of science underlying the physical sciences to the behavioral sciences. Not only is it highly contentious that a philosophy of science adequate for one set of sciences is ade-

1. On this point, Garner is worth quoting in some detail since he has put the point extremely well:

We would never consider drawing conclusions from an experiment based on a single subject, because we know that a different subject might give us a different answer. And we have even learned to be cautious about drawing general conclusions based on the use of a single task. Continuous and discrete reaction-time tasks, for example, might just lead to different results. Why, then, are we apparently so happy drawing sweeping conclusions about how information is processed when we have used only one kind of stimulus? I am not about to argue that we forgot what we have learned about subject and task variations in our studies: I do want to argue that we become equally concerned about the nature of the information input.

Multiprocess Organism

Another point that has not been given sufficient consideration in our studies of information processing is that the organism is truly capable of many different types of information processing. Yet psychologists have often assumed that only a single information process operates for the organism. How many papers have been published in the past two years in which the question being asked was whether the organism processes information in serial or in parallel? Why must the organism do just one or the other? Very probably it can do either, depending on the task and the stimuli. And even as likely is that the organism frequently does both, not in the sense of doing first one and then the other, but in the sense of doing both simultaneously.

Now the particular pertinence of arguing for a multiprocess organism when I want to emphasize the role of stimulus concepts in information process is this: One of the most important variables in determining how the organism will process information is the nature of the input itself, the very thing that has received so little attention in our research. If we want to understand how the organism processes information, we must be prepared to ask under what circumstances a particular form of processing is used. And one of the most important of these circumstances has to do with the nature of the stimulus input, the information to be processed (Garner 1970, pp. 350-351).
but it is not even clear how a philosophy of science even if acceptable should be carried over from one science to another. For example, the definitions of S/R in psychology have varied greatly depending on what was being "carried over" from physics. If it was the concept of "force" that was being transported, then S/R were defined one way; if "energy" or "event," then S/R were defined in other ways (Gibson 1960; Wolman 1971).

The purpose and plan of this paper is four-fold. First, we show that the concepts S and R can be defined in systems terms. This will enable us to show that the terms S and R can be given a purely abstract, conceptual interpretation in addition to their having a concrete (i.e., physical) interpretation. This will also enable us to make the point that S/R never exist as isolated, independent entities. They can only be defined and given meaning by reference to, (i.e., in relation to) the surrounding system of which they are a part or in contact with.

Second, we discuss some basic (i.e., archetypal) kinds of systems. This will allow us to indicate explicitly the range of possible kinds of distinct S's and R's that can exist. Since these systems derive from epistemology, this step will also allow us to indicate some philosophies of science particularly appropriate for the behavioral sciences.

Third, we develop a matrix of S and R permutations. This step indicates the basic kinds of distinct S/R experiments that can exist.

Fourth, we indicate some of the archetypal kinds of experiments that result from our classification.

In effect, the first two steps allow us to acknowledge explicitly the fact that there are a multiplicity of ways for defining S's and R's and, that each of these ways is equally valid; the last two steps provide us with a rich enough conceptual framework to see how each of these ways relate to and build on one another. Thus, while each of the ways is radically distinct from the others, they all nevertheless relate to one another. The last two steps also allow us to show how the range of conceivable experiments has been greatly limited by our failure to recognize, develop, and utilize the full range of conceivable S/R types.

We begin by defining the concept of a system:

2.1 SYSTEM(S): An entity consisting of two or more elements (E) and a non-empty set of relations (R') holding among the elements.

Note that this definition makes the concept of a system exceedingly general. The definition places no limitation at all on the kinds of entities (or elements) that may constitute a system. The entities can be physical as well as abstract (i.e., symbolic) or purely conceptual. For example, the elements could be actions, individuals, organizations, discussions, etc., and the relations could be interactions, communication, exchange of resources, information, orders, or any other conceptual or physical construct which bind or connect the elements (E). This initial concept of system does not involve aims, functions, or purposes; it is merely a "confining" concept which divides the all inclusive set of elements and relations each into two subsets: those of direct concern to the observer or experimenter and those of no interest to him. Notice that the identification of a system, its properties, its boundaries, its structure, and its function will depend on the purposes of the observer.

The set (E) of elements belonging to the system can be subdivided into two mutually exclusive and exhaustive subsets: a subset called the environment (E_e) and another called the object (E_o). The set of relations can be divided into three subsets: those relations between elements belonging to the object (R_o), those between elements of the object and the environment (R_e) and those between elements of the environment (R_e'). Environment and object can be defined in terms of elements, relations, and the interests of the observer.

2.2 ENVIRONMENT (E_e): A subset of elements of (E) and their relations, such that the relationships between them (R_e') are of no direct concern to the observer. Only the relations between elements belonging to this subset (E_e) and the members of the object (E_o) are of primary interest to him (i.e., R_e).

2. The discussion in this section follows the approach to systems suggested by Ackoff (1971).

3. As Ackoff has put it:

Although concrete systems and their environments are objective things, they are also subjective insofar as the particular configuration of elements that form both is dictated by the interests of the researcher. Different observers of the same phenomenon may conceptualize them into different systems and environments [emphasis ours]. For example, an architect may consider a house together with its electrical, heating, and water systems as one large system. But a mechanical engineer may consider the heating system as a system and the house as its environment. To a social psychologist a house may be an environment of a family, the system with which he is concerned. To him the relationship between the heating and electrical systems may be irrelevant, but to the architect it may be very relevant (Ackoff 1971, p. 609).

See also Churchman and Ackoff (1947) for a further explication of how the purposes of a researcher define his objects or elements of inquiry.
2.3 OBJECT \((E_0)\): A subset of elements of \((E)\) and their relations, such that the relationships between them \((R_0')\) are of direct and primary concern to the observer.

A system is thus defined in terms of an "object-environment" couple, which implies that the environment forms an integral part of the system, and that object and environment are complementary subsets with respect to the system.

The observer's "interest" may be derived from his possibility of controlling or modifying the elements in the system. In this case, the environment would be defined as those elements, together with their relations, which affect the object but over which there is no possibility of exerting direct control.

Next we bring the concept of "function" into our conceptual framework. In order to do this, we require some preliminary definitions. Given a specified class of entities \(Y\) (which can be objects, concepts, relations, events, etc.) and a particular system \(S'\), we define (following Churchman and Ackoff 1947):

2.4 NECESSITY: A state of the system is necessary for the existence of \(Y\) at a later time if, (1) under a specified environment and set of environmental relations, the state is always followed by \(Y\) at a later time, and (2) given a different state and no change in the environment and set of environmental relations, \(Y\) would not appear at a later time.

2.5 SUFFICIENCY: A state of the system is sufficient for the existence of \(Y\) at a later time if the state is always followed by \(Y\), under any environment or set of environmental relations. (A state of the system is insufficient for the existence of \(Y\) at a later time if there is a particular environment and/or set of environmental relations, under which \(Y\) never follows the state under consideration.)

2.6 PRODUCER-PRODUCT: A state of the system is said to be the producer of a class of entities \(Y\), the product, if this state is necessary and insufficient for the existence of \(Y\) at a later time.

2.7 POTENTIAL PRODUCTION: A system is said to be a potential producer of \(Y\) if more than one state has produced \(Y\) and there exists at least one state which has not produced \(Y\).

Necessity and sufficiency refer, in effect, to deterministic causality, as typified by, say, Mills' Canons of Induction (cf. Ackoff 1962). Producer-Product refers to probabilistic causality (Churchman and Ackoff 1947, Singer 1959). In terms of these definitions we can define the concept of "function":

2.8 FUNCTION OF A SYSTEM: The class of entities \(Y\) is said to be the function of a system \(S'\) if \(S'\) is a potential producer of \(Y\). For example, consider a transportation system that has a purposeful function, \(Y\), e.g., "the transporting of people from point A to point B within some specified time interval \(T\) and acceptable cost range \(C\)."

Then the system \(S'\) consisting of the entities "cars, buses, bicycles, etc., plus surrounding city environment" is only a "potential producer" of \(Y\) because not all "states" of \(S'\) will produce \(Y\) (a fact that is only too apparent and characteristic of modern urban transportation systems). The reason we require that not all states of \(S'\) produce \(Y\) is that it makes no sense to talk about the "function" (and especially "purposeful function") of a deterministic system, i.e., a system that can and will produce one and only one effect no matter what the state of the environment for a given cause. Indeed, Churchman (1965b) goes so far as to reserve the term "system" only for those entities which satisfy definition 2.8, i.e., a system is an entity if and only if it satisfies both 2.1 and 2.8. Thus, according to Churchman:

We postulate that systems are teleological entities. This means that if something is to be called a system, there must be alternative systems (i.e., at least two alternative means), and there must be a designer whose intentions are expressed in terms of the common potential products of the sets of systems (i.e., the common set of functional purposes or goals that all the systems are pursuing) (Churchman 1965b, pp. 3-4).

Given these notions, we can define the concepts "stimulus" and "response." We consider a system in the Churchman sense (defs. 2.1 and 2.8). Thus, we consider a system composed of defining elements \(E\), and relations \(R'\) and a class of functional entities \(Y\).

2.9 STRUCTURAL STIMULUS: A structural stimulus is one or more modifications of a system's initial defining elements \(E\) and/or relations \(R'\).

2.10 FUNCTIONAL STIMULUS: A functional or teleological stimulus is one or more modifications of a system's initial defining functional entities \((Y)\).
2.11 STRUCTURAL RESPONSE: A structural response is a subsequent modification of a system's defining elements (E) and/or relations (R') in response to or as a result of an initial modification of E, R', and/or Y; i.e., a structural response can result from either a structural or functional stimulus.

2.12 FUNCTIONAL RESPONSE: A functional or teleological response is a subsequent modification of a system's defining functional entities (Y) as a result of an initial modification of E, R', and/or Y.

Notice that a system may be subjected to any combination of initial structural and/or functional stimuli and respond in any combination of structural and/or functional response patterns.

With these notions and definitions, one can respond to a number of Gibson's queries regarding the nature of stimuli and responses. For example, Gibson (1960) asks, "Can a stimulus be taken as the sufficient cause of a response, or can it not?" (p. 695). The answer is clearly "no" since a system, and hence the concepts S/R, are defined through a producer-product relationship. By definition a particular stimulus (or producer) is necessary but not sufficient for the production of a given response (or product). Gibson further asks, "Must a stimulus be defined independently of the response it produces— in physical rather than terms of behavior or sensory terms?" (p. 696) and "Do stimuli exist in the environment or only at the receptors?" (p. 697). Like the terms "object" and "environment," the definitions of S/R emphasize that they are, if only in part, dependent upon the interests and background of systems observer (conceptualizer) or experimenter. In this sense, the terms S/R will always have an ineradicable conceptual element. Thus, stimuli may be simultaneously conceptualized or analyzed into their physical, behavioral, and/or sensory components. (And they may exist in the environment or at the boundary of the object.) Recalling Garner's (1971) earlier response to Gibson, it is not an either/or. The appropriate questions are: Under which conditions is a physical conceptualization more appropriate than a behavioral one? For which conditions are both equally appropriate? These questions we take up in the next section where we consider various kinds of systems, i.e., archetypal kinds of conceptualizations.

Further systems concepts could be defined, e.g., "adaptiveness" (Ackoff 1971, Sagasti 1970) and "separability" (Churchman 1965b). These would allow us to respond even further to Gibson's questions, e.g., "When is a pattern or relation to be considered a single stimulus and when a number of separate stimuli?" However, at this point, it is more fruitful to proceed to an examination of different kinds of systems, since this examination will not only allow us to respond to Gibson's concerns but give us a unique perspective on systems research. Before proceeding, we emphasize that stimuli and responses NEVER exist as isolated entities. They take on their basic meaning and definition ONLY by reference (i.e., through R') to a set or system of other entities (E).

3. PHILOSOPHICALLY DERIVED SYSTEMS OF INQUIRY

The systems we shall examine come from the history of Western epistemology. We use epistemology because (1) it represents the most general attitude of man towards the problem of how to build models (conceptualizations) of the system called the "real world"; (2) the history of epistemology is anything but unanimous in its choice of a single, "best" model; hence, (3) a survey of Western epistemology provides us with the broadest possible survey of the most disparate, widely conflicting attitudes that man has taken toward the problem of conceptualizing fundamental models.

Our models derive from the recent efforts of C. West Churchman to formulate some of the major systems of epistemology in such a way that they could become of direct relevance to the information-systems needs of the practicing scientist. The title of Churchman's effort, The Design of Inquiring Systems (1971), emphasizes that to conceptualize or to model a problem is to conduct an inquiry into its nature and that to conduct an inquiry is to gather or produce some information on the problem's nature. Thus, in this sense, "information" is a function of "epistemology." What we know about a problem (i.e., the information we have on its "nature") is a function of how we have obtained that knowledge, i.e., of some system of inquiry. Because information is a function of inquiry, The Design of Inquiring Systems is an exploration into the design of archetypal philosophically-based information systems. To model a problem is to present information on its nature to some decision-maker who is (or may be) required to take action on the problem (Ackoff 1956, Churchman and Ackoff 1947).

Of necessity our discussion of each inquirer must be brief. For more extensive discussions the reader must be referred to Churchman (1971). See also Mitroff (1970, 1971, 1971a, 1972) and Mitroff and Betz (1972). Also, it should be understood that we are not claiming that these systems are exhaustive of the class of philosophic systems. Such a claim would be as silly as it would be pretentious. We merely claim that these inquirers are representative of some basic attitudes towards modeling. Further, we do not claim that our labeling of these models necessarily corresponds with historical usage. The labels represent Churchman's characterization of the major "spirit" of each Inquiring System (IS) and of the historic system.
to which each most nearly corresponds. Ours is an exercise in systems design, not in historical analysis.

Given a problem, each IS will, in general, produce a radically distinct "representation" or conceptual model of it. The reason is that each inquirer starts from radically distinct types of fundamental building blocks, i.e., primitive elements or "elementary units of information." In addition, each inquirer embodies a radically distinct kind of guarantor for insuring the validity of the "final information content" that is built up from the elementary building blocks. As a result, the final information which is generated from the blocks will be fundamentally (i.e., characteristically) different for each IS. In effect, the final information content of an inquirer with respect to a problem is that inquirer's representation of that problem.

Leibnitzian IS are the archetype of formal, symbolic systems. On any problem, they will build a formal, mathematical or symbolic representation of it. They start from a set of elementary, primitive "formal truths" and from these build up a network of ever-expanding, increasingly more general, formal propositional truths. The guarantor of such systems has traditionally been the precise specification of what shall count as a proof for a derived theorem or proposition; other guarantor notions are those of internal consistency, completeness, comprehensiveness, etc. The final information content of Leibnitzian IS is identified almost exclusively with its symbolic content. Leibnitzian IS are best suited for working on clearly defined (i.e., well-structured) problems for which there exists an analytic formulation as well as solution.

Lockean IS are the archetype of experimental, consensual systems. On any problem, they will build an empirical, inductive representation of it. They start from a set of elementary empirical judgments ("raw data," observations, sensations) and from these build up a network of ever-expanding, increasingly more general, sets of "facts." (Where in a Leibnitzian IS the networks are theoretically, deductively derived, in a Lockean IS they are empirically, inductively derived.) The guarantor of such systems has traditionally been the function of human agreement; i.e., an empirical generalization is judged objective, true, or factual if there is sufficient widespread agreement on it by a group of "experts." A beautiful example of a Lockean IS is that of a Delphi (Turoff 1971). The final information content of a Lockean IS is identified almost exclusively with its empirical content. Lockean IS are best suited for working on well-structured problem situations for which there exists a strong consensual position on the nature of the problem situation. If there does not exist a strong consensual position, or if the consensual position is suspect, no matter how strong it might be, Kantian and Hegelian IS may be called for.

Kantian IS are the archetype of multi-model, synthetic systems. On any problem, they will build at least two alternate representations or models of it. (If the alternate representations are complimentary, we have a Kantian IS; if they are antithetical, we have an Hegelian.) The representations are partly Leibnitzian and partly Lockean, i.e., Kantian IS make explicit the strong interaction between scientific theory and data. They show that in order to collect some scientific data on a problem a posteriori one always has had to presuppose the existence of some scientific theory a priori, no matter how implicit and informal that theory may be. Kantian IS presupposes at least two alternate scientific theories (this is their Leibnitzian component) on any problem or phenomenon.
From these alternate Leibnitzian bases, they then build up at least two alternate Lockean fact nets. The hope is that out of these alternate fact nets or representations of a decision-maker's or client's problem, there will be one that is "best" for representing his problem. The defect of Leibnitzian and Lockean IS is that they give only one view of the problem. Kantian IS attempt to give many explicit views. The guarantor of such systems is the degree of fit or match between the underlying theory (theoretical predictions) and the data collected under the presumption of that theory. Since in a Kantian IS information is neither purely theoretical nor experimental, the final information content is a function of both. Kantian IS are best suited for handling problems of "moderate" ill-structure. Problems that are "wickedly" ill-structured are best handled by Hegelian IS.

Hegelian or Dialectical IS are the archetype of conflictual, synthetic systems. On any problem, they build at least two, completely antithetical, representations of it. Hegelian IS start with either the prior existence (identification) of or the creation of two strongly opposing (contrary) Leibnitzian models of a problem. These opposing representations constitute the contrary underlying assumptions regarding the theoretical nature of the problem. Both of these Leibnitzian representations are then applied to the same Lockean data set in order to demonstrate the crucial nature of the underlying theoretical assumptions, i.e., the point that the same data set can be used to support either theoretical model. The point is that data is not information; information is that which results from the interpretation of data. It is hoped that out of a dialectical confrontation between opposing interpretations, e.g., the opposing "expert" views of a situation (Mason 1969), the underlying assumptions of both Leibnitzian models (or opposing policy experts) will be brought to the surface for conscious examination by the decision-maker who is dependent upon his experts for advice. It is also hoped that as a result of witnessing the dialectical confrontation between experts or models, the decision-maker will be in a better position to form his own view (i.e., build his own model or become his own expert) on the problem that is a "creative synthesis" of the two opposing views. [For a case study of this process, see Mason (1969); for a theoretical model, i.e., for a Leibnitzian model of a Dialectical IS, see Mitroff (1971); for a treatment of decision-theory from a dialectical point of view, see Mitroff and Betz (1971a).] Where in the Lockean IS the guarantor is agreement, in the Hegelian, it is intense conflict, i.e., the presumption that conflict will expose the assumptions underlying an expert's point of view that are often obscured precisely because of the agreement between experts. Hegelian IS are best suited for "wickedly" ill-structured problems. These are the problems for which, precisely because of their ill-structure, there will be intense debate over the "true" nature of the problem. Conversely, they are extremely ill-suited for well-structured, clear-cut problems because here conflict may be a time-consuming nuisance.

Singerian-Churchmanian IS are the most complicated and, hence, the most difficult to describe. We must of necessity refer the reader to Churchman (1948, 1971) and to Singer (1959) for an explication of their properties. In a very brief word, Singerian IS are best suited for studying all of the other IS. They do this by converting "wicked" problems into "structured" and "structured" problems into "wicked." Singerian inquiry shows how Leibnitzian and Lockean IS can be modified to work on wicked problems. How they do this is a most fascinating story that must be left to Churchman's book, The Design of Inquiring Systems (1971).
It would take a whole series of papers to analyze and classify properly the kinds of S's and R's that have been typically presented in behavioral science experiments. In an unpublished paper, Mitroff and Williams (1972) have attempted to show for the specific field of Information Science that the majority of studies have been overwhelmingly Leibnitzian and Lockean in character. The typical Leibnitzian study in information science has proceeded on the assumption that the fundamental problems are mathematical or symbolic; i.e., the fundamental problem is to find and then to study appropriate formalisms for the expression of "information." Examples of this kind can be found in those attempts to build abstract models for the retrieval of documents classified according to the formal system being investigated (cf. Baker 1969). Under this kind of study, little or no attention is paid to the thorny problem of how one gets the initial or "given" data that the system assumes is available for operation. It is either assumed that the problem of "data collection" is so trivial that nothing needs to be said about it or it is so complex that nothing can be said about it. Data collection is apparently a problem for somebody else or, in our terms, some other Inquiring System (IS).

The typical Lockean study in information science, on the other hand, has been undertaken on the crudest of empirical bases. Examples are the plethora of studies on "document relevancy" (Resnick and Savage 1964, Richmond 1963). The stimuli in these experiments are usually documents themselves or functions of them, like abstracts or various abstracting schemes; the responses are the "judgments" (e.g., rankings) of a group of "experts" as to the "relevancy" of the documents for a specified user, class, or purpose. More precisely, the responses are taken to be some "satisfactory agreement between the judgments of some group of experts." It is this feature which makes these studies "predominantly" Lockean in character, although not exclusively so, since the stimuli are also partially Leibnitzian. While we have no quarrel with the use of either agreement or expert judges in behavioral science—indeed, they are indispensable throughout all of science, physical as well as behavioral sciences (cf. Helmer and Rescher 1959)—we do question the validity of studies that have accepted "agreement" without much reflection on the nature of "agreement." In order for "agreement" to remain a valid epistemic rule in any study, we must ask, and in a deep and probing way, why the judges agreed. Would they have agreed if we had given them some other task?

The number of Kantian studies in information science is almost non-existent. One of the very few is Salton's SMART System at Cornell, which explicitly presents multiple models for performing automatic information retrieval (Salton 1968).

It would take an equally extensive survey to establish our point that in psychology the overwhelming number of studies and experiments have been Leibnitzian and Lockean in design and execution if not in their underlying philosophy. The underlying Leibnitzian and Lockean philosophical presumption of psychology is readily apparent from the analysis by Mischel:

These limitations on psychological research were reinforced by the belief that a science of behavior must begin with colorless movements and mere receptor impulses as such [NOTE: This is a Lockean Feature], and proceed to explain purposive human behavior in terms of "postulates involving mere stimuli and mere movement" (Hull, 1943, pp. 25-26). This belief about what "a satisfactory natural-science theory of behavior" must look like led behaviorists to criticize the use of cognitive concepts involving consciousness, like Lewin's "expectancy," "life space," and so on, on the ground that they are "subjective" and "introspective," "sheer anthropomorphism" rather than science (Hull, 1943b, p. 287). And these epistemological prescriptions were justified by an appeal to the philosophy of "logical positivism," "physicalism," and "operationalism" (Hull, 1943b, p. 273). Here again there is a striking parallel between S-R theory and association psychology: they rely on similar epistemological assumptions. For "physicalism" and related doctrines transpose into logical terms [NOTE: This is a Leibnitzian Feature], the thesis Hume expressed in psychological terms when he held that ideas must be derived from impressions; this now becomes the thesis that meaningful terms must be explicable in terms of (objective) experience, either directly, by pointing to something that can be observed, or indirectly, by using other words that can be explicated in this way (Mischel 1969, p. 30).
On the design and execution side, the Lockean basis of psychological experiments is most readily apparent. The majority of experiments have been predicated on the presentation of simple, isolateable, sensory qualities as stimuli, e.g., dots, figures, patches, colors, sounds, etc. (Garner 1970), with simple physical reactions as responses. A classic example of simple Leibnitzian stimuli in psychology are the nonsense syllables of Ebbinghaus (1913). Although Ebbinghaus's experiments were not exclusively Leibnitzian (there are also strong Lockean features), as simple symbolic entities, his stimuli would be classified as Leibnitzian. On the other hand, the responses required of subjects were mixtures of Leibnitzian (symbolic) and Lockean (experiential) components. The Leibnitzian component of the responses was again the symbolic nature of the nonsense syllables—i.e., the fact the subjects were required to respond to symbols with symbols; the Lockean component was the fact the subject had to learn "through trial and error" (i.e., experience) certain "associations" that were not a natural, formal property of the stimuli themselves.

More modern and complex Leibnitzian analyses of S/R can be found in the work of Garner (1962). His work is not only interesting but extremely important, because it represents an attempt to give a formal (information-theoretic) analysis of how systems of S's and R's relate to themselves and to one another.

One of the most interesting and important examples of Kantian inquiry in the behavioral sciences is reflected in the work of Newell and Simon (1963). Their General Problem Solver (GPS) is a prime example of Kantian inquiry in the sense that the GPS explicitly embodies several distinct means-ends strategies (models) for proving theorems in the sentential calculus. We also cite the GPS for another reason. The notion of a Kantian inquiry should not be confused with the notion of a Kantian experiment or that of a Kantian stimulus. In a Kantian inquiry, multiple models are applied to the design and analysis of an experiment. This means the stimuli or responses in that experiment will necessarily be Kantian in character. In order for an experiment to be Kantian, either the S's or the R's must be Kantian themselves. This means more than the fact that there must be multiple stimuli or multiple potential responses made available to a subject. In a Kantian experiment, at least two of the stimuli and/or the available responses must be designed under alternate conceptualizations (e.g., alternate Schools of Psychology) of the experiment. (If the alternate conceptualizations are completely antithetical, i.e., contrary, we have Hegelian stimuli and/or responses. For further details and elaboration of this point, see Mitroff 1971.) Thus, while we would not contend that the behavioral sciences have been bereft of Kantian inquiry, we would contend that there has been little, if any, systematic (and especially, sophisticated) use or development of Kantian experimental design. The situation is even worse for Hegelian and Singerian experimental design. In short, because of our predominant emphases on simple Leibnitzian and Lockean stimuli and responses, we have not developed the capability for experimenting with more complex S's and R's, especially S's and R's as complex as entire systems. As a result, many of our experiments must be regarded as deficient in external validity (Campbell and Stanley 1969) since we can not generalize from the simple S's and R's presented in most experiments to the complex S's and R's we face in everyday life. Indeed, in everyday life, the S's and R's we face are exceedingly complex. We rarely encounter the simple physical entities of the laboratory, especially in isolation from other entities or systems. More often than not, the most persistent stimuli that we are continually forced to respond to are complex entire systems of linguistic and conceptual stimuli which demand equally complex patterns, i.e., systems, as responses.

It is for this reason that we have developed the notion of Inquiring Systems (IS) as generalized stimuli and responses. Indeed, if our formulation is correct, IS would represent the most generally conceivable kinds of S's and R's available for experimental design. In their most complex and sophisticated form they would appear most fruitful for those classes of experiments particularly concerned with complex decision-making behavior. It is with this end in mind that the notion of IS as stimuli and responses has been developed.

In Table 1, we summarize briefly the essential defining characteristics of each of these IS both as stimuli and as responses. It is extremely important to emphasize that the summary represents the most complex and complete form that each of these IS as S's and R's can take. Not every experiment will or need embody every defining characteristic or feature of these IS.

5. A TAXONOMY OF IS EXPERIMENTS

Combining S's and R's in all possible ways results in 5 x 5 or 25 archetypal kinds of potential experiments. It is our contention that to date the behavioral sciences have systematically explored but a very small subset (mainly, the upper left-hand corner of Leibnitzian, Lockean S-R combinations in Table 2) of the total kinds of experiments that are generally available. In Table 2, we indicate briefly the kinds of problem situations for which the IS experimental combinations would seem most appropriate for studying.

A brief elaboration of a few of the cells in Table 2 is in order. The cells (Leibnitzian, Kantian) and (Lockean, Kantian) are labelled Idea Generation, Brainstorming, because they involve the transformation of a single system (formal or experiential) into a set of multiple systems.
### ANALYSIS AND DESIGN OF STIMULI, RESPONSES, AND EXPERIMENTS

#### Table 1

<table>
<thead>
<tr>
<th>IS</th>
<th>As STIMULI—a Subject is presented with:</th>
<th>As RESPONSES—a Subject is required to respond with or to produce:</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Leibnitzian</td>
<td>1. a set of primitive (undefined) formal elements (e.g., symbols); 2. a set of explicit, formal rules or operators for forming; 3. a model or a single set of complex relations (propositions) that can be explicitly shown to follow from 1. by means of 2.</td>
<td></td>
</tr>
<tr>
<td>II. Lockean</td>
<td>1. a set of primitive (undefined) experiential elements (e.g., sensory qualities, elementary observations, or &quot;raw data&quot;); 2. a set of agreement producing experiential operators (e.g., individual perception, expert judgment, group decision-making, T-groups, etc.) for forming; 3. a consensual position or a single set (linkage) of &quot;factual&quot; propositions that can be explicitly shown to follow from 1. by means of 2.</td>
<td></td>
</tr>
<tr>
<td>III. Kantian</td>
<td>1. at least two alternate (i.e., &quot;somewhat&quot; complimentary, but potentially divergent) sets of Leibnitzian elements and operators; 2. a set (s) of Lockean primitive, experiential elements such that... 3. when each of the Leibnitzian element and operator sets is applied to the Lockean element set(s), a set of multiple Lockean-Leibnitzian (i.e., integrative) models (interpretations) is produced such that... 4. the interdependence between the Leibnitzian and the Lockean elements is explicitly demonstrated and the multiple interpretations are made available for explicit consideration (i.e., choice).</td>
<td></td>
</tr>
<tr>
<td>IV. Hegelian or Dialectical</td>
<td>1. at least two completely antithetical (i.e., strictly contrary) sets of Leibnitzian elements (i.e., differing fundamental assumptions) and operators such that when they are applied to... 2. a common Lockean set of primitive, experiential elements (i.e., a data set),... 3. the underlying Leibnitzian model assumptions are brought up to the surface for a decision-maker's conscious inspection for the purpose of... 4. allowing (aiding) the decision-maker to form his own model assumptions that are a synthesis of the antithetical Leibnitzian model assumptions.</td>
<td></td>
</tr>
<tr>
<td>V. Singerian-Churchmanian</td>
<td>1. a set of Leibnitzian, Lockean, Kantian &amp; Hegelian models such that... 2. any model may be recursively applied to any other model (including itself) for the purpose of... 3. elucidating the distinctive features (characteristic) assumptions underlying each model and the entire inquiry process itself so that... 4. the more nearly strictly scientific (technical) features of inquiry (Leibnitzian, Lockean, Kantian) can be integrated with the ethical features of inquiry (Hegelian/Churchmanian).</td>
<td></td>
</tr>
</tbody>
</table>

(Kantian). The cells (Leibnitzian, Hegelian), (Lockean, Hegelian), and (Kantian, Hegelian) we labelled Conflict Generation, because they either involve the transformation of single systems (Leibnitzian, Lockean) or a set of complementary multiple systems (Kantian) into a set of strongly conflicting systems (Hegelian). The cells (Hegelian, Leibnitzian) and (Hegelian, Lockean) are labelled Conflict Resolution because they involve the transformation of a set of strongly conflicting systems (Hegelian) into a single, well-defined system (Leibnitzian or Lockean). The (Hegelian, Kantian) cell is labelled Conflict Management because it involves the attempt to manage conflict by giving a Hegelian stimulus a number of alternate (but not strongly conflicting) interpretations so that the

9. Note that in the typical experiment it is only this one characteristic of the entire set of IS features that usually constitutes the stimulus. Note that under our formulation, there exist enumerable other ways by which to constitute (form) a stimulus or a response. For example, any or all of the four features can constitute the stimulus and/or response. Thus, the total number of ways that one can form stimuli and responses is \( \frac{4}{!} + \frac{4}{1} + \frac{4}{1} + \frac{4}{1} = 2^4 - 1 = 15 \).

10. Thus the stimuli are already somewhat more complex to begin with than in the pure Leibnitzian case although the elements and operators can be extremely simple in any particular experimental design.

11. The purpose of this last step is to allow a decision-maker to examine consciously the ethical (and/or aesthetic) aspects (consequences) of his policies by subjecting them to an intensive dialectical cross-examination by opposing moral or ethical viewpoints. The point is that if agreement is the decision-rule of ordinary or in Kuhn's (1962) terms, "normal" science, then in Churchman's terms, disagreement is the decision-rule of ethics. That is, since the one thing men do not universally agree on is what "ought to be," if we are ever to have a "science of ethics" (Churchman, 1961), it will have to be based on a process of decision-making that explicitly allows for conflict and debate. Thus, in order to derive the "ethical" implications of any technical or scientific model, we explicitly incorporate a dialectical mode of examining (or testing) models. This constitutes a minimal condition for ethical evaluation. It obviously does not insure that the evaluation will be "sufficient." Indeed, there may be no known "sufficient" conditions at this time.
### Table 2

**A TAXONOMY OF IS EXPERIMENTAL PROBLEM (DECISION-MAKING) SITUATIONS**

<table>
<thead>
<tr>
<th>IS as STIMULI</th>
<th>IS as RESPONSES</th>
<th>Leibnitzian</th>
<th>Lockeian</th>
<th>Kantian</th>
<th>Hegelian</th>
<th>Singerian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leibnitzian</td>
<td>Traditional Experimental Design: The Formal Testing of Well-Defined Experimental Hypotheses</td>
<td>Idea Generation: Production of Possibilities; Hypothesis—Production; Brainstorming</td>
<td></td>
<td></td>
<td>Conflict</td>
<td>Derivation of the Ethical Implications of Scientific/ Technical Ideas (Models)</td>
</tr>
<tr>
<td>Lockeian</td>
<td>Detailed Development and/or Analysis of Multiple Ideas/Hypotheses</td>
<td>Alternate Futures: Elaborative Idea Generation: Elaborative Production of Possibilities</td>
<td></td>
<td></td>
<td></td>
<td>Ethical Implications of Futuristic Planning</td>
</tr>
<tr>
<td>Kantian</td>
<td>Conflict Resolution and/or Suppression</td>
<td>Conflict Management</td>
<td></td>
<td></td>
<td></td>
<td>Conflictual Futuristic Planning</td>
</tr>
<tr>
<td>Hegelian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A decision-maker or subject of the experiment can choose between the interpretations, i.e., he can choose the one “most suitable” for his problem.

We hope in future papers to report on experiments currently underway utilizing these S's and R's in complex decision-making tasks. As preparation for these experiments, we are first finding it necessary to train our subjects in the handling of these complex S's and R's. While subjects deal with exceedingly complex S's and R's in everyday life, they do not necessarily deal with them in as pure and as explicit a form as our S's and R's present them. Thus, for example, we are finding it necessary to give training in the handling of Hegelian S's and R's since in our culture people do not normally think in an explicit Hegelian mode. The experimental question we are pursuing is, under which conditions are conflict resolution, generation, and perpetuation hindered and/or aided by explicit training in Hegelian S/R handling (where one of the experimental conditions is differential training in the handling of these complex S/R patterns)?

### 6. CONCLUDING REMARKS

The tone and spirit of this paper has been explicitly philosophical. This is because the problems associated with adequately defining the terms "stimulus" and "response" are fundamentally con-

---

12. It should be emphasized that research on Hegelian IS is very different from previous research that has been done on conflict, persuasion, and attitude change. In previous research (Schunk and Goodman 1938; Sears, et al., 1965) where debates have been used to effect attitude change, both sides could introduce whatever arguments and data they wished to support their position. In a Dialectic debate, both sides are constrained to the same data set in order to make the differences between the two sides even stronger and sharper. Likewise, their positions are laid out in the form of side by side contrary propositions (Mason 1969, Mitroff 1971). In previous research, the effects of the arguments and data sets are confounded. Thus, we have no way of knowing whether it was the different data sets or the arguments that did the persuading.
ceptual, not technical. In this sense, they cannot be broached without recourse to philosophy at some point.

The purpose of this paper has been to show that philosophical analysis acting in concert with modern systems theory is of direct concern both to the philosopher and to the behavioral scientist; i.e., that both philosophical analysis and systems theory can be used to give a new perspective on the difficult problem of conceptualizing behavioral science experiments.

In his paper, Gibson (1960) raises a number of fundamental questions regarding the nature of the stimulus. It should be obvious from a reading of this paper that there are as many responses to Gibson's questions as there are kinds of Inquiring Systems and types of experiments based upon them. In the long run, the most important consequence of this exercise may not be that it provides us with a single, satisfactory answer for every important question that we can raise regarding the nature of S's and R's but that it provides us with an (not the) interesting way for posing questions and for contemplating how to go about settling them. The fundamental questions of any science are too important to be settled by any one mode of inquiry. When will we realize that to raise the question of the definition of fundamental terms is to raise a policy-information question? For, in effect, Gibson's questions are policy matters (Helmer and Rescher 1959) whose "answers" affect the fundamental course and development of the behavioral sciences. It is our contention that we need to broaden the ways that we have typically used (i.e., Leibnitzian, Lockean) to settle policy questions in the sciences. We need to inquire into the nature of inquiry. This is what the concept of Inquiring Systems is all about (Churchman 1971).

REFERENCES


Gupta, Shiv K., and MINAS, J.


--- "On Whole Systems," Internal Working Paper #31 (May 1965b), Space Sciences Laboratory, University of California, Berkeley.


Mason, Richard O. "A Dialectical Approach to Strategic Planning," Management Sci., 15, 8
IAN I. MITROFF AND FRANCISCO SAGASTI


MITROFF, IAN I. and BLANKENSHIP, VAUGHN. "On the Methodology of the Holistic Experiment: An Approach to the Conceptualization of Large-Scale Social Experiments." J. Technological Forecasting (forthcoming).


E.

STRESS IN ORGANIZATIONS
The concept of stress has received a great deal of attention in recent years among psychologists, leading to a large number of theoretical and experimental articles. This fast-growing literature was last reviewed several years ago (Lazarus, Deese, & Osler, 1952), although a few reviews emphasizing the applied aspects of this concept have appeared in military reports (Katchmar, 1954; Holtzman & Bitterman, 1952; Harris, Mackie, & Wilson, 1956; Miller, Bouthilet, & Eldridge, 1953).

The primary problem in compiling a review of the literature on stress is that this concept has been used with so many different meanings that there no longer seems to be one thing called "stress" which can be the subject of an over-all review. And not only is stress defined in various ways but it also repeatedly appears in articles without any mention of what the author means by that term. This confusing state of affairs leaves the reviewer the choice of either considering all experiments which are called investigations of stress by their authors, or considering only those experiments which fit the reviewer's definition of stress. Following the second choice may bring some order to this mass of literature but it may be an order based on a definition unacceptable to most psychologists. However, an attempt to organize this review without benefit of some specific definition seems so unwieldy that the second choice will be followed by this reviewer.

Since a great deal depends on selecting a good definition of stress, it might be profitable to consider first the terminology and definitions of stress prevalent in those fields where it has been productively studied, i.e., physics and physiology. Moreover, since psychologists have been influenced, either explicitly or implicitly, by the theories on stress of physicists and physiologists, a clearer understanding of these concepts would be valuable.

BASIC TERMINOLOGY

Physicists are in good agreement on their basic stress terminology: "stress" is the force applied to an object and "strain" is the resultant change in the object's dimensions. Since stress can be represented as small forces through the cross sectional area of an object, it is said to be in a "state of stress" (Sears, 1950). However, some physicists are reluctant to refer to "states" of an object and avoid using this latter term.

Unfortunately physiologists, like psychologists, have failed to agree on their use of basic terms. Selye has proposed a set of terms which has been accepted by many other physiologists and even won the approval of some psychologists. He applies the term "stress" to a state of the organism, "stressors" to stress producing agents, and "stress reactions" to those which are characteristic of a stressed organism (Selye, 1956). On the other hand, other physiologists, and psychologists, apply the term stress to agents (stimulus situations) and still others speak of it as a response. Since decisions on basic terminology are so arbitrarily made, this failure to agree on what general class of variables to call stress is unfortunate. Without better agreement at least at this level, effective communication between researchers on stress is severely handicapped.

While it might be preferable to fall in line with the earlier terminology of physics, Selye's terms seem to be in a more appropriate form for studying the living organism under stress and will be used in this review. Therefore, the term "stressor" will be applied to stimuli eliciting stress, "stress" to a state of the organism, and "stress responses" to those responses which are characteristic of or reflections of this state of stress. The term "strain" could be easily substituted for "stress responses" but psychologists are accustomed to identifying a response with an adjective which also describes the organism's drive or state, e.g., fear response, hunger response.

OPERATIONAL DEFINITION OF PHYSICAL STRESS

The physicists define stress as the force applied to an object and strain as the resultant change in dimensions occurring in an object under stress (Sears, 1950). At first glance, a valid operational measure of stress seems to be a change in an object's dimensions. But not all changes in dimensions indicate an object has been stressed, as heating a nonconstrained object will change its dimensions without stressing...
it. Conversely, an object can be stressed without a change in dimensions, as heating a constrained object will stress it without changing its dimensions. Therefore, strain is at best only a rough indicator of possible stress.

Since an object is stressed when subjected to a force, stress could be defined by an adequate operational definition of a force. A force is defined as either a push or pull accelerating an object or as energy stored in an object. With a definition operationally grounded on these measurable effects, acceleration or energy storage, the productive study of stress by physicists has been greatly facilitated. Arguing by analogy, psychologists would similarly benefit from a definition based on measurable dependent variables by which stress can be recognized.

OPERATIONAL DEFINITIONS OF PHYSIOLOGICAL STRESS

General definition

The general physiological concept of stress is derived primarily from theoretical notions of Bernard and Cannon. This concept is that homeostatic mechanisms maintain within a limited range factors in the internal environment which are crucial to the life processes of an organism. An organism is stressed when its homeostatic mechanisms are unable to maintain these factors in a state of dynamic equilibrium.

The question is, how can this concept best be put into suitable operational terms? One possibility is to define stress by any change in the internal environment, a definition somewhat analogous to accepting strain in a definition of physical stress. The difficulty is that since a living organism is subjected almost constantly to conditions which alter these variables, the organism would be always in a state of stress, with homeostatic mechanisms reducing, but not preventing, stress.

To avoid stretching the concept to this extent, the "limited range" or "dynamic equilibrium" for these variables can be emphasized, with displacement beyond that range indicative of stress. This places the definitional load on determining "normal" ranges and necessitates measuring these variables under diverse conditions to specify values for their normal limits on an actuarial basis. Certainly such a definition lacks precision, and such problems as species, population, and individual differences, as well as adaptive changes for a single individual across time, complicate attempts to make this definition more exact. Nevertheless, it does give the physiologist some measurable dependent variables which allow him to recognize the presence of stress with a statistically specified degree of accuracy.

Selye's definition

Selye (1956) has defined his concept of stress in still another way, in terms of a few specified reactions of the organism. Selye selected these reactions empirically by noting that after various noxious agents were injected into an organism, these reactions took place: (1) enlargement of the adrenal cortex; (2) atrophy of the thymus; and (3) ulcers in the gastrointestinal tract. Selye concluded that these tissue damages and functional changes were part of the organism's general adaptive response to stress. Stress was therefore defined as the state manifested by these reactions, stressors as the agents eliciting stress, and stress reactions as the reactions (General Adaptation Syndrome) of an organism in a state of stress.

The stress reactions specified by Selye do provide him with a suitable operational definition of stress which has generated many productive experiments. He and his colleagues have demonstrated the stressful nature of many substances and shown how to combat these stressors with such agents as hormones. They have also investigated the time course of reactions following the administration of a stressor, separating them into: (1) alarm reaction; (2) stage of resistance; and (3) stage of exhaustion. It is in this time course of events that Selye has run into operational difficulties. As the reactions to stressors have been plotted backwards in time, a point is reached where, for example, the size of the adrenal cortex does not differ significantly from normal. Since the reactions specified in his definitions are not present, can it be said that the organism is in a state of stress at that time?

In his most recent book (1956), Selye has preferred to handle this problem by stating that during these earlier stages, stress is manifested by other reactions which are being determined by further experimentation. However, if these studies reveal that early reactions to a stressor are like those to a non-stressor, then Selye's concept might be broadened to include a stage prior to the alarm reaction. This period would be characterized by reactions which are local and peculiar to each physiologically active stimulus. During this stage the organism would be in a state of dynamic equilibrium and not in a state of stress. Only when a stimulus of a certain quality and intensity is applied for a sufficient length of time does it become a stressor, causing a disturbance in the internal environment great enough to precipitate the second or alarm stage. This alarm stage would be characterized by reactions involving the total organism, reactions which would be similar for all stressors. The inclusion of a non-stressed stage of dynamic equilibrium would naturally bring Selye's concept more closely into agreement with the
physiological concept of stress previously discussed.

PSYCHOLOGICAL STRESS

If Selye's or the more general physiological definitions are satisfactory definitions of stress in the living organism, is either also acceptable as a definition for psychologists? One possibility is to assume that an organism physiologically stressed is also psychologically stressed, and then proceed to discover which psychological variables produce these physiological reactions and what psychological responses accompany this state. This assumption seems to be implicit in those psychological experiments which use physiologically defined stressors to induce stress in their subjects, e.g., anoxia, hypothermia. Certainly the probability is high that there is a relationship between psychological and physiological stress, especially in cases where the site of action of the physiological stressor is the central nervous system. However, if the extent of this relationship is to be experimentally verified, rather than accepted as a basic assumption, this necessitates a definition of stress in terms of psychological variables. And, with non-physiologically oriented psychologists, the need for a separate definition of psychological stress is even more obvious. To underscore the fact that stress is defined in terms peculiar to each field, the expressions "physical stress," "physiological stress" and "psychological stress" will be used in the remainder of this review to indicate what general types of variables were the determinants of a given state of stress. The desirability of such a differentiation has been pointed out by several investigators of stress (Pascal, 1951; Lazarus et al., 1952).

Stress research by psychologists has suffered more from a lack of definitions than has research by physicists and physiologists. Stress has become a catch-all term applied to research varying from the simplest stimulus, e.g., a buzzer, to situations in which the individual's life is threatened, e.g., a combat mission. Perhaps there is no need to define psychological stress when a situation is used which seems obviously stressful to most individuals, such as a combat mission. However, when the researcher attempts to set up mild experimental facsimiles of these situations, their stressful aspects may be highly debatable. Unless all stimuli are to be considered stress-inducing forces, a definition of psychological stress is needed to see what the author means by stress and to judge how successfully he has operationalized his definition.

DEFINITIONS OF PSYCHOLOGICAL STRESS

Current definitions of psychological stress may be classified into two general categories: (1) definitions emphasizing the stimulus situations to which the individual is subjected, e.g., anxiety or frustration-inducing situations; (2) definitions emphasizing both the stimulus situation and the occurrence of responses or changes in the subject's responses. There are further definitional differences depending on whether stress is considered a state, drive, stimulus, or response, but these terminological variations will not be considered.

Definitions based on stimulus situations

Definitions which fall under the first category may specify different stimulus conditions at the definitional level, but they all seem to be derived from a similar concept of psychological stress. The individual is conceptionalized as responding to changes in the environment so as to maintain equilibrium in his relationships with his environment. The individual is stressed when some aspect of this relationship is threatened: a threat to his self-esteem (Glixman, 1949); to a balance between needs and satisfactions (Fox, 1955); to attainment of some goal (Lazarus et al., 1952); or to his integrity (Basowitz, Persky, Korchin, & Grinker, 1955). Stated in these general terms any of these threats can be made to subsume all the others. It is when these threats are operationalized, e.g., anxiety for threat to integrity (Basowitz et al., 1955) or frustration for the threat to attainment of a goal (Lazarus et al., 1952), that differences become more apparent. While there are similarities between anxiety and frustration, distinctions are made between both the stimuli arousing either of these states and between their characteristic responses. Therefore, at the operational level, these theorists seem to be defining different kinds of stress, rather than a unified concept of stress.

Since these definitions are restricted to one type of stressor, they fail to suggest other kinds of potential stressors to the research planner. Perhaps their greatest inadequacy is that they do not indicate a relationship to similarly restricted definitions. Without a more unified definition, experimental results will tend to remain as unrelated bits of information, rather than serving as empirical supports for a growing concept of psychological stress.

A second criticism of these definitions is that they are not stated in terms which can be operationalized with any precision. They are not specific about the exact stimulus conditions which will induce stress, and of course they don't mention any response by which the presence of psychological stress can be demonstrated. For
example, the threat to attainment of a goal is handled operationally by subjecting an individual to failure on a task. There is no criterion for testing the assumption that successful completion of that task was a goal, with failure a threat, and the individual in a state of stress. Lazarus et al. (1952) discuss this problem and conclude that there is no adequate method of testing these assumptions. Basowitz et al. (1955) come to the same conclusion but go further in stating that ultimately stress situations will be defined by the occurrence of a given response.

Can these definitions lead to productive research on psychological stress unless they include a measurable response as a criterion variable? We have already discussed the extent to which physicists and physiologists have operationally based their definitions of stress on variables which are analogous to behavioral dependent variables. And in psychology most intervening variables have had to be grounded on the response side by an appropriate response before high agreement could be reached as to which experiments were interpretable as investigations of a given intervening variable. For example, if the experimentalist is interested in hunger he will naturally try to set up conditions which will arouse hunger in his subject, such as food deprivation. He can then demonstrate this hypothetical state is present under the conditions he selected by letting his subject make food-oriented responses. After the subject's behavior has shown he is hungry the experimenter can go on to find out about the relationship of this intervening variable to other experimental conditions. Unless the presence of an intervening variable is exposed to the litmus paper test of an appropriate behavioral response by the subject, shouldn't an investigator limit his conclusions to the relationship of his specific independent variables to the dependent variables, rather than attributing the changes in behavior to a shadowy intervening variable?

Some psychologists have maintained that psychological stress is a special intervening variable since it can be induced in a subject without affecting his behavior. The stressed individual is said to mask the presence of stress by such factors as increased attention, increased motivation, or some adaptive defense mechanism. Therefore it would be useless to look for a regular response to stress since none may occur. However, this viewpoint should remain an unsupported hypothesis until experiments can be designed to tease apart these variables of stress and reactions against stress to show their individual, rather than combined, effects on behavior. It is scanty evidence when an effect and a reaction against an effect are both hidden from the observer.

The literature on psychological stress is filled with conflicting conclusions about the effects of "psychological stress" on the same behavioral variable. Some of these results can be attributed to the psychologist's nemesis, individual differences, this time in responsivity to stressors. But many of these conflicting conclusions are due to the failure of the experimental conditions to induce stress in the subjects; therefore the experimenter's conclusions on the effects of psychological stress are invalid and misleading to his fellow researchers. Granted that there is a need for a criterion behavioral variable in a definition of psychological stress, is any such variable presently available? Or is our present knowledge so inadequate that a selection of an appropriate response would be premature? This naturally leads us to a consideration of the second category of definitions.

Definitions based on both stimuli and responses

These emphasize both the experimental conditions likely to induce stress and the behavioral responses which reflect this state. Freeman (1945) defines psychological stress as a threat to the integrity of the total organism which is manifested in physiological measures of bodily arousal. Patton (1955) defines psychological stress as a threat which causes behavior to oscillate beyond its normal bounds of variation and return slowly to its mean. Pascal (1951) sees it as a threat to gratification of needs, measurable in terms of functional deficit. Schaffer (1954) defines stress as occurring when an organism is unable to make an adaptive response to a disruption in its relation to its environment. This state is reflected in the rate and range of general activity, by altered rates of acquisition and extinction of learned responses, and by fixation on a dominant response.

The experimental conditions suggested by these theorists are similar to those proposed by theorists who define psychological stress only by the stimulus situations likely to induce stress. As to the responses specified in these definitions or suggested by experiments on stress, they may be placed into the following general categories: (1) physiological responses; (2) defense responses; (3) subjective reactions; (4) changes in goal-oriented behavior; and (5) decrements in goal-oriented behavior. As in the case of Selye's specific reactions, the fruitfulness and acceptability of any of these responses depends on the theorist's ability to pick out the crucial response variables across a variety of stressful situations.

It should also be mentioned that in this search for an appropriate response, some psychologists have also defined stress by more than one response. For example, as criteria for determining that a laboratory situation did induce...
stress, Funkenstein, King, and Drolette (1953) proposed subjective reports of emotional changes, physiological reactions, and decrements in task performance.

Physiological responses. Taking up these responses in order, physiological responses are selected usually as either indicators of the functioning of homeostatic mechanisms or of emotional reactions. The interest in homeostatic mechanisms is based on the assumed relationship between physiological and psychological stress. The rationale for the use of physiological measures of emotional reactions is that psychological stress is often equated with the emotional states of fear, anxiety, frustration. The main difficulties with this class of variables are the technical problems in measuring and analyzing physiological reactions as correlates of psychological variables, as years of research on fear and anxiety correlates have pointed out. Add to this the reluctance of some psychologists to use or at least lean that heavily on physiological measures, and it is apparent that a definition based solely on this type of response will not be generally acceptable.

Defense responses. Defense mechanisms are important indicators of psychological stress for clinicians and psychiatrists who conceptualize the individual as reacting to stress-inducing threats by means of defense responses. The problem is to define those characteristics of defense mechanisms by which psychological stress is to be recognized. Accepting the occurrence of any defense response as an indicator of stress would make the definition too general. Another possibility is to recognize a normal range of adaptive defense mechanisms for a non-stressed individual and require an increase in intensity of these responses or a change to maladaptive responses as evidence the individual was stressed. A related procedure is to accept the defense mechanisms characteristic of psychotic patients as also characteristic of any individual under stress. It is not possible for this reviewer to evaluate effectively the validity of this class of response variables. While realizing the need for clinicians to use variables of importance in their area, the complex behavioral patterns which constitute defense mechanisms are subject to such varied interpretation that they are not precise enough to serve as a satisfactory criterion for stress.

Subjective reports. The third type of response by which psychological stress is definable is subjective reports. These reports are designed to tap the distressful, emotional experience of a subject, either by questioning the subject or by subjective observations of his general behavior. This type of response is open to the usual objections against its lack of objectivity and quantification, so that it seems best suited as a rough indicator of psychological stress for preliminary investigations.

Changes in goal-oriented behavior. The fourth type of response by which psychological stress is definable is similar to that by which many interviewing variables are defined—either an increment or decrement in goal-oriented behavior. Unfortunately, stress is not like such intervening variables as thirst, for which stimulus conditions can be specified as constituting a related goal (water), with behavior oriented towards this goal an indicator of this hypothetical state. Instead of being related to one kind of motivation and therefore one kind of goal, psychological stress may be elicited whenever an organism is trying to reach any goal which will satisfy its psychological needs. The defining criterion is that the attainment of this goal and thus the satisfaction of the need be threatened.

However, it is possible to maintain that there is always a minor obstacle and thus a threat to the attainment of its goal by any motivated organism, even if it is just a long runway for an unhappy rat to traverse or a heavy lever for him to press. Therefore all motivated organisms could be considered in a state of stress and psychologists would merely have to choose between using either the term "motivation" or its equivalent "stress."

If psychological stress is not to become synonymous with motivation, it is necessary to differentiate between the changes in goal-oriented behavior of a stressed and of a non-stressed motivated organism. Perhaps one could be more specific about the added motivation (e.g. anxiety, frustration, fear) induced by the threat to reaching a goal. Some behavioral or physiological measure of these motivational states could be selected to show that the experimental situation constitutes such a threat. Thus the total definition would contain two behavioral criteria of psychological stress: a change in goal-oriented behavior, as compared to control conditions; and the occurrence of an emotional response, indicating that attainment of the goal is threatened.

There are two criticisms which I would like to make of such a definition: one practical and the other theoretical. On the practical side, there is the problem of selecting good measures of emotional responses, be they subjective reports, behavioral observations, or physiological reactions. The theoretical criticism is that a definition based on these two behavioral criteria is not adequate to operationalize a more extreme concept of the psychologically stressed organism. This concept emphasizes the maladaptive, disorganized behavior of a stressed organism, as differentiated from the adaptive, organized behavior of a non-stressed organism. This
adaptive behavior may occur even when the non-stressed organism is subjected to "threatening" conditions which arouse an emotional reaction.

Decrement in goal-oriented behavior. This concept is better operationalized by the fourth and last type of response variable, a decrement in goal-oriented behavior. It is the thesis of this reviewer that this variable provides the best criterion by which psychological stress can be recognized for several reasons. First of all, it seems adequate for operationalizing our concept of stress derived from observations of an organism's behavior in non-laboratory, stressful situations. Confused, disorganized, inappropriate behavior leads us to call a given situation a stressful one. And this disorganized behavior can be operationally defined in an experimental situation as decrements in goal-oriented behavior.

Secondly, emphasis on this type of behavioral criterion suggests a relationship of psychological stress to levels of motivation rather than to certain types of motivation. Thus a high level of any motive state may be just as much a source of psychological stress as external conditions which threaten the attainment of a goal for lesser motivated organisms. This notion is similar to the distinction Leeper (1948) made between high and low levels of fear. He held that high levels of fear produce disorganized behavior but that low levels can facilitate organized behavior. Hebb's theoretical notion that high levels of cortical arousal contribute to disturbed behavior while normal levels contribute to organized behavior, suggests a neuro-physiological mechanism which may play a role in psychological stress (1955).

Finally, this measure provides a fairly rigorous test to insure that psychological stress has been induced in the experimental subject. It is possible that a disruption or decrement in behavior may occur during later stages of stress, and that other responses may be characteristic of its earlier stages. Therefore, stress will not be recognized in experiments which tap only these earlier manifestations of stress. However, it might be preferable in preliminary investigations of an intervening variable to be extremely rigorous, determining stressors and their effects only in cases where stress definitely has been induced. Once the stressors and their important characteristics have been identified, it would be simple to proceed as Selye has done in his later work, investigating the earlier aspects of this state.

On the other hand, this response criterion could be criticized for not being definitive enough to exclude cases in which the organisms show a decrement in goal-oriented performance without being in a state of psychological stress. Some conditions, such as decreases in motivation or in incentive, also lead to drops in performance while the subject is not necessarily stressed. Perhaps an erroneous labeling of these conditions as stressors can be avoided if other measures of behavior are also used which satisfactorily test the disrupted, agitated behavior of a stressed individual. For example, errors in response, blocks or bursts of responses, or emotional responses should characteristically accompany the drop in performance.

PROPOSED DEFINITION OF PSYCHOLOGICAL STRESS

What is the concept and definition of psychological stress which has developed out of all these considerations? It is that psychological stress is a state which occurs when an individual is subjected to conditions which disturb or threaten to disturb crucial psychological variables from within their normal limits. These variables are of two general types: (1) those variables whose displacement from normal constitute a threat to the continued existence of the individual, thereby arousing high levels of biological drives or fear; (2) those variables whose displacement from normal constitute a threat to some aspect of his personality structure or ego, thereby arousing anxiety, failure, frustration. When either of these conditions occurs, the individual is in a state of psychological stress, the stimulus conditions arousing this state are psychological stressors, and the responses characteristic of this state are stress responses. The behavioral manifestations of stress are disorganized, disrupted, emotionally distressed responses. The operational definition of these stress responses is an impairment of psychological function or a decrement in performance on a goal-oriented task. Measures of this decrement in performance should be supplemented by other measures of the disrupting effects of stress, such as increased errors, bursts of responses, or emotional reactions.

The above reference to "crucial" psychological variables, an obvious derivation from the concept of physiological stress, is stated in rather vague terms. The further accumulation of experimental evidence should enable us to become more specific about the characteristics of these crucial variables, much as the physiologist has been able to delimit crucial internal variables, such as pH, CO₂ and O₂ levels. However, the comparative simplicity, limited number, and limited range of these variables in an organism's internal environment have permitted the physiologist to pin down this aspect of the concept much more easily than the psychologist can hope to for some years to come.
Therefore, an operational definition of psychological stress must presently be based on measures of impairment of function, from which the limits of a crucial psychological variable were exceeded and the individual was psychologically stressed. This definition is somewhat related to Selye's selection of specific kinds of responses by which stress can be recognized. It suffers from the same advantage of providing an operational measure by which fruitful preliminary investigations may proceed.

Advantages of this approach

What are the advantages of this approach to the above definition? First of all, the definition is based on an examination of the concepts underlying the many diverse definitions in the psychological literature, and attempts to fit these conceptually related definitions. In addition, a consideration of the definitions which have been productively used in physics and physiology emphasized the need for the inclusion of an operational measure to test for the presence of stress, and such a measure was proposed. Finally, some relationships between physiological and psychological stress were pointed out. Perhaps an understanding of these relationships may enable psychologists to make more fruitful generalizations from the findings of physiological investigations concerning characteristics of stressors, stress, and stress responses.

Theoretical considerations

The question might also be asked at this point, of what importance is the study of psychological stress? Theoretically, as well as practically, the study of stress shows promise for psychologists. When a psychologist builds a theoretical model of the individual, or the social psychologist of a group of individuals, the model builder is faced with the problem of representing the complexity of his system with a minimum of variables. The more significant these variables are to the real system, the more representative his theoretical model will be. A living system involves variables which are crucial to the survival of that system and are therefore kept relatively invariant by the system. Stressing a system may be a suitable experimental tool for the theorist to use in differentiating core variables from those less important variables which vary greatly without affecting the functioning of that system.

As a speculative example, perhaps Freud was better able to select variables of great significance for a conceptual picture of the individual because he studied those who were psychologically stressed. Observations of neurotic and psychotic patients have continued to generate theories of the variables which are crucial to the development of the ego and its relationship to its environment. When a system is considered in which the conceptual units and their relationships are more abstract, such as a group of individuals in social psychological studies, the use of stress may be even more promising. The theorist would construct a model of group processes from factors which remain relatively invariant when the group is stressed, but would discard those factors which vary widely without disturbing group processes.

Practical considerations

In addition to these considerations, investigations of psychological stress can shed experimental light on a variety of practical factors. On the stimulus side, the stress-inducing properties of various conditions can be determined so that they may be avoided or altered to reduce their stressful potential. On the response side, what psychological factors or functions are most likely to reflect the characteristic impairment of stress? There is the problem of specificity of effect; is a function sensitive to a variety of stressors or differentially sensitive to various conditions? Naturally other manifestations of stress should be examined, both of the physiological and psychological varieties. It has already been discussed how measures of errors, increased rates, and sporadic response may be necessary to catch the disturbed aspect of stress in order to exclude decrements in performance which are attributed to such non-stressors as decreased motivation.

Fatigue and stress. A concept that is closely intertwined with psychological stress is that of fatigue and an important question is can the decremental effects of fatigue be differentiated from those of psychological stress. The types of responses previously suggested are also characteristic of fatigue so they won't make a satisfactory differentiation. One possibility is to emphasize that stress is a state involving the total organism, as Selye has suggested. Therefore, impairment of performance is a reflection of stress only if the experimental conditions allow one to make an inference of this type of total involvement. For example, if visual function is impaired by extensive use, intense stimulation, or even lack of stimulation, this could be considered a localized fatigue effect on that specific function. To demonstrate that more than this one system has been affected and the organism thereby stressed, tests of other
functions also should show some decrement. General systemic fatigue induced by loss of sleep is more likely to be a stressor than specific fatigue induced by physical work involving some of the somatic musculature. Even experiments on specific fatigue are of interest, however, because the information they can give on the differential sensitivity of different tests or functions to this condition may be generalizable to stress experiments.

Also there is the practical question of determining adaptive responses to potential stressors. The training of individuals in the use of these reactions to prevent stressors from impairing function would be analogous to hormone injections to combat physiological stressors.

Finally, there is the large problem of individual differences in response to potential stressors. What are the appropriate tests for measuring personality traits which would predict these differences? Again there is the question of specificity for these traits, do they measure differences in responsivity to specific stressors or to diverse kinds of stressors? And do they offer predictability of differences in impairment of a specific function or to all functions? The development of adaptive reactions is certainly pertinent to this question of individual differences. Do such experiments as those on "gentling" indicate that the amount of environmental change to which an organism can adapt is dependent on his having been subjected to controlled, gradually increasing specific and non-specific stimulation as it matures? Perhaps a psychotic individual has either been exposed to large environmental changes before he could handle them or he is physiologically handicapped by the malfunctioning of his nervous system in his ability to adapt to changes. And finally can the factors of sensitivity (lowered threshold) be separated from responsivity (over-reaction or prolonged reaction) to stressors?

These and similar questions may be answered by well-planned investigations of psychological stress. Experiments which seem to this reviewer to propose some answers to these questions will be reviewed in the succeeding section. However, at this stage in research on stress the three problems of major interest seem to this reviewer to be: the sensitivity of testing instruments to the response manifestations of stress, the sensitivity of measures of individual differences to responsivity to stressors, and the stressor potential of stimulus conditions. Possible solutions to these problems will receive primary emphasis. Finally, it should be kept in mind that the primary criterion for the selection and evaluation of experiments in this review is their success in showing a decrement in performance attributable to psychological stress.

PART II. EXPERIMENTAL RESEARCH ON PSYCHOLOGICAL STRESS

ORGANIZATION BY FUNCTIONS

Tests which measure the effects of psychological stress all involve to some extent perceptual, motor response, and cognitive or decision making functions. However, performance on a given test also may be differentially attributed to just one of these broad functional categories, by the degree to which performance depends mainly on that function. This review was organized to begin with a test rather equally loaded on all three functions, i.e., reaction time, and then move on to tests more heavily loaded with perceptual, then motor response, and finally cognitive factors. Decisions concerning the major factor loading of some tasks had to be made quite arbitrarily at times. This is especially evident in the more complex tasks, such as pursuit tracking, which are loaded with a multitude of factors but finally had to be placed arbitrarily under one of the major functional categories.

STRESS SENSITIVE FACTORS

Under each of these broad headings an attempt was made to point up the factors which may play a part in making a given test stress-sensitive. In general, it seems that a task is better able to reflect the effects of psychological stress if it successfully extends the functions contributing to performance to their limits. Such a task may thus preclude the subject's masking the disrupting effects of stress by some compensatory mechanism.

Some of these stress-sensitive factors are: (1) complexity of presented stimuli; (2) complexity of cognitive manipulations required; (3) degree of response coordination required; (4) rate of presentation of stimuli; (5) rate at which responses must be made; (6) reversal of habitual stimulus-response associations. Tests which tend to measure the limits of one or more of these factors for an individual would necessitate a high level of attention as well as broad, flexible attentive processes.
PSYCHOLOGICAL STRESS

REACTION TIME

This test emphasizes the requirement for making a fast but simple response to easily discriminable stimuli.

Simple reaction time

This involves only one response and one stimulus. The general finding concerning this test is that it is insensitive to conditions which have evoked a stress-like decrement in performance on other tests. General systemic fatigue induced by deprivation of sleep for periods even up to 100 hours has had no effect on simple reaction time in two experiments (Cooperman, Mullin & Kleitman, 1934; Edwards, 1941). An early experiment of Patrick & Gilbert (1896) did report a marked slowing and increased variability of simple reaction time after 90 hours loss of sleep, but only three subjects were tested and these results have not been replicated.

Potential physiological stressors have also had little effect on simple reaction time. Forlano, Bermack & Conkley (1948) reviewed the effects of ambient and body temperature on reaction time and concluded that temperatures between 50°F and 117°F do not change reaction time. Short duration anoxic conditions also were found to be ineffective as stressors as measured by this task (Waldfogel, Finesinger & Verzeano, 1950). Carbon monoxide further points up the relative insensitivity of simple reaction time since blood saturation of CO reached 45% before even a slight impairment resulted (Forbes, Dill, DeSilva & VanDeventer, 1937).

In a series of experiments subjects were tested on a variety of tasks under moderate and heavier doses of LSD-25, a psychomimetic drug (Abramson, Jarvik, Hirsch & Ewald, 1955a; Jarvik, Abramson & Hirsch, 1955a, 1955b; Levine, Abramson, Kaufman & Markham, 1955; Jarvik, Abramson, Hirsch Ewald, 1955c; and Abramson, Jarvik & Hirsch, 1955b). This series of experiments can be criticized on several counts; for instance, tests were always given in the same order with no control for either an order effect or for change in drug reaction across time. In addition, no non-drugged control group was included to permit a good statistical evaluation of possible drug effects. Nevertheless, within these limitations, these experiments do give some rough indications of the relative sensitivity of a variety of tests to the same potential stressor. Abramson, Jarvik and Hirsch's experiment (1955b) indicated that while the moderate dose of LSD-25 did not change reaction time, there did seem to be a tendency for the larger dose to slow both audio and visual complex reaction times while slowing only simple audio reaction time. This differential affect on simple audio versus visual reaction may be due merely to difference in intensity or attention arousing characteristics of the stimuli used, but it may indicate a differential sensitivity of those two perceptual systems to the same stressor. At least this experiment gives some support to the conclusion that more complex functions are better indicators of psychological stress.

Individual differences. The Taylor Manifest Anxiety Scale (TMAS) has been used to measure individual differences in reaction time under various conditions. When electric shock was the stimulus to which responses were given, high anxieties on the TMAS had faster reaction times and squeezed a hand dynamometer harder than low anxieties (Wenar 1954). Increasing the intensity of the shock speeded up reaction times for both groups, but it failed to produce a differential effect on these groups. Castaneda (1956) used an auditory stimulus of varying intensity and he did find at least a tendency for high anxieties to be faster to the stronger stimulus and slower to the weaker stimulus than low anxieties. A tentative conclusion might be that speed of response to different intensity stimuli is directly related to level of TMAS anxiety. However, when the meaning of the shocks was changed by giving them at different intensities after a slow response, Farber & Spence (1956) could find no relationship between changes in reaction time and level of TMAS anxiety.

One final experiment on individual differences with this technique measured the changes in reaction time of psychoneurotic patients and normals under short duration anoxic conditions (Waldfogel et al., 1950). No differential changes were discovered for these two groups.

Randomized vs. regular procedure. By using a slightly different technique, Rodnick & Shakow (1940) were able to demonstrate a difference between the reaction time of schizophrenic patients and normal controls, a difference which has persisted through succeeding experiments (Huston & Singer, 1945; Huston & Senf, 1952; Tizard & Venables, 1956). They found that schizophrenics were slower in their reaction time than normals, which isn't too surprising and could be attributed to such factors as lack of motivation or poor contact with the environment. However, Rodnick & Shakow also varied the length and regularity of the interval between a preparatory signal and the stimulus for response. When this interval is long (10-25 sec.), schizophrenics respond faster if the preparatory interval is randomly varied than if the interval is held constant over a series of trials, while normals respond slower in the randomized procedure than in a regular procedure. Only Knehr (1954) has failed to replicate these findings and he found neither a difference in reaction time between his schizophrenics and normals, nor a
faster reaction time of the schizophrenics to the longer interval in a random procedure. His first finding may indicate his schizophrenic sample differs from the others, since Huston & Serf (1952) have shown that chronic schizophrenics have slower reaction times and show the randomized effect more than early schizophrenics. Moreover, his experimental conditions were also different in that he used no intervals over 10 seconds, while the other experimenters used intervals up to 25 seconds. While they did find an effect at the 10 second interval, it may have been the interspersion of the longer intervals in a random procedure which accounts for these differences in reaction time in the random and regular procedures.

Several other variables which may play a part in this effect have been experimentally investigated. Besides finding that the results are more pronounced in chronic schizophrenics than in early schizophrenics, Huston & Senf (1952) found the effect was entirely lacking in neurotic and manic-depressive patients. Apparently the cognitive defects similar to those of mental defectives are not responsible, since a sample of this group were similar to normals in their reaction time under these procedures (Tizard & Venables, 1956). Whatever function is involved, the combination of amphetamine sulfate and sodium amytal were effective in altering the reaction time of schizophrenics to the extent that they also responded faster to all preparatory intervals when regularly presented (Huston & Singer, 1945; Huston & Senf, 1952).

These experimental findings have been attributed to failures in mental set of schizophrenics (Rodnick & Shakow, 1940; Tizard & Venable, 1956) or to their difficulty in maintaining attention (Huston & Senf, 1952). While failures in set or attention may explain why the schizophrenics respond slower than normals in a regular procedure, these do not seem to be adequate explanations of why they respond faster at long intervals in the irregular than in the regular procedure. The tendency for their attention to drift during a long wait should be the same whether the previous trial was of a different interval (random procedure) or a similar interval (regular procedure). This finding seems to be worth further investigation to determine what disturbance in a normal process accounts for the results.

Choice reaction time

The stimuli are still qualitatively simple and easily discriminable, but they are increased in number as compared to simple reaction time. The responses also remain simple in that they do not require a high degree of muscular coordination. Naturally, the additional stimuli introduce the problem of deciding how to respond as well as when to respond.

Broadbent has investigated the stressor potential of noise intensity and frequency during a choice visual reaction test of some 25 minutes duration (1957). He found that random noise should be at least 100 dbs in intensity and above 2000 cps to be an effective disrupter of choice reaction time. Since this type of noise when used as a stimulus for responding also evoked a faster reaction time than low frequency and low intensity noise, Broadbent concluded that its success as a distractor depends on its ability to compete with other stimuli to control responses in a limited capacity receptor system. Or in the phraseology which is popular among physiological psychologists, this depends on the stimuli's attention-arousing ability. This experiment seems noteworthy for both its description of the physical properties of what constitutes an auditory stressor and for the attempt to determine why it is a distracting stressor.

One of the variables which may be important in reflecting the stressor potential of sleep deprivation is duration of the measuring task. Loss of sleep up to 90 hours had no effect on a brief task (Tyler, 1947; Patrick & Gilbert, 1896), while there was a decrement in a 10-minute test (Tyler, 1947). Pepler (1959) found that after just one night's loss of sleep there was a decrement in choice visual reaction test of 20 minutes if the subject first performed on a 30 minute tracking task. He also found a drop in choice reaction time performance under high temperatures which was different from the change produced by fatigue. Under the thermal stressor, subjects responded with increased errors, while under the fatigue stressor they tended to omit responses altogether. The first seems to reduce accuracy, the second to reduce activity.

Threats of shock for slow responses had no effect on choice visual reaction time (Farber & Spence, 1955). Moderate anoxic conditions had no effect on choice visual reaction time (Waldfogel et al., 1950). Experiments with the psychomimetic drug LSD-25 indicate this test may be more sensitive to a potential stressor than simple reaction time. Both choice visual and audio reaction times seemed to be affected by heavy doses of LSD-25, while only simple audio was affected (Abramson, Jarvik & Hirsch, 1955b). However, even this test may not be as sensitive as other tests which were disrupted by lighter doses of LSD-25.

Re-pairing stimuli and responses. One variable which may be of interest in stress research with this test is the re-pairing of stimuli and responses after initial training on the original pairings. One of the characteristics of stress is similar to the effects of high motivation; an interference with the effects of high motivation; an interference with the flexibility needed to reverse learned responses while facilitating the performance of established habits. Several
experiments have indicated that under fast experimenter-paced conditions there is a tendency for subjects to make more errors on the re-paired items and fewer errors on the original pairings (Castaneda & Palermo, 1955; Castaneda, 1956; Palermo, 1957). This experimental procedure might be worthwhile for testing various stressors of interest and as a condition for showing individual differences in stress responsivity.

Individual differences. Farber & Spence (1956) found no differential effect of threat of shock for slow responses on the choice visual reaction time of high and low anxieties on the TMAS. Also, under moderate anoxic conditions, changes in choice reaction time of psychoneurotic patients were similar to those of normal controls (Waldfogel et al., 1959).

PERCEPTUAL TESTS

Tests which measure various aspects of perceptual functions (discrimination, generalization, thresholds, acuity, flicker fusion, perceptual span, etc.) are all defined as perceptual tests.

Discrimination reaction time

This task differs from choice reaction time in that discrimination of the stimulus or discrimination between stimuli is more difficult, with perceptual factors playing a larger role in performance. Quick discriminations may be encouraged by the instructions, by shocks for slow responses, or a fast rate of serial presentation of the stimuli, but the important dependent variable is accuracy of discrimination.

Vigilance. Mackworth (1950) has emphasized the factor of vigilant attention in this type of task. Vigilance may be sensitive to stressors external to the task, but it is also susceptible to fatigue, making task duration an important variable. For example, Mackworth (1950) has reported an increase in missed signals on the clock test (responding to each double jump of a clock which usually moves in single jumps) after the first hour of a two-hour test, and Bakan (1955) found an increase in discrimination threshold in a one and one-half hour test.

Flexibility of attention. Jerison has pointed up another factor of interest to stress research on this test, that of flexibility of attention (1957). He contrasts the vigilance needed to watch for one changing stimulus with the flexibility required to scan a broad field in which several stimuli may change at one time or during successive periods. His results indicate that high intensity noise was not a stressor in a vigilance task (1957), while it did disrupt performance in a scanning task (Jerison & Wing, 1957).

Warren subjected a group to a ride on a radial accelerator and reported that this novel, anxiety-arousing experience caused a decrement in discrimination reaction time, a result which he doesn't feel is due to the physiological effects of such a ride (1950).

Eriksen & Wechsler (1955) used random shocks in an attempt to induce anxiety in their subjects but were unable to show any effect of this condition on discrimination accuracy. They did find an increase in stereotyped responses and concluded from this that response thresholds were affected rather than a sensory process.

Individual differences. Stimulus generalization of psychiatric patients was compared to normals under strong shocks, weak shocks, or a buzzer for slow responses (Rosenbaum, 1953; Malm, Shagass, Belanger & Smith, 1951). Only the strong shock was effective in increasing stimulus generalization but not even this condition differentiated between the generalization of psychiatric patients and normal controls. In a replication of this experiment, Rosenbaum (1956) did find a differential effect on generalization. He rated the patients and normals for their level of anxiety and found that the strong shock evoked more generalization in high anxieties than in the low.

Using the MMPI to select psychasthenic and hysterical subjects, Eriksen (1954) found more generalization in the hysteric than in the psychasthenic under threat of shock for slow responses or random shocks. This finding is contrary to the usual prediction by the supporters of the theory that high TMAS anxiety increases generalization since it has been found that subjects scoring high on the hysteric scale score low in TMAS anxiety.

Grice (1955) has offered some data which further question this expected relationship between high TMAS scores and increased generalization. Using a test which added the variable of maximizing conflict among the required responses, he did find that high anxieties generalized more than low anxieties. However, he also reported a high correlation between the intelligence level of his subjects and degree of discrimination. This led him to conclude that his findings, and perhaps other findings, are primarily attributable to intelligence differences and not anxiety level.

Kamin introduced one further variable in this search for personality correlates of performance on this task when he reported a correlation between changes in discrimination reaction time under threat of shock and mechanical aptitude scores (1955). He found no relationship between TMAS scores and changes in performance so he postulates his test as a better measure of susceptibility to experimental apparatus
stress since it measures familiarity with mechanical apparatus.

Perceptual accuracy and acuity

These tests differ from discrimination reaction time in that there is no emphasis on speed of responding, only on accuracy.

Focus of attention. The threat of an occasional shock improved visual discrimination, a result which may be due to an increase in level of attention and a concomitant narrowing of attention under this condition. Some support for this possibility is indicated by the further finding that learning of irrelevant auditory stimuli was decreased by this threat. Finally, the combination of shock-threat plus irrelevant stimuli did impair discrimination, as if although attention is narrowed by the threat, the subject also has difficulty keeping his attention focused in the presence of distracting stimuli (Silverman, 1954). If subjects studied detailed pictures while under threat of shock, a later test showed a decremental effect of this threat on recall of the details (Kohn, 1954). Unfortunately, this test involved other factors beside perception, such as the abstraction of details, retention, and later recall of these details so it is impossible to attribute the results merely to disturbance of perception.

Callaway & Band (1958) have attempted to examine the variable of focus of attention by measuring the effects of atropine on two general types of tests: (1) those requiring narrow attention (Hidden Figures Test and the Stroop Color Naming Test); and (2) those requiring broad attention (Luchin's Water Jar Test and Raven's Progressive Matrices). Hypothesizing that atropine broadens attention, they predicted that it would act differentially on these tests, impairing performance on the first type of test and facilitating performance on the second type. They were able to show the postulated impairment of performance on these two perceptual tasks, but were less successful in finding a facilitation of performance on the broad attention, cognition tasks. While one may disagree with the authors, both on the extent to which their tests fit their hypothesis and on the support their results give to the hypothesis, this experiment does attempt to investigate a factor of interest in stress research. It is possible that stressors differentially affect attention processes, narrowing or broadening them and thus impairing performance on only those tasks requiring a similar type of attention. Furthermore this effect may be mediated by hormonal discharges acting on the brain stem reticular formation and on, or through, the autonomic system, as Callaway & Band have suggested.

Reports to subjects that their projective tests indicated neurotic tendencies had no effect on a test of brightness contrast, although several tests reflected increased perceptual rigidity under this condition (Moffitt & Stagner, 1956). Davis tested infantrymen under different combat conditions without finding any change in several tests of perceptual functions (1956).

Visual acuity was insensitive to the fatigue induced by strenuous visual inspection (Brozek, Simonson & Keys, 1950), to general fatigue induced by as much as 100 hours sleep deprivation (Edwards, 1941; Patrick & Gilbert, 1896; Cooperman et al., 1934), and to prolonged semi-starvation (Franklin, 1946). Sleep deprivation also did not change measures of depth perception (Edwards, 1941).

Individual differences. In the area of individual differences, Stevenson & Iscoe reported that high anxieties on the TMAS required more trials and were more variable on a discrimination task than low anxieties (1956). However, there was no difference between these two groups in a test of brightness contrast after it was suggested to the subjects that they had neurotic tendencies (Moffitt & Stagner, 1956), nor in a transposition test following training on a discrimination task (Bendig & Vaughan, 1957). Finally, Callaway & Band could find no correlation between changes in performance on the Hidden Figures Test under atropine and scores on the MMPI or the Rorschach (1958).

Flicker fusion frequency

This indicator of a perceptual system's characteristic responsiveness to a rapidly changing stimulus is believed to indicate an individual's level of articulation (Saucer & Sweetbaum, 1958), or the lability of his perceptual system. This test may reflect the presence of psychological stress better than a variety of other perceptual tasks because the functional limits of the underlying factors may be more nearly approximated, leaving less room for compensatory factors to cloud the issue.

Surprisingly enough, FFF has been shown to be relatively insensitive to a variety of potential physiological stressors. Heavy doses of LSD-25 and mescaline (Landis & Clausen, 1954), a light dose of epinephrine (Basowitz, Korchin, Oken, Goldstein & Gussack, 1956) or deprivation of sleep for 114 hours (Tyler, 1947) did not change visual FFF. Prolonged visual inspection did decrease FFF, but this occurred only after four hours and not after two hours of inspection (Brozek et al., 1950). Rather extreme conditions of physical work, physical work plus high temperature or plus a caloric deficient diet did not alter visual FFF. Only physical work plus three days of total starvation produced a consistent drop in visual FFF (Brozek & Keys, 1944). This decremental effect of food deprivation was supported by the finding of Franklin.
that prolonged semi-starvation also decreased visual FFF (1946).

Induced anxiety stressors. Several studies indicate that FFF may be more sensitive to conditions arousing emotional reactions than it is to these more physiological conditions. For example, while physical exercise had no effect on FFF, the anticipation of participating in a competitive boxing match decreased visual FFF (Ross, Hussman & Andrews, 1954).

Davis (1955) found that among the many tests administered to combat troops, only visual and auditory FFF showed a decremental effect under these combat conditions. Krugman (1947) reports that among returning veterans, those suffering from anxiety reactions had lower visual FFF, although the anxiety and non-anxiety groups overlapped a great deal.

A similar effect of anxiety on FFF was demonstrated by Goldstone (1955). He found that the visual FFF of his high anxiety psychiatric patients and high anxiety normals, as measured by Lorr's Scale and anxiety symptoms, was lower, more variable, and decreased more under repeated testing than that of the comparable low anxiety groups.

Auditory vs. visual FFF. Several experiments have pointed out other variables which should be considered when using FFF. Davis (1955) demonstrated that after only one hour of solving mathematical problems auditory FFF was decreased, while after two hours visual FFF still showed just a tendency towards lower fusion. These results may be attributable to intensity, duration or similar differences in the auditory versus visual conditions, but at the moment it is tempting to relate these findings to others comparing visual to audio functions (Abramson et al., 1955b) and postulate that the audio system is more sensitive to psychological stressors.

Procedural variations. Suzuki (1950a) has reported a technique in which he holds the frequency of his flashing pulse constant at about 20 cps. He measures FFF by increasing and then decreasing voltage to determine where flicker first appears and then disappears. He has found that the difference between these two voltages is a measure which varies with severity of physical work performed by the subject. His conclusion that this is a valid measure of physical fatigue was supported further by a later experiment which demonstrated that it varied with the oxygen requirements of subjects doing physical work (1950b).

Saucer & Sweetbaum (1958) used a still different approach. They held the length of the stimulus pulse constant and shortened the dark intervals between flashes until subjects reported fusion. With this technique, schizophrenic patients reported fusion at longer dark intervals than normal controls, indicating, for these authors, poorer articulation for schizophrenics.

Perceptual span

The digit symbol subtest of the Wechsler-Bellevue is a measure of an individual's perceptual span, with an emphasis on the factors of narrow attention and immediate memory.

Induced failure stressors. The most frequent potential stressor investigated by means of this test has been induced failure, either on a different task or on the digit symbol test itself. Close observation through a one-way screen and shocks for failures on false norms produced a decrease in auditory span in two experiments (Williams, 1947; Carlson & Lazarus, 1947). Reports of failure on an earlier task decreased auditory span, while a test involving over-learned responses (a vocabulary test) was unaffected (Moldawski & Moldawski, 1952). A similar condition increased errors, increased variability, but also increased speed at which the experimental group worked to compensate partially in their total scores for their errors (Lazarus & Eriksen, 1952).

Auditory span was not susceptible to disruption by doses of LSD-25 which seemed to impair performance on such tasks as choice reaction time (Abramson et al., 1955b), complex cancellation tests (Jarvik et al., 1955b), or learning tasks (Jarvik et al., 1955a), and a low dose of epinephrine, though inducing anxiety-like feelings, did not change auditory span (Basonwitz et al., 1956).

Individual differences. Possible measures of individual differences in changes of auditory span under stress have been studied with interest ever since Williams (1947) reported a relationship between Rorschach responses and decreases in auditory span under close observation and shocks for failures on false norms. Westrope (1955) and Carlson & Lazarus (1953) have unsuccessfully attempted to replicate these findings. Westrope (1953) also failed to find a correlation between high and low anxiety of the TMAS and changes in auditory span under these conditions.

Stopol (1954) tried to induce stress by two different experimental situations, by either failure reports on an earlier test or by distracting stimuli. He, too, found no relation between Rorschach scores and changes in digit span. Stopol was also interested in the consistency in a subject's changes in digit span across these two situations as a measure of his tolerance to these potential stressors. Stopol assumed both conditions were stressors and since he found no consistency in digit span changes for his subjects, he concluded there was no consistency in stress tolerance. Unfortunately the basic data were not published so the extent to which these
situations differentially stressed his subjects (as defined in this review) cannot be determined. His hypothesis that individuals are differentially sensitive to stressors should receive much more experimental attention, but, as published, his experiment doesn't aid in confirming or rejecting it.

Mandler & Sarason (1952) reported no correlation between anxiety scores on their questionnaire and changes in digit span following a report of earlier failure. However, they do report a greater variability in the performance of their high anxieties as opposed to low anxieties. Heilbrun (1958) gave this test under neutral conditions to brain damaged, schizophrenic, psychotic, neurotic, and physically ill patients as well as to normal controls. He found a tendency for the brain damaged patients to have the shortest span, the remaining patients a longer span, and normal controls a still longer span, but with so much overlap between groups that he questions the sensitivity of this test for diagnostic purposes.

Perceptual rigidity

Rigidity of percepts is pertinent to psychological stress since one of the postulated characteristics of stress is that it interferes with the flexibility needed to acquire new percepts while at the same time facilitating entrenched, habitual percepts. Experimentally, a perceptual set is built up by instructions or training, or else established, cultural percepts are used. Then a test stimulus, differing from this set, is presented to measure the degree to which the subject sticks to his perceptual set in the face of incompatible cues.

Closures of open circles. This test attempts to measure rigidity by the subject's failure to perceive open circles when they are periodically given, interspersed with closed circles.

Parachute training, which should induce anxiety in the trainees, increased the number of closures consistently over those of a control group (Kochin & Basowitz, 1954). Also during the first week of training there was an increase in number of closures in the morning before the day's jump compared to their evening scores. Moffitt & Stagner (1956) found an increase in closures following a suggestion of a neurotic tendency in their subjects. This condition also led to increased rigidity in other tests: a reduced tolerance of ambiguous figures and an increased acceptance of the distorted Ames room as rectangular. However, another measure of perceptual rigidity, perception of incomplete figures, was resistant to change by this situation.

Airmen trainees were given an ego-involving set but this had no effect on the closure test (French, 1955). Since this condition failed to produce an increase in rigidity on several other tests, it did not seem to be an effective stressor. This experiment questioned the relationship between these several tests of rigidity (closure test, changing figures, design preference, Luchin's water-jar test, California F test) and failed to find any general correlation under either neutral or ego-involving conditions. If there is a common factor of rigidity underlying these examples of tests of rigidity, this experimenter could not find it.

In the area of individual differences, neither ego-involving instructions (French, 1955) nor suggestion of a neurotic tendency (Moffitt & Stagner, 1956) led to any correlation between closure test performance and TMAS scores.

Ambiguous figures test. This test is felt to measure tolerance of ambiguity, with low tolerance individuals resolving the ambiguity by means of premature hypotheses which interfere with correct perception. Perceptual rigidity is indicated by the resistance of these premature hypotheses to change as correct ones become more apparent.

Reports to subjects that their projective tests indicated neurotic tendencies decreased their tolerance of ambiguity, just as it increased closures and acceptability of the Ames room as rectangular (Moffitt & Stagner, 1956). Reports of failure on this task resulted in only a tendency for subjects not to perceive incongruous figures (Smock, 1955a) and the same condition also only tended to decrease tolerance of ambiguity (Smock, 1955b).

There was no correlation between anxiety as measured by the TMAS and performance on this test following the suggestion that the subjects had neurotic tendencies (Moffitt & Stagner, 1956).

Changing Figures. This test measures rigidity by establishing a perceptual set and then changing the stimulus gradually to determine the subject's ability to perceive the changed figure.

Ego-involving instructions had no effect on this measure of rigidity (French, 1955). This test was not correlated with the California F test, Luchin's water-jar test, closures test, design preference, and TMAS.

Distorted Ames room. The non-rectangular Ames Room situation is a measure of another aspect of rigidity, the degree to which an individual perceives either the broader environment as distorted or the figures as distorted, when the two are in conflict.

When subjects were informed that they showed neurotic tendencies, this increased their acceptance of the distorted Ames room as rectangular and this effect was related to scores on the TMAS, with high anxieties showing this greater environmental constancy (Moffitt & Stagner, 1956).
Reversible figures
This is a test of perceptual reversals induced by fixating such figures as the Necker Cube. This measure may also be related to perceptual rigidity.

Prolonged semi-starvation had no effect on perceptual reversals (Franklin, 1946). Bills (1931) attempted to induce perceptual fatigue by prolonged performance on this test. As the test continued he hoped to show that these fatigue effects were reflected in increases in “blocking” responses, which he defined as pauses equivalent to the average time of two or more control responses. He found only a tendency for blocks to increase in frequency and duration when subjects were tested for some seven minutes.

MOTORIC TESTS

Included under this general heading are those tests requiring a more co-ordinated or skilled motor response than those previously considered, or those tests designed to measure some aspect of a motor response such as strength, persistence, steadiness.

Matching tasks
These tasks differ from complex reaction time in that a more co-ordinated motor response is required.

A factor of interest on the stimulus side is the complexity of the display, with those of a limited display requiring narrow vigilance and those with many independent signals requiring flexible attention (Jerison, 1957), thus more likely to reflect stress effects. The rate at which stimuli are presented or changed may be determined by the subject or by the experimenter. If experimenter paced, higher rates may be more stress sensitive as the subject will not be able to compensate for temporary blocks in his responses, as he can if he works at his own pace.

On the response side, the degree of response co-ordination may be an important factor of stress sensitivity together with the rate at which responses must be made.

Discontinuous matching tasks. Tasks in which the response is made intermittently rather than continuously.

Broadbent found that steady noise increased errors on a prolonged vigilance task only in the experimenter-paced condition. In the subject-paced condition there was an increase in intermittent pauses, but this was compensated for by periods of rapid responding to mask any effect on total score (1953).

Broadbent compared the effect of noise on two broad attention tasks, one of which required an adjustment response to dials and the other an on-off response to lights. He found a decrement in performance on the adjustment response task but not in the on-off response. Part of this effect may be attributed to the easier discriminability and better attention arousing properties of lights versus dials, but some is probably due to differences in response co-ordination (1951).

Prolonged performance on a dial matching task reflected no stress effects even after four hours (Hoffman & Mead, 1946). General systemic fatigue induced by 30 hours of sleep deprivation resulted in performance decrements on a vigilance task after ten minutes (Wilkinson, 1958). Since tasks of less than ten minutes duration were not affected, one important variable again seems to be task length. However, this is not the complete story since an addition task showed no effect after 25 minutes.

Low body concentration of alcohol resulted in a consistent decline in performance in a Link Aircraft Trainer, which requires broad attention and almost continuous highly co-ordinated responses (Aksnes, 1954).

Continuous matching tasks. An example of this type of task is the pursuit meter task requiring an adjusting response to a continuously changing stimulus.

Pursuit tracking seems to be sensitive to thermal stressors with temperatures above and below normal decreasing performance over a five-day period (Teichner & Wehrkamp, 1954). The control subjects exhibited learning improvements in this experiment, but a second experiment in which the subjects were tested under 55°F conditions after long pretraining indicated that the results were due to a reduction in performance level and not to a reduction in learning rate (Teichner & Kobrick, 1955).

Pepler compared the effects of high temperatures and loss of sleep on pursuit tracking performance. Both conditions resulted in performance decrements but with the qualitative difference that high temperatures tended to decrease accuracy of response while loss of sleep reduced activity level, increasing failures to correct errors (1958). Specific fatigue induced by two hours of tracking tended to increase errors and the duration of errors, a response change analogous to Bill’s blocking, as defined previously on page 44 (Siddal & Anderson, 1955).

A decline in performance has been shown by other experimenters in a prolonged tracking task (Pincus & Hoagland, 1944; Hauty & Payne, 1953, 1955). Adams experimentally isolated the visual discrimination component of this task and found that prolonged performance on this component did produce a later decrement on the complete task (1955).

Even a relatively heavy dose of LSD-25 had no effect on a short test of pursuit tracking,
while this dosage seemed to impair hand steadiness (Abramson et al., 1955c).

Very little work has been done on testing the individual differences in stress responsivity with this task. Under neutral conditions, Matarazzo & Matarazzo (1956) found no correlation between TMAS scores and pursuit tracking performance.

Motoric dexterity tasks
These tasks require finer and more dexterous movements than in the matching tasks, e.g., letter cancellations.

Subjects tested for 30 minutes under noise attempted more items but had a lower percentage of accuracy because of an increase in errors (Smith, 1951; Corso, 1952). Prolonged experimental semi-starvation resulted in a slight decrement in a short duration Ball-Pipe Test (Brozek, Franklin, Guetzkow & Keys, 1946). Mild anoxic conditions led to only a temporary decrement in finger dexterity and arm-hand coordination followed by rapid adjustment and continued learning (Russell, 1948). During a submarine patrol, with its confinement, boredom, and increases in CO, the crew made increased errors in letter-cancellation tests (Schaeffer, 1951).

A moderate and a heavy dose of LSD-25 seemed to have little effect on letter, figure, and number cancellation tasks while performance on cancellation tasks involving cognitive manipulations was apparently disrupted, e.g., cancellation of consecutive numbers whose sum is nine (Jarvik, Abramson & Hirsch, 1955b).

Auble & Britton tested their subjects on a cancellation task under noise to determine if TMAS scores could account for individual differences in performance (1958). They found an inverse relationship between these two measures, with high anxieties doing better under noise and low anxieties tending to do better under quiet.

Pattern tracing
This task emphasizes a particular kind of perceptual-motor co-ordination rather than dexterity.

McKenzie (1957) attempted to determine what aspects of auditory noise play a part in the impairment of performance on this task. He found that a difference of 8 cps in two signals given independently to each ear decreased performance while an externally produced beat of 8 cps or the same frequency to each ear had no effect. Surprisingly, this internal beating hindered measures of manual co-ordination but did not affect performance on Thurstone's PMA tests.

Prolonged semi-starvation plus fatigue induced by physical work resulted in increased errors and length of contact time (Brozek et al., 1946).

Mirror tracing. This novel test requires a reversal of the habitual perceptual-response co-ordination.

Beier attempted to induce anxiety in his subjects by interpreting their Rorschachs as showing maladjustment and this produced an increase in time to complete this test (1951). Using a similar situation Wiener could show no effects on this task (1955).

In a later experiment Beier replicated these conditions to investigate the relationship between Rorschach measures and changes in mirror tracing scores (1953). No correlation was found between these two scores and unfortunately Beier did not publish any data concerning changes in mirror tracing under this condition. Under neutral conditions no correlation was demonstrated between anxiety as measured by Rorschachs or the Illinois Personality Inventory and mirror tracing (Ausubel, Schiff & Goldman, 1953).

Blindfold Stylus Maze. This is also a novel task in which the subject is denied accustomed visual information.

Perhaps due to the novelty of this task, most of the experimental work has been done on measures of individual differences in performance under neutral conditions. Farber & Spence found that while in eyelid conditioning high anxieties on the TMAS were superior to the lows, the lows were superior to the highs in the stylus maze, with the more difficult choice points differentiating best between these groups (1953). Axelrod, Cowen & Heilizer attempted to replicate these findings but could discover no correlation between TMAS scores and stylus maze measures (1956). Matarazzo used the entire range of TMAS scores rather than just the usual extremes and found only a tendency for middle anxieties to do better than both highs or lows (1955).

When anxiety level was determined by either Rorschachs or Illinois Personality Inventory, there was no differential effect of anxiety level on performance except that the lows were superior to highs on the first trial (Ausubel et al., 1953).

Vocal control
Speaking under delayed feedback conditions has been shown to be an extremely effective means of disrupting speech control. A series of experiments have examined some of the pertinent variables and a general conclusion seems to be that a delay of some 0.2 seconds is most effective at intensities of about 80 db. Disturbances were indicated by an increase in duration and intensity (Black, 1951; Spilka, 1954); by errors in reading digits (Butler & Galloway,
PSYCHOLOGICAL STRESS

1957); and by increasing loudness, increasing fractures, and decreasing intrusions (Korobow, 1955). Random delays had no effect on speaking while fixed delays did (Butler & Galloway, 1957). If delay and undelayed feedback were mixed, articulation and duration increased (Fairbanks, 1955).

Individual differences in responsivity to this stressful condition have been measured with a variety of personality tests by Spilka (1954) with changes in vocal intensity variation being correlated to scores on several of these personality traits. Korobow was also able to relate personality characteristics as determined by introspective Rating Scale A and B and the Jenkin’s Personality Questionnaire to different measures of speech disruption under delayed feedback (1955).

Tapping speed.

This task measures the speed of doing repetitious response.

Again McKenzie (1957) found that an internally produced beat of 8 cps resulted in a drop in performance on a tapping test, while neither an external beat of 8 cps nor a signal of 8 cps had any effect. General systemic fatigue induced by 65 hours of sleep deprivation only tended to decrease number of taps. There was no indication of response blocking on this task while math problems and a color naming test did show this effect (Warren & Clark, 1937). Patrick & Gilbert (1896) failed to demonstrate any change in tapping after 90 hours without sleep.

Ross et al., attempted to investigate the effects of muscular exercise and induced anxiety on this test but were unable to do so because they found the test too unreliable (1954).

Prolonged semi-starvation tended to decrease tapping (Brozek et al., 1946). Malmo & Finan found a significant decrement in rate on this task under anoxic conditions (1944), while Waldfoget et al. (1950), only found a tendency for similar anoxic conditions to slow tapping. They also failed to show any differential effect of anoxia on the performance of psychoneurotic patients as compared to normals.

Motoric steadiness.

Body sway and especially hand steadiness have been of interest in stress research perhaps because of the analogy to failures in muscular control in individuals under emotional tension.

Hand steadiness has been found to be sensitive to the Cold Pressor Test or shocks for failure on a task (Lewinsohn, 1956), as well as to a variety of disturbing and distracting stimuli, such as air blasts, flashing lights, horns, bells (Lofchie, 1955).

Hand steadiness was not affected by the anxiety aroused before a competitive boxing match, while body sway did increase (Ross, Hussman & Andres, 1954). The possibility that this measure of gross body steadiness may be more easily disturbed than hand steadiness is supported by Edward’s findings that 100 hours of sleep deprivation had no effect on hand steadiness while body sway increased (1941). Cooperman et al. (1934) did find that with their smaller sample of subjects both measures showed effects of 60 hours loss of sleep. Muscular fatigue induced by boxing exercises or a boxing match also led to decrements in both of these tests (Ross et al., 1954).

Hand steadiness was responsive to milder anoxic conditions than the tapping test or several learning tasks (Malmo & Finan, 1944). On the other hand Waldfoget et al. (1950) used a more severe anoxic situation and found only a tendency for hand steadiness to decrease. Finally, epinephrine at doses which induce anxiety-like feelings decreased hand steadiness (Basowitz et al., 1955).

Individual differences. Rodnick, Rubin & Freeman (1943) measured hand steadiness while distracting-type stimuli (air blasts, gongs, lights) were presented whenever subjects touched the side of the hole. Under these conditions hand steadiness decreased only for those subjects rated low in social adjustment, while it increased for those rated high. Lofchie (1955) tested his subjects while similar stimuli were presented randomly and he found that those scoring low on the Rorschach Index of Perceptual Maturity were more disrupted than those scoring high on this index.

Anoxic conditions did not result in differential changes in the hand steadiness of psychoneurotic patients and normal controls (Waldfoget et al., 1950). Shocks for failure on a task evoked more hand tremor in high TMAS anxieties and in high scorers on the MMPI subscales for hypochondriasis and depression than in low scorers on these tests. In this same experiment both shocks for failures and the Cold Pressor Test resulted in more hand tremor for patients with ulcers and psychiatric patients with anxiety symptoms than for a control group. However, neither TMAS nor MMPI scores were correlated with hand steadiness following the Cold Pressor Test (Lewinsohn, 1956).

Parsons, Phillips, & Lane (1954) attempted to correlate changes in blood eosinophil level with changes in hand steadiness under distracting type stimuli and induced-failure conditions. They found a drop in eosinophil level under both conditions, but these drops were not correlated to changes in hand steadiness. Thus there was no consistency in an individual’s reaction to a potential stressor as measured by a physiological and a psychological type of response.
Persistence

Persistence is measured by the time a response which requires muscular effort is maintained, such as an elevated leg, and probably measures motivation as well as energy expenditure.

Anxiety-arousing doses of epinephrine decreased time of leg elevation (Basowitz et al., 1950). Psychoneurotic patients were less persistent than normal controls in holding an arm or leg extended (Jones, 1948), which may be attributed to motivational variables. However, a correlation has been found between persistence in leg elevation and history of past crimes among medical patients. Those patients who had committed violent, homosexual crimes and thus, the author hypothesized, should be anxious in a hospital setting were less persistent than those who had committed heterosexual crimes and should be less anxious in a hospital setting (Tong, 1957).

COGNITIVE TESTS

These tests emphasize cognitive factors by requiring that the subject solve problems by means of logical rules, form concepts, or make cognitive associations to stimuli.

Problem solving

The usual problems given are rather simple mathematical problems. Osier (1954) found that reports of failure lowered performance on long division problems, while attempts to arouse the student's anxiety by a complaint about his conduct had no effect on his performance on this task.

Broadbent (1958) reported that random noise interfered with solutions of subtraction problems. Angelino & Mech (1955) used a lower intensity noise than Broadbent and got only an initial decrement on rapid, simple addition problems.

General fatigue induced by deprivation of sleep for 30 hours had no effect on an addition task although performance on a vigilance task of shorter duration suffered a decline (Wilkinson, 1958). Bills had his subjects perform on a variety of mental tasks in order to fatigue his subjects. This condition led to an increase in the length and the frequency of blocking of responses in the solutions of addition and subtraction problems (1931). Warren & Clark (1937) also used addition and subtraction problems and found a similar increase in blocking of responses after 65 hours without sleep.

Short duration, mild anoxic conditions had only an initial decremental effect on the solution of simple addition problems, followed by a rapid recovery and continued improvement (Russell, 1948). Medium and heavier doses of LSD-25 seemed to impair performance on addition and subtraction problems (Jarvik et al., 1955c).

Anxiety, as measured by the TMAS, had no differential effect on solution of long division problems under neutral, failure, or anxiety-arousing conditions (Osler, 1954).

Problem solving rigidity. This rigidity is inferred from the subject's adherence to a given set for solving problems when a shorter, more efficient solution is available, as in the Luchin's water jar problems. Psychological stress is postulated to increase this rigidity.

Pally (1955) found that reports of failure increased measures of rigidity on the Luchin's problems. Cowen found that failure on an insoluble problem or reports to subjects that their projective tests had maladaptive features led to increased rigidity (1952a). Cowen (1952b) also showed that a group which was informed that their projective tests had questionable features were more rigid than a group whose tests were interpreted as being favorable.

Maltzman, Fox, & Morrisett (1953) hypothesized that high anxiety, as measured by the TMAS, would interfere with the shift to a more appropriate solution in the water jar test, while facilitating performance in an anagram test in which the given set was appropriate for the test trial. Their results supported this hypothesis, with high anxieties being more rigid on the water jar problems but making fewer errors on the anagrams than did the low anxieties. However, French (1955) found no correlation between scores on the TMAS and rigidity on the water jar test, when given under either neutral or ego-involving conditions.

Concept formation

Included under this heading are tests which generally require that the subject form concepts by abstracting out a common element or principle from a complex situation or series of situations. In addition, in the Wisconsin Card Sorting Test the subject must shift his concepts on succeeding trials.

Random electric shocks had a detrimental effect on performance on the Wisconsin Card Sorting Test, while auditory distractors and heckling had no effect (Ross, Rupel, & Grand, 1952). Reports to subjects of maladjustment features in their Rorschachs had no effect on the Wisconsin Card Sorting Test (Wiener, 1955) but this same type of situation did have a detrimental effect on the Holsopple Sorting Test, as well as on another concept formation task (Beier, 1951).

Corso (1952) investigated the effects of random noise on performance in the Minnesota Paper Form Board Test. This test is perhaps the reverse of the Wisconsin Test, since it
requires the combination of isolated figures to form a concept of the whole rather than the abstraction of elements from a complex stimulus. Corso found that noise led to more items attempted, more errors made, and, surprisingly, less variability of performance. Smith (1951) used the same task and the same potential stressor but added monetary incentive to do well. His results show that under these conditions the changes in performance reported by Corso do not occur.

Moderate and heavier doses of the psychomimetic drug LSD-25 seemed to impair performance on this Form Board Test (Abramson et al., 1955a).

Individual differences. In the area of individual differences on this type of test, Wesley (1953) scored his subjects on anxiety and rigidity by means of special questionnaires before testing them on a task similar to the Wisconsin Card Sorting Test. Those scoring high in rigidity and low in anxiety were faster in shifting conceptual sets than the medium scorers on his questionnaires. On the other hand, those scoring high in anxiety plus low in rigidity did not differ from this medium group in their performance on the sorting task. Thus his test of high rigidity seemed to be a better predictor of performance than his measure of high anxiety.

Mandler & Sarason (1952) investigated the relationship between anxiety as measured by another questionnaire and performance on the Kohs Block Test. They found that under neutral conditions the high anxieties only showed more variability than the low anxieties. However, a failure report led to improved performance for the low anxieties while the high anxieties did worse.

Coding tests

Under this rather vague classification are tests of color naming and verbal associational learning, both of which require that the subject make a cognitive association to a stimulus.

Color naming. Electric shocks for failures, criticisms for failure, or time pressure increased errors of omission on a high speed color naming test (Davidson, Andres, & Ross, 1956). Sleep deprivation for 65 hours increased blocking of responses on a color naming test (Warren & Clark, 1937). On the other hand, Cooperman et al., (1934) reported there was no decrement in performance on such a task after 60 hours without sleep unless the duration of the task was prolonged.

As for measures of individual differences, Davidson et al., (1956) failed to find any differential effect on a high speed color naming task of high and low anxiety as determined by the TMAS. However, they did find some interaction between anxiety and performance under electric shocks for failure, criticisms for failure, or time pressure, leading them to conclude anxiety may have a "priming" function.

Conflictual color naming. The Stroop Test has one section which presents each color word in the wrong color for that word. This introduces the added factor of interference between responses, and the subject must ignore the conflictual word and name the color. Stressors should decrease the subject's ability to handle this interference, thus depressing his level of performance.

Callaway & Band (1958) postulated that if atropine broadens attention processes it would have a detrimental effect on tests which require narrow attention—in this case the Stroop Test. Their results give some support to this hypothesis, since atropine did impair performance. On the other hand, they have also postulated that epinephrine should narrow attention, but Basowitz et al., (1956) found that this drug had little or no effect on this task although the dosage level was sufficient to arouse anxiety-like feelings.

Callaway & Band (1958) were unable to discover a correlation between MMPI factors and changes in performance on this test under atropine.

Verbal association learning. Experiments on verbal learning which are of interest in the area of stress research have been the attempts to define the characteristics of this task which contribute to poorer performance for individuals under neutral or potentially stressful conditions. Investigation has primarily centered on objective measures of the difficulty an individual experiences in associating a list of stimulus words with the correct responses. There are two kinds of measures of this difficulty: the association value of each stimulus to its response, and the degree of intralist similarity as a measure of competitive response tendencies. Individual differences in performance are usually predicted on the basis of TMAS scores, with high anxiety facilitating performance if the test has many correct and few competitive response tendencies but interfering with performance if there are many competitive, incorrect tendencies. Naturally stress would more likely be reflected in performance on tests which have a high degree of competitive response tendencies.

Montague (1953) used lists in which he controlled for both association value and similarity. In accordance with the general hypothesis, he found that high anxieties on the TMAS performed worse than low anxieties on the list with many incorrect response tendencies, but performed better on a list with many correct response tendencies. Nicholson (1958) used lists of high vs. low association value and compared high and low TMAS anxieties under task-orienting vs.
ego-orienting instructions. The high anxieties were similar to the low anxieties on both types of lists in the task-oriented situation. However, his results do indicate that the high anxieties suffered a large decrement in performance on the low association value list under ego-orienting instructions.

High anxieties on the TMAS were no different from low or medium anxieties in their learning of highly associated items under neutral conditions, although the usual prediction is that their performance should be facilitated (Sarason, 1956). Nor did reports of failure have a differential effect on the performance of these different groups. However, under ego-orienting instructions, the low and medium anxieties improved while the high anxieties did worse. This is compatible with Nicholson's findings, although his effect was noted on the low association list, while Sarason's effect was on a high association list. Sarason (1957a) again used a list of high association value to investigate further the differential effects of failure reports on TMAS anxiety. Again he found no differences between his high and low anxieties under neutral conditions but under failure the performance of the high anxieties declined. In addition, the low anxieties even showed a significant improvement when they were given failure reports on a different task rather than on a preliminary list of nonsense syllables.

Korchin & Levine (1957) compared the performance of high and low anxieties on a list of high association value and on false math equalities. While the two groups did not differ on the verbal learning task, the high anxieties did worse on the false equations than did the low anxieties.

Finally, Spence, Farber, & McFann (1956) corroborated Montague's results by controlling both association value and similarity. On lists of low competitive response tendencies, the high anxieties did better than low anxieties, and on lists of high competitive response tendencies, the low anxieties were superior.

Moving on to experiments in which lists of low association value were used, Taylor (1958) found that high anxieties learned this type of list better than low anxieties under neutral conditions. However, in a failure situation she failed to find the predicted greater drop in performance for the high anxieties as compared to the low anxieties. Both groups did worse but they were not differentially affected.

Hellizar, Axelrod, & Cowen (1956) used a slightly different technique in selecting the stimulus items. They chose conflict and neutral words for each individual and paired these words with nonsense syllables of low association value. Their results do not indicate any differences in performance between high, medium, or low TMAS anxiety groups on these items.

Silverman & Blitz (1956) compared the effects of threat of possible shock and threat of a definite shock during learning on performance in a later test. They found these threats had no effect on the performance of the high anxiety group but that under the threat of possible shock the low anxieties did better and under threat of a definite shock they did worse. Silverman and Blitz also included a measure of learning incidental numbers and their results indicate that high anxieties showed little incidental learning, perhaps due to their high level of fixed attention under threatening conditions. Another variable of interest pointed up by this experiment is that of the possible differential effect of a threatening condition on learning vs. performance—perhaps after the removal of the threat during the test performance high anxieties react differently from low anxieties.

Sarason (1957b) supported Taylor's results by also showing that high anxieties do better than low anxieties on a low association list, under neutral conditions. He also included a group of medium anxieties, who did worse than even the low anxieties. Under ego-orienting instructions low anxieties did better than under task orienting instructions—a finding which fits in with Nicholson's (1958) and Sarason's (1957a) results. Sarason tested his subjects for retention after 24 hours and found that the high anxieties were able to perform better than all other groups, regardless of the previous experimental conditions. This result Sarason attributes to a dissipation of anxiety over time, permitting the high anxieties to do well as they gain control of the situation.

Another group of experiments have been concerned mainly with the degree of competition between response tendencies as measured by the level of similarity or duplication among the items used. Lucas (1952) varied the level of duplication in his lists and found only a tendency for high anxieties on the TMAS to drop in performance as the duplications increased. Lucas also used a varying number of failure reports and as the reports of failures increased the low anxieties improved while the high anxieties suffered a decrement in performance.

Spence, Taylor, & Ketchel (1956) found that on low competition lists high anxieties did better than low, while on high competition lists low anxieties tended to do better than the high anxieties. Taylor & Spence (1952) had reported previously that on a list using only two words as responses and thus highly competitive, high anxieties made more errors and took more trials than low anxieties.

However, Lazarus, Deese, & Hamilton (1954) also used high duplication lists and they were unable to find any difference between the high and low anxieties on their learning of these lists.
Random shocks or shocks for failures to respond had no differential effect on these two groups. Finally they scored these subjects on hysteria and psychasthenia in the MMPI, but these measures also were unrelated to performance on the learning task.

Finally, on lists of low similarity and thus low competition between response tendencies, Taylor & Chapman (1955) found high anxieties did better than low anxieties. Deese, Lazarus, & Keenan (1953) used a list of low duplication and their results indicate little difference between the performance of those scoring high and low in neuroticism on the Winne Scale. Random shocks had little differential effect on these two groups but under shocks for failures the high neurotics improved in performance while the low neurotics suffered a decline. Since neuroticism scores are correlated with TMAS scores, these results are not compatible with the usual prediction for high and low TMAS anxiety groups.

**DISCUSSION AND SUMMARY**

It is obviously difficult to draw specific conclusions about psychological stress from experiments which differ widely from each other on their experimental subjects, experimental conditions, and experimental tests. Since few of these experiments can be equated on a satisfactory number of these variables, it would seem premature at this time to suggest more than a few general conclusions to serve as tentative hypotheses for further experimentation.

This review was organized on the basis of the psychological functions, i.e., vision, cognition, which were postulated as underlying an individual's performance on the experimental task. One hope was that this organization might point out differences in the sensitivity of these functions to stress. The first general conclusion reached by this reviewer is that this hope was not realized. All of the promising results can be more easily attributed to the effects of other variables rather than to differential effects on these postulated functions. Some possible functional differences were suggested by the data to this reviewer, e.g., audition may be more sensitive than vision to disruption by stress. However, it is felt that even these differences are better accounted by a group of other variables; variables which run like a recurring theme through many of the reported experiments.

What are these promising variables? It is rather difficult to label some of them, probably because they are not neat, unitary variables. Instead they are more often characteristics of what are usually accepted as more specific variables. It might be mentioned also that these variables are related to the concept that psychological stress is a state which disrupts or disorganizes behavior. Therefore tests which are designed on the basis of these variables are more likely to provide adequate measures of these disruptive effects. The variables are as follows:

1) **Functional limits:** This term, more than any of the others which follow, is descriptive of a general aspect of many different variables. Regardless of which variables or functions contribute to performance, the disruptive effects of stress can be demonstrated best if the task used successfully measures the limits of these functions under neutral conditions. The closer a task comes to fulfilling these optimal conditions, the more likely it is that even states of mild stress will be reflected in decrements in a subject's later performance under stress or of a stressed comparison group.

Examples of tests which are somewhat successful in measuring functional limits are the perceptual tests of flicker fusion frequency and perceptual span. However, tests of most functions can be varied in a similar method of limits procedure, altering variables which will extend the individual to his functional limits, e.g., by increasing the number or complexity of the stimuli presented, or of the motoric responses to be made, or of the cognitive manipulations to be carried out.

A few words might be said about situations in which the subject is tested beyond his capacity to perform, rather than at or just below his limits. A simple example of this is the presentation of stimuli at a rate faster than the subject can possibly respond. In these "overload" conditions the experimental task itself becomes the primary source of stress, on condition that the subject is motivated to accomplish everything demanded of him. A percentage decrement in performance at the overload rate as compared to performance at a rate within his limits is not a good indication of stress, it would merely show that more is being asked of the subject than he can handle. If it is to be inferred that the subject is trying to meet these excessive demands and is stressed by his failure to do so, other behavioral indicators would be preferable. One measure of this would be to show that his absolute level of performance has decreased; in other words, he is missing stimuli he could respond to when operating within his limits. Other behavioral reflections of stress would be bursts of sporadic responding and emotional responses under these conditions.

Task duration is another overload condition which is often used as a stressor. This type of situation is also especially sensitive to decreases of motivation, but if motivation can be maintained, the differential sensitivity of various
functions to fatigue-stress can be determined. As might be expected, fatigue induced by sleep deprivation interacts with this task stressor, so that most of the effects of sleep deprivation have been demonstrated on long duration tasks.

2) Level of attention: or level of vigilance (Mackworth, 1950). Since an individual may mask the disruptive effects of stress by increasing his motivation to perform better, motivation is a crucial factor in stress research. Attention level is important in this respect because one effect of motivation may be to increase attention to stimuli, especially those relevant to that specific motivational state. Therefore, if a subject has been performing in a task which already required a high level of attention, he is less able when stressed to compensate for the resulting disruption by increasing his level of attention.

Examples of this kind of test are predominately perceptual tasks in which the subject attends to discrete stimuli appearing for brief intervals of time: reaction time, perceptual span, and flicker fusion frequency.

3) Focus of attention: In addition to attention level, focus of attention is another aspect of attention processes which is important in research on stress. As the individual becomes stressed, an early reaction may be a narrowing of the focus of attention, both across time and space. Then, a later reaction or a reaction to greater stress may be the loss of ability to control even this limited range of attention, as the individual becomes overly responsive to all stimuli. A facile conceptualization of stress is that the internal stimuli produced, act as irrelevant, distracting stimuli, preventing effective control of attention. Thus tests which demand flexibility of attention (Jerison, 1957) or broad attention (Callaway & Band, 1958) will reflect an earlier effect of stress than will tests requiring a limited focus of attention. Examples of this type of test are: complex reaction time; discontinuous matching tasks with a broad display; complex, integrative problem solving tasks.

4) Strength of habit or level of learning: This variable is related to psychological stress by the notion that stronger habits or overlearned responses are less easily disrupted by stress than are corresponding responses which are weaker or underlearned. Performance on an overlearned verbal association test will not demonstrate a disruptive effect as easily as performance on a novel task, such as the blindfold stylus maze. Using the term "novel" task naturally brings up the factor of the flexibility needed to learn a new response, which introduces the next stress-sensitive variable.

5) Reversal of habit or set: This variable is descriptive of a variety of tests whose one common characteristic is the factor of flexibility needed to reverse a set or habit acquired in some fashion, e.g., by instructions, experimental pre-training, or environmental pre-training. If stress is conceptualized as facilitating overlearned responses, it will also interfere with the flexibility needed to inhibit these responses and acquire new responses. This flexibility can be made the primary experimental variable by measuring the amount of interference introduced by learning a new response following pre-training on another response. The more the test response conflicts with the earlier response, the greater will be the interference and the more this interference will be enhanced by stressors. Examples of tests in which the habitual set or habit conflicts with the new one are: changing figures, Ames distorted room, mirror tracing, choice reaction time with re-pairing of trained associations. The Luchin's water jar test measures flexibility in a slightly different way in that the original set is still adequate for solving the test problem, but it is not as efficient as a new approach would be; while in the other examples the original set is incorrect for solving the test problem.

These, then, are the variables which seem to play an important part in making a test stress-sensitive. In addition to task-stressors, many different experimental situations have been used as potential stressors. These stressors can be grouped under three broad headings: 1) motivational arousing, i.e., anxiety, frustration, failure, or fear arousing; 2) distracting, irrelevant, or novel stimuli; 3) disrupters of attention processes which disrupt performance on diverse physiological homeostasis, i.e., epinephrine, LSD, anoxia, fatigue. Of course, most stressful situations can be shown under analysis to include elements of two, if not all three of these categories.

The question of whether there are general stressors which disrupt performance on diverse functional tasks has not been adequately answered since the tasks are not usually equated on their stress sensitivity. Broadbent (1957) has pointed up a type of characteristic which may contribute to making a situation a general stressor. After finding that noise at a specific intensity and frequency was an effective stressor on a test of reaction time, Broadbent went on to show that this type of noise as a stimulus evoked a faster reaction time than noise at a lower frequency and intensity. He concluded that its success as a stressor was a function of its ability to compete with other stimuli to control responses, or, in other terms, to compete for attention.

In the area of individual differences in sensitivity or responsivity to stressors, the findings in experiments using measures of individual differences have been very inconclusive. The TMAS has received the greatest amount of attention in stress research and most of the
promising results in these experiments have been counterbalanced by negative or even contrary findings. Other measures of individual differences have been used, but they have not been tried in a sufficient number of experiments to assess their validity.

A few experiments have been concerned with consistency in an individual's responsivity to different stressors (Stopol, 1954). The results were negative although this may have been due to the use of non-stressful situations rather than to inconsistency in response. Perhaps a satisfactory measure of individual differences can be derived by finding first those individuals who are stressed by different situations and then devising a test which will measure characteristics differentiating them from those who are consistently resistant to stress.

ACKNOWLEDGEMENT

Work included in this paper was done in connection with research supported by the Office of the United States Army Surgeon General under Contracts No. DA-49-007-MD-575 and DA-49-007-MD-684 and a United States Public Health Service Grant No. M-1871, Dr. James G. Miller, Principal Investigator. The opinions expressed in this publication are not necessarily those of the Office of the Surgeon General, Department of the Army, or of the United States Public Health Service.

The author wishes to thank the members of the seminar on Individual Stress at the Mental Health Research Institute, The University of Michigan: C. C. Foster, S. Kornblum, L. Uhr, and especially R. W. Gerard, J. G. Miller, and H. M. Jackson for their helpful and stimulating discussions.

BIBLIOGRAPHY

15. BEIER, E. G. The effects of Rorschach


34. CONRAD, R. Some effects on performance of changes in perceptual load. J. exp. Psychol., 1955, 49 (5), 313-322.


42. EDWARDS, A. S. Effects of the loss of 100 hours of sleep. Amer. J. Psychol., 1941, 54, 80-91.


48. FORLANO, G., BERMACK, J. E. and CONKLEY, J. D. The effect of ambient and body


74. KNEHR, C. A. Schizophrenic reaction time responses to variable preparatory intervals. Amer. J. Psychiat., 1954, 110 (8), 585-588.


84. LAZARUS, R. S. and ERIKSEN, C. W. Effects of failure stress upon skilled performance. 1952, 42, 100-105.


112. RODNICK, E. H. and SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. Amer. J. Psychiat., 1940, 97, 214-225.
114. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. J. exp. Psychol., 1953, 45, 35-43.
140. SUZUKI, K. Oxygen consumption during and after exercise and its relation to the degree of fatigue as measured by the method of electric flicker. Tohoku J. Exp. Med., 1950b, 52, 9-16.


151. WALDMAN, M. Personality factors and performance under stress in schizophrenics. Amer. Psychol., 1951, 6, 314.


CURRENT ISSUES IN GENERAL SYSTEMS THEORY AND PSYCHIATRY

William Gray

We in the mental health professions live in very exciting times—times of great abundance and great problems. We have a richness of approaches and we have gotten over our feelings of hopelessness about mental illness and emotional disturbance and even over the hopelessness that each of us had about the possibilities for further growth and development. We have transactional analysis, psychoanalysis in many forms, group therapy, encounter therapy, existential therapy, assertive therapy, and many others.

Our interest in general systems theory lies in the help that this rather profound theory can give us in understanding the relationships between the many forms of psychotherapy presently in practice. We are also interested in what help general systems theory can give to the practicing psychotherapist. This interest has arisen as a result of progress in group and family therapy, which has thrust into the center of our attention the effect of memberships in superordinate groups on our function as individuals, and what is of vital importance to the mental health profession, the indication of the origin of psychopathology in the needs of the supersystem.

The brand new question is, "Does Johnny's psychosis, or psychosomatic illness, or depression, serve a homeostatic function in the family system, the group system, or the network system?" Should we now go about our task from a systems perspective, that is, should we focus our intervention on changing the family, group, or marital system, and if we do this, will the individual's symptoms disappear? The answer is a booming "yes," as the family therapists, group therapists, and transactional therapists will tell you. So we must learn new terms, like homestasis, boundary function, positive feedback of runaway type, subsystem formation, information codes, game theory, and leadership function, and we must learn to observe sequences of behavior and to intervene at appropriate times.

But all of this is somewhat mechanistic, a matter of shifting and pushings of superstructures, and we are in danger of becoming master puppeteers who maneuver our patients into health by changing the rules of the systems in which they exist. This, however, would not work, since passivity corrupts, and total passivity totally corrupts. Thus all of our good group and family therapists add to their approach elements aimed at fostering individual growth and development, at undoing neuroses; and, in order to do this, they apply elements derived from psychoanalysis, transactional analysis, gestalt therapy, and many others.

Our challenge now is to see whether we can extend the general systems framework, so very congenial to the study of supersystems, to the study and treatment of intrapsychic phenomena. This is a beginning task and a difficult one, for, to see ourselves as systems, akin somehow to the larger systems that surround us and the smaller systems that construct us, is not easy. It is here that the mechanistic flow of much systems theory leaves a bitter taste or is simply indigestible. However, there is such a thing as humanistic systems theory, and we have an opportunity to begin to work on it to further free our own humanity and that of our fellows.

The objection that might be raised, that it will introduce us to the notion of constraints, is not valid, as it has become quite clear that freedom and growth issue from constraint, not from its absence. This is the result of the well known theorem of W. Ross Ashby who has shown that any complex system under constraint, when faced with an insoluble problem, can only solve this problem by increasing its complexity. This is in accord with what Roy R. Grinker, Sr. says in his plea that we must become unitary thinkers, when he comments that the pathway to conflict resolution is to rise above conflicts, although, of course, we must be clearly aware of what they are ("The Continuing Search for Meaning," in W. Gray and N. D. Rizzo, eds. Unity Through Diversity. New York: Gordon and Breach, 1973).

Ludwig von Bertalanffy, a humanist from his earliest years, was not blind to the contributions that mechanistic systems theory made to humanity and to systems theory. He insisted only that the machine model was a defective case, in the mathematical sense, of a more humanistic general systems theory. He was gracious enough to give me credit for the formulation of at least four humanistic general systems principles. These are to be found in his paper, "The Unified Theory for Psychiatry and the Behavioral Sciences," published in Adolescent Psychiatry (April 1973), and in a paper of mine, "Bertalanffian Principles as a Basis for Humanistic Psychiatry," in E. Laszlo, ed., The Relevance of General Systems Theory (N.Y.: Braziller, 1972).

The four Bertalanffian or humanistic principles include the organismic system or non-reductionistic approach; the conception of the organism and personality as an active system, rather than reactive or robotic; the need to deal with what is specifically human, that is, man's capacity to engage in symbolic activities as well as moral, ethical, and creative pursuits; and the principle that anamorphosis or morphogenesis be recognized as a natural law, and that it be coupled with a realization that new organizational laws can be expected to be found at new organizational levels. Perhaps the latter should be divided into two principles. The first half contains the principle that there are three types of equilibrium: (1) entropic, in which equilibrium is maintained at the expense of structure, (2) homeostatic, in which structure is maintained in the face of disturbance, (3) morphogenetic, in which disturbance is dealt with through internal restructuring and new growth, and it is the morphogenetic that is
characteristic of man and his institutions and, incidentally, of life itself. The second half is the fairly simple principle that one must expect different organizational forms at different levels; that is, the family or group system will have somewhat different laws than the individual, while, at the intrapsychic level, organizational laws will be different at the Oedipal level than at the anal.

I am not satisfied with my original listings, for, although true, they do not go far enough. I have been more interested in detailed exploration of morphogenic systems in personality and therapy and how these balance out against a homeostatic background. What appears to be helpful in therapy occurs along these lines, that is, the occurrence of growth in a setting of safety. It is the psychotherapeutic situation that offers the setting of safety in our work, and, if we are mindful, offers the opportunity for morphogenic growth. I think we need to be aware that defensiveness and dependency bubble up the structure of our interpretations, and this is the way that homeostatic systems operate. Moreover, depending upon where we are at in a particular therapy, we should either be happy about it or take steps to undo the homeostating mechanisms. This, in general, can be accomplished by a mismatch between the codes of the patient and his neurosis and those of the therapist, introducing a phase of adaptive self-organization in the patient in which he gains a new state of equilibrium by changing his codes. This is Laszlo's term and complements his second form of regaining of equilibrium, that is, through manipulative self-stabilization which, if done by the patient, is homeostatic in type. These are complementary terms; in general, the therapist is engaged in manipulative self-stabilization while the patient is engaged in adaptive self-organization or reorganization. Of course, it cannot always be this way, for the patient would be reduced to a manipulated robot, and, as therapists, we all realize that there are times when adaptive reorganization of our own codes and ideas is necessary. There are some beautiful examples of this in Emanuel Peterfreund's book, Information, Systems, and Psychoanalysis (N.Y.: International Universities Press, 1971), in which he traces in a very sensitive and intuitive way the changes in his own internal codings as the therapy proceeds and he attempts to be in tune with the patient. These are lovely examples, I believe, of my own theory of emotional cognitive structuring, which states that our own cognitive understandings of a situation are organized by the feelings that we have and our ability to process these feelings into delicate and yet persisting emotional nuances.

But I want to stick to the main theme of a morphogenetic-homeostatic analysis of personality and therapy. My basic view is that man is a morphogenetic system encapsulated in a homeostatic capsule, and the implications of this are that we must know something about both types of systems—what it is that increases their strength and what it is that permits them to change. We have covered the homeostatic elements to some degree, so let us now proceed to morphogenic systems. The problem of morphogenesis is the problem of the origin of differentness from sameness and is, in theory, explainable either through the concepts of Maruyama and his notion of positive feedback or through the concepts of Rosen's two-factor theory. (M. Maruyama, "Mutual Causality in General Systems" in J. H. Milsum, ed., Positive Feedback. N.Y.: Pergamon, 1968; R. Rosen, Dynamical System Theory in Biology. N.Y.: Wiley, 1970, and "A Survey of Dynamical Descriptions of System Activity" in Gray and Rizzo, op. cit.). The principles are simple enough, Maruyama believing that if you have a deviation amplifying feedback system, then differentness will be produced as long as it is in operation. Thus, if identical twins decide they want to be different, or if their environment demands it from them, we do not have to look for innate differences to explain their differences. Rosen's two-factor theory really concerns itself with situations in which the operation of different forces results in unstable equilibria which respond to a sufficient degree of perturbation, with the system then taking off in pursuit of a new position of metastability, with the cycle being repeated as many times as one wishes. This is typical of group therapies, where an equilibrium position (for example, a defensive posture of seductiveness to cover jealousy) becomes unstable and then, with a perturbation (for example, being rejected by the group) proceeds on to a new metastable equilibrium of, for example, hostility covering love or pain. It is my opinion that keeping these humanistic systems principles in mind will help us in doing better psychotherapy and in studying what it is that is occurring in the psychotherapy we carry out.

I do not want to neglect the process by which overwhelming morphogenesis can be limited, for this is just as important, since the world is composed of those willing to change too rapidly as well as those unwilling to change at all. In general, this is carried out in two ways; by bounding or focusing discussion and interaction, and by cognitive interpretations that let people know where they are at and thus provide a metastable point for temporary closure. J. E. Durkin has a theory complementary to my theory of emotional cognitive structuring ("Encountering: What Low Machs Do" in R. Christie and F. Geis, eds., Studies in Machiavellianism. N.Y.: Academic Press, 1970). His theory holds that change always takes place in the emotional field, while stability occurs from the simple process of thinking about it. In the example just given, if you interpret the patient's seductiveness as always occurring when he appears anxious or, even better, when the ordinary person might be jealous, you will tend to stop the morphogenetic process and allow the homeostatic fields to take over, although such interpretation will also lead to a later burst of morphogenic activity. Stabilizing activity on the part of the therapist is well described by Durkin and Glatzer who state that the neutrality of the therapist alone is not enough to maintain therapeutic movement, but that it is essential to bring about insight into the transference behavior. Otherwise, what is referred to in the growing general systems parlance of psychotherapy as a runaway, or a continually escalating series of positive feedbacks, would occur. ("Transference Neurosis in Group Psychotherapy: The Concept and the Reality" by H. E. Durkin and H. T. Glatzer in Wolberg and Schwartz, eds., Group Therapy: 1971. N.Y.: Intercontinental Medical Book Corp., 1973.)

Before leaving what to me is the most important part
GENERAL SYSTEMS THEORY AND PSYCHIATRY

of the general systems view of personality, I would like to comment that morphogenetic-homeostatic balances within the personality are, of course, in a system relation with a similar aspect of their surrounding system. I think a study of this would be intriguing and is an urgent necessity for such matters as the design of health care systems whose design often ignores such factors. For example, untrammeled homeostasis in the surrounding system may well lead to morphogenetic outbursts in individuals, with various types of characteristic time delays. The opposite may be true, while patterns of fluctuation may turn out to be most conducive to patterns of health in individuals.

I would like to make it more evident than I have, that the sine qua non of morphogenesis is primary spontaneous activity, and it is this that we watch and treasure in our patients; that is, when they say something or behave in a way that is spontaneous and new, although, of course, it must be relevant. To hear a burst of assertive comments from a previously passive person is always pleasurable and can be built on.

The second area that I have been intrigued with is that of boundary functions and how these become more stable or more permeable, and the even more important issue of the type of transport that takes place across boundaries in living systems. This is an issue taken up by H. Durkin in her paper on, “Catalyzing Self Regulatory Processes in Group Therapy”, presented recently at a meeting of the International Society for Group Therapy in Zurich. Durkin stresses the role of the group therapist in serving a regulatory function for the group as a whole. O. Kernberg writes about the boundary function of the therapist in having to serve not only the patient but also the surrounding systems, in “A Systems Approach to Priority Setting of Interventions in Groups”, presented at the 30th Annual Conference of the American Group Psychotherapy Association, Detroit, Michigan, February, 1973. H. P. Laqueur, the father and developer of multiple family therapy, deals in particular with the different boundaries involved in individuals, dyadic, couple, peer group, multiple family, and network therapy, and how an awareness of the relationship of each system to its supersystem and subsystem can be incorporated into the operating codes of the therapist, no matter which type of therapy he is carrying out, leading to improvement in his effectiveness (personal communication, 1973). A child therapist doing his work with family therapy principles in mind will do better work, and the opposite is true.

My own view of the importance of boundary functions in a general systems view of personality is the realization that, without them, life is impossible; that without an ability to shift permeability, life is also impossible, and that rate changes, upon which rests the whole concept of how change occurs, take place at boundaries. Thus a formation of a dyad in a group, a pulling away of a patient in a therapy session, or the occurrence of manifestations of positive or negative transference are all very important because they signify boundary changes. These are usually the spontaneous boundary shifts and correspond to spontaneous morphogenetic movements. It can be assumed that active work has gone on preceding this development, because of the active transport boundary principle. In heavily defended boundaries, the system function of the therapist as an active transport mechanism becomes crucial. To hold a patient on a line—in, for example, dealing with sadness—requires active therapeutic work by the therapist and lays to rest, if it was ever believed, the concept that the role of the therapist can be a passive one. It is, of course, important to be aware of the type of active transport that maintains a rigid boundary and to conceptualize how such a self-perpetuating system can be interfered with, as we do when we investigate the active work a patient takes on to maintain an image of supermasculinity. So, the general systems principle here is that boundaries are to be recognized, their function understood, and ways of changing them conceptualized, with the rather glorious understanding that boundary changes mean system changes, thus giving a sort of short cut to our understanding of how patients change. This corresponds to the Bertalanffian principle devoted to the organismic function of the human being and to the second Bertalanffian principle of primary activity.

The third aspect of the general systems view of personality that has recently intrigued me is consideration of whether a particular human being has "systemness" qualities of sufficient degree. The equivalent sociological problem is whether a particular institution has sufficient qualities of "psychiness," a term that probably means the particular system qualities of the individual, for, if supersystems and systems have not the qualities of reaching out and meeting each other, then, by and large, the world and its future are in trouble. It is helpful to think of psychopathology in terms of systemness, because it serves us as a guide to what changes and developments are necessary for our patient to become a person among people or an individual in an institution. It is an amplification of the psychoanalytic view of object relations, which deals with the openness of system attachments of one person to another, but it suggests that more direct approaches can be made to its solution by encouraging people to say "hello," through group therapy and other measures by which the systemness qualities can be measured. It is a simple but profound insight of the systems view of personality that says that a system is the first necessity for functioning in a system world.

A natural follow-up to this point is the study of information theory as a model for what human functioning is all about. As J. G. Miller points out, what systems process is matter, energy, and information ("Living Systems: Basic Concepts" in W. Gray, F. J. Duhl, and N. D. Rizzo, eds., General Systems Theory and Psychiatry. Boston: Little, Brown, 1969.) And it is information that leads to patterning, and patterning is what systems are all about. A prepublication paper by G. Swogger, Jr. (1973) deals brilliantly with the application of information theory concepts to group process. The function of a group can be understood in terms of the competing information programs being processed, in terms of the ability of the group to come to a more unified information processing system, with the development of codes in the form of group norms, and in terms of the way in which the interplay between the group norm and the residual
individual norms sets up a two-factor morphogenic process, as described earlier. The concepts of input, throughput, and output are congenial to an information processing view, and these are, of course, basic observables about human function. The processes of positive and negative feedback are also congenial to such a view. Less well known, but crucial to a systems view of man, are the concepts derived from information theory, that it is the internal coding of a system that determines its stability and determines its potential for existing in the world, the latter depending on there being a considerable isomorphism between the internal codes and the codes of the environment. Here is the lovely insight that meaning, or understanding, or ability to coexist depends upon the ability to establish isomorphism between the two codes. This is what occurs when we say to another person, "Ah, I see what you mean," or when a person says to his world, "We are alike enough so that I am able to exist." It also provides the insight that newness and sameness are both essential qualities in our relationship to our environment and that there are optimal levels of each. Meaning can also be understood as the digestion of newness into sameness, and it is well described in the adaptive self-organization diagram of E. Laszlo. (System, Structure and Experience, N.Y.: Gordon and Breach, 1969). It is easy to see that, if sameness is total, stagnation results, while total variety or newness would simply be not understandable. Again, we are back to the fundamental nature of stability and change, or homeostasis and morphogenesis. It is intriguing indeed to view life itself as a variety processing system in which growth, development, and evolution are part of the natural laws of our universe. This brings us around to the fourth Bertalanffian principle.

The next topic in my more recent views of how to establish a general systems theory of personality deals with the topic of transformation. Part of the problem is not difficult, for we have already been talking about slowly occurring changes that take place as the result of the operation of sameness to difference principles. The intriguing part, however, is those phenomena included in the more formal theory of transformation; that is, those situations in which sudden and often dramatic shifts in system functioning occur, often without input. The classical example is that suddenly you learn to ride a bicycle, or suddenly you understand democracy, or suddenly you undergo a process of conversion, in this case with the transforming process continuing over a period of time, apparently without further input.

I have approached this topic from a different point of view. What I am really intrigued with is the problem of what is the nature of the relationship between emotions and cognitions; for, without such knowledge, our dreams of a general systems theory of personality will fade in the dust. One could, of course, include action and conation, or the problem of will power, as elements that need to be included and integrated in a system view but, since I consider that emotions organize cognitions into thought and that thought governs conation and action, I am satisfied with my own views. In brief, my views are that emotions were the original orienting system; that is, that one behaved or concluded as the result of a feeling assessment, and that what has happened is that in evolutionary development our feelings, which used to organize our conception of our environment, now organize our cognitive fragments into wholes or parts. This model of thinking would postulate what is ordinarily thought of as creative thinking as the paradigm, for here it is most clear that the thinking process originates with an emotional process. A real thought, then, would have an emotional precursor, with the general paradigm being that an emotion is felt deeply and intensely, is then encountered and recognized, following which it undergoes processes of modulation, differentiation, and precisng, reaching a stabilized nuance form during a meditational phase, and, in this state, the organization of cognitive fragments into a thought or gestalt takes place. Further references to this can be found in W. Gray, "Emotional-Cognitive Structuring: A New Theory of Mind" in FORUM for Correspondence and Contact, Vol. V, No. 2, Oct., 1973, published by International Center for Integrative Studies, N.Y.; in W. Gray, "Emotional-Cognitive Structuring" in General Systems, Vol. XVIII, 1973, and in a paper presented at the American Group Psychotherapy Association meeting in New York in February 1974.

In summary, what I have to say amounts to a plea that we desperately need a general systems view of personality an intrapsychic general systems theory, as it were, so that the individual has an opportunity to survive and to grow in our increasingly system-oriented world. This is epistemologically correct, for the evidence is incontrovertible that nature and the universe operate according to systems principles, and it is inconceivable that we as individuals should be denied this natural and shared mode. It is here that going beyond mechanistic systems into humanistic systems, morphogenic systems, and systems that have an evolutionary future is so important, and it is this that I have tried to outline here. The fact that we operate under constraints is not the issue, for, without constraints, patterns could not endure. The crucial issue is, rather, that constraints are necessary for morphogenesis, and the crucial issue for us as human beings and psychotherapists is to see to it that we develop a theory and ability to deal with the issue of constraint in such a morphogenic fashion. The beginning of this will be our willingness to develop a general systems theory of personality.
THE CONCEPT OF STRESS
IN RELATION TO THE DISORGANIZATION OF HUMAN BEHAVIOR*

Sir Geoffrey Vickers

It seems clear that at present the phenomena which the psychiatrist studies are more readily comparable with those which concern the student of animal behavior than those which concern the physiologist. The animal experimenter, like the psychiatrist, starts from an observed correlation between a "situation" and a "disorganization of behavior." The disorganization may take many forms. The unbearably frustrated man may break down into rage or tears or lethargy; he may begin to act at random or he may begin some apparently irrelevant behavior or he may cease to act at all. And so, within its compass, may the unbearably frustrated rat. Differences in the form of breakdown are interesting as pointers to differences in temperament but they are irrelevant as indices of disorganization.

How then are we to recognize this characteristic disorganization? Shall we take as our index the fact that the behavior under review has become "non-adaptive"—inept to the situation which has evoked it? The judgment may be only superficially true. Neurosis may be protective in man, perhaps also in other animals. In any case, is it really "adaptive" for the man in the bewitched cockpit or the rat in the bewitched maze to go on forever ringing the changes on a repertory of responses which have clearly become irrelevant? I recall the comment of a psychiatrist when a colleague had described some particularly devilish design of animal frustration. He asked—"What would be a non-neurotic response to a situation like that?" and I do not recall that he received any convincing reply.

It would seem that what appears as disorganization may in fact be defensive reorganization. Some kinds of psychological stress response appear to have a protective value. Yet this protection is bought at a cost. Field Marshall Lord Wavell, in some famous lectures in the last war, pointed out that stupidity in generals should never excite surprise, since they are chosen from the exceedingly small class of human beings who are tough enough to be generals at all. Their essential qualification is that they should continue to function, even if not very well, in conditions of stress which would cause less stable organisms to break down. His point seems equally well made whether by "breakdown" he meant "disorganization" or "protective reorganization."

I remain in doubt about the criteria for determining disorganization. They seem to imply the passing of some threshold beyond which external relations can no longer be handled at the previous level, either because internal coherence has been lost or because it has only been preserved at the price of some withdrawal. For the purposes of this discussion, however, it is sufficient to raise the question of their definition and significance.

If pressed next to define the situations which provoke such disorganization, both the psychiatrist and the student of animal behavior would be driven to some tentative formulations. The animal experimentalists in particular could not refuse to formulate the assumptions which govern the design of their experiments. Pavlov, I understand, used four main types: one, a progressive increase in the intensity of the signal to which the animal was conditioned; a second, a progressive increase in the delay between the giving of the signal and the arrival of the satisfaction to which it was attached; a third, confusion, by the introduction of anomalies in the signals themselves; and a fourth, interference with the animal's physical condition. Leaving the fourth aside for the moment, it seems useful to subdivide the third, in the light of other experimental work, into at least three subclasses—doubt, where the signals are ambiguous; conflict, where they are mutually inconsistent; and frustration, where response to a clear signal does not produce its accustomed result. Professor Liddell's paper names several other situations, including loneliness, monotony and self-imposed restraint, which may need categories of their own.

It is interesting that the factors common to these animal experiments can be most readily described by words drawn from the subjective vocabulary of humankind—apprehension, suspense, doubt, conflict, frustration. Psychiatrists, I think, would agree that these categories also cover the situations which they recognize

as provocative of disorganization in man. This suggests that these states, which we know by introspection, reflect conditions of neural excitation which are not confined to conscious states or to nervous systems so complex as our own.

Furthermore, many of these states involve the concept of expectation, whether it be the hateful expectation which can be neither accepted nor escaped, the pleasurable expectation which is intolerably deferred, the conflict of expectation which evokes conflicting responses or the sheer confusion of expectation which destroys the basis for action. The frequent presence in experimental situations of something which we can only describe in terms of expectation suggests that this also is something we need to be able to describe in terms of neural activity not necessarily related to conscious human states.

At this point the psychiatrist and the student of animal behavior may legitimately seek the help of the physiologist. The organism is not functioning as it was. What is this change in its manner of functioning? How and why has the change occurred? Proper questions, surely, to put to a specialist in organic function.

Until recently the physiologist would have had little to say. He could supply an impressive model of the internal relations of the organism. He could describe the homeostatic mechanisms which neutralize the impact of external variables, such as heat and cold and the course of their defensive activity, as they are overcome by forces too strong for them—forces which, following Dr. Selye's classic formulation, he knows as stressors. He could describe the adaptive changes which enable the organism to hold its own for a while against such forces—the changes by which, for example, a man chased by a bull mobilizes for a few minutes an abnormal amount of energy. But why should the approaching bull function as a stressor, no less than a fall in temperature? Why, stranger still, should the hope of winning an Olympic mile function as a stressor, no less than an approaching bull?

These were questions physiology was not prepared to answer. Nor is it clear that the collapse of the runner after he has leapt the gate or burst the tape has anything in common with the breakdowns studied by the psychiatrist and the student of animal behavior.

Clearly the stressful situation which concerns the psychiatrist is given not by events but by the organism's interpretation of events in relation to itself. This in turn is a function of the way in which the individual personality is organized; and any model of this must include the organization of experience.

We have several conceptual models of organization. We have the physiological model; but this until recently had little to say about the organization of experience. We have the body of psychological concepts dealing with perception and learning theory. We have the multi-dimensional models implicit in psychodynamic theory. All these models deal with systems constructed by abstracting different sets of variables from the bewildering, observable entity, man. Their better integration is an urgent need.

Some recent developments already make for better integration. The anatomist finds in the nervous system paths by which past experience may modify the sensory input even before it reaches the associative areas. Observations with the electroencephalograph may provide a physiological measure of some aspects of personality and personal integration, and have suggested to Dr. Grey Walter a physiological theory of learning. Both psychiatrists and animal experimentalists are adding to the correlations, hitherto curiously rare, which relate physiological and biochemical to psychological changes at the threshold of breakdown: and pharmacologists are placing at their disposal an increasing repertory of drugs of known composition which have predictable effects on mental state. A genetic difference has been shown by animal breeding to underlie variations in vulnerability to stress. And animal ethologists like Mr. Tinbergen have provided us with language and concepts which everyone from the physiologists to the psychoanalysts seems to find acceptable.

A further important factor making for integration is the current development of language and concepts apt for describing open systems generally. The concept of stability, for instance, which Dr. Ross Ashby has done so much to generalize seems equally applicable to the systems studied by the physiologist, the psychiatrist and the sociologist. The same concept is applicable to the much more numerous cases in which the governing controls of the system are not fixed but change with time, such as the pattern of growth and maturation, or a changing cultural norm. Homeostasis is a special case of the much wider process which Professor Waddington has called homeorhesis.

One idea, which recurs in all these fields, may be of central importance in the study of stress. It is the idea of matching with a pattern. It seems clear that the raw material of experience is not the whole of the "blooming, buzzing confusion" which beats on us but a selection of the regularities which we can detect; and that these in turn provide the categories for the classification of further experience. Conditioning—and perhaps more beside—depends on recognizing regular relations between recurrent events in the categories thus distinguished. Judgments of value—most refined of tropisms—are linked to situations recognized by their correspondence to some pattern, however complex.
Thus if I say that A has given me a fair deal, I must first have selected for attention a particular group of features of my relations with A; I must have judged these to be a deal, rather than a fight; and applying the standards of fair deal, rather than fair fight, I find them to match. The evaluative judgment "fair" no less than the cognitive judgment "deal" is an act of classification, of matching with a pattern. Thus pattern governs throughout. The mind may boggle at the thought of a black box no bigger than our heads which can group, regroup and handle shifting configurations of symbols so complex, so plastic and yet so enduring; but we boggle rather at the complexity than at the principle.

It is not surprising that expectation should figure so largely in the organization and the disorganization of personality; for what we know consciously as expectation is only the exposed part of an iceberg that floats very low in the water. In pursuing, maintaining and eluding the external relations by which we live and die, we are guided by symbolic representations of what is happening and what ought to be happening. We need a model of this process to understand the nature and noxious operation of stress; and contrariwise, in seeking to understand the working of stress, we may well contribute to our understanding of controlled behavior. Organization and disorganization are opposite sides of the same penny.

Meaning derives from relationship to the familiar. When experience does not provide us with a templet, we have to invent one. This is exemplified at the conscious level by another model, with which I happen to be familiar; and I will venture to describe it briefly, because it presents a clear and simplified picture of control by expectation.

In the practical affairs of life we assume that conduct is controlled to an important degree by structures of expectation; and the implications of this have been worked out both in theory and practice with a high degree of refinement in the field of administration. No one today would try to run a business without maintaining and continually comparing two running representations, both projected into the future. One of these is a representation of what is happening and what is likely to happen next. The other is a representation of the course of events which we want to bring about or prevent. I will refer to these as the "actual" and the "standard" or, sometimes, as the "is" and the "ought to be." The comparison of the two representations yields a stream of mis-match signals on which we act.

Thus the building contractor plots on charts against time the planned course of many interdependent operations; and, as work proceeds, he plots against these what is actually being achieved and projects these trends also into the future. Divergence of the two is a signal for action to bring the actual into line with the standard; or, if this proves impossible, to revise the plan, so as to provide a realistic and hence an effective control.

This homely illustration has features which are equally relevant to individual organization.

The controls used in business are representations of relationships which we seek to maintain, to alter or to escape. Some are relationships between the organization and its environment, such as the rate of intake and the outflow of money, materials and men; we might call these metabolic relationships. Some are relationships within the organization itself. In either case the control may be negative or positive; it may be directed to bringing the relationship continually back to some optimal position or to preventing it from straying beyond some critical threshold. And it must represent in the same code both the actual and the standard; it becomes meaningless if either term is absent.

Thus what I have called the actual is a highly artificial construct. It is hypothetical; for our information about what is happening is never complete or exact or direct. It is selective; for we can attend only to a few aspects of it at a time. It is represented in a code, be it in writing or figures or graphically, which limits and distorts what can be represented; and this is necessarily the same code in which we represent the standard—how else could we compare the two? Finally, it has an inescapable time base. In all this, I suggest, it closely parallels the working of the individual mind.

The controls used in business, like those of the individual mind, do not always give clear or correct or adequate guidance. First, they apply only to those aspects of experience to which we have chosen to attend. We may have chosen the wrong variables for attention. Again, the signals may be ambiguous; or, being clear experience may supply no apt response; or, responses proven apt in the past may let us down. But their most inescapable embarrassment is their conflict. It is a feature of practical life that any projected action is relevant not merely to the purposive sequence in which it arises but to many others also; and it cannot be equally congenial to all. If apt for some, it will be inept for others. Whatever doors it opens, others it will shut. In business, short term profit, long term stability, internal coherence, public relations—each of these disparate standards, when compared with an appropriate selection from the actual, maintains its own stream of warning and advice; and these are no more consistent in business than in private life.

Moreover, their inconsistency is inescapable. The highly organized business, like the highly organized personality, necessarily
generates more inconsistency in its governing expectations than one less highly organized. This curious fact, which deserves closer analysis than I can give it now, may account for the fact that stress of war produces relatively so little civilian neurosis. Life in wartime may be harder but it is also simpler.

Apart from the inadequacies already described, the controls of business are themselves partial and intermittent; for either the actual or the intended may be inaccessible. Like the fogbound navigator, we may know where we ought to be but not know where we are; or, like the climber following an unfamiliar route, we may know where we are but have no assurance that it is where we ought to be. And experience cannot be guaranteed to remedy either case; for the results of our actions may return for judgment after so long an interval and with so large an admixture of other variables, that they provide neither validation for the past nor guidance for the future.

In the board room, then, stress is associated with doubt and conflict of clearly definable kinds. We may have to live for years with deafening streams of mis-match signals, either because we can devise no suitable response or because the action they invite, if taken, would elicit even more violent protests in another context. Alternatively, we may have to live in an eerie silence, because either the actual or the standard is not registering in the appropriate code. And either of these states may lead to disorderization (or protective reorganization) suggestive of the forms familiar to the psychiatrist and the animal experimentalist.

Far flung analogies deserve to be considered with suspicion; yet this one seems to me to be useful in three ways.

First, controls of the sort which I have been describing seem to be closely related to controls at simpler levels. The pecking response of the herring gull chick is pre-set to be released by information sent in a particular code. The chick responds to the red patch on the feeding beak—which is itself an abstraction—because it corresponds to the pre-set signal, a correspondence not absolute but falling within limits of tolerance which are themselves built in. Learned responses can, of course, become attached to much more elaborate cues, controlled by patterns distilled from past experience. It would seem then that the process of matching with a pattern is inherent in the control of behavior from the simplest cue-governed response to the most sophisticated act of cognition, for all cognition is recognition. Moreover, the patterns which our brain can record are patterns in time as well as in space. An instrument which can symbolize change with time can presumably represent the future as easily as the past.

Then, secondly, the model seems useful to me in its representation of conflict as endemic and necessarily increasing with more complex organization. By more complex organization I mean not increase in size or increase in the power and variety of responses available or even increase in sub-division of function. I mean increase in the number and diversity of objectives to be sought and thresholds to be avoided, of those positive and negative standards by which behavior is regulated; in other words, in the norms which the system is set simultaneously to seek and the limits which it is set simultaneously to avoid. This is the dimension in which, as I see it, organisms and organizations tend to develop and in which they tend to fall back whenever they set themselves a task of reconciliation which proves in face of events to be too much for their powers.

Thus, thirdly, the example is useful in stressing a new dimension in which the higher organisms, no less than organizations, are adaptable. Most biological work treats as given the acceptable and unacceptable states which act as governors, positive and negative, of a system’s behavior; and in so far as this is so, the only scope for adaptation within the individual life span is the development of responses and skills to serve these needs. But at higher levels and conspicuously at the level of human life, the individual’s wants and needs, no less than his responses and skills grow and multiply within his life span. Many different but mutually exclusive possibilities compete for realization within the framework of the biologically given. We win increasing—and embarrassing—scope in the setting of our own systems.

So the psychologist needs a model of organization more complex than would meet the needs of the physiologist. The biologist, at grips with the code built into the structure of the gene, is seeking the key to a program. But the psychologist needs a model not only of the organism’s program from birth to death but also of its self-programming capacity and its self-programming propensity; a model which will represent our goal-setting as well as our goal-seeking. For though these two activities are so closely related that they cannot be considered separately, they are also, as I believe, so distinct that they cannot usefully be simplified by resolving one into the other.

Finally, I remind myself that a model of conflict does not necessarily tell us anything about pathologic stress; for we cannot assume that conflict in itself is noxious. Conflict-solving is the normal activity of our brains. And if the threshold is quantitative, it is also relative. We need to understand both the noxious nature of a given stress in relation to the organization of a particular individual and the vulnerability of a given individual to a particular situation.
THE CONCEPT OF STRESS

You may feel that a model of conflict in such conscious and cognitive terms as I have used has little bearing on our subject matter. In fact, of the governors of corporate behavior, some of the most powerful of them seldom register in the board room and it is a major task of administration to bring them to the central consciousness—often despite the censorship of powerful repressing mechanisms. But I will not pursue the analogy any further. If it seems an unduly cognitive approach to a problem essentially dynamic, it is at least in tune with the times. When men's basic conceptual models were mechanical, people seeking the explanation of some happening looked first for the force which made it happen. In these servo-mechanical days we seek first not the force but the program, not the source of energy but the source of information. The world is full of energy, waiting to be borrowed; the weakest signal may release force to move a mountain. So naturally interest centers on the signalling system. And it is here, it seems, that the development of our mental powers has endowed us with so alarming a capacity for organization—and disorganization.

REFERENCES

F.

HEALTH AND STABILITY OF ORGANIZATIONS
Social Systems Analysis and Industrial Humanism: Awareness Without Revelation*

Robert A. Smith, III

Worries come only if you believe in a future. Believing in ghosts, you are freed from burden, all the worries enter the ghost's head. Yet the ghost has no head, so it does not suffer from headaches either, not for anybody's sake.

Those who out of sheer wrong habit still attempt to think for themselves, get their ears boxed by the ghost. From this ghostly boxing you can neither free yourself nor can you escape it; against it is neither appeal nor any judgment at all.

The entire population, ghost ridden, now walks with eyes shut. "The most ancient form of movement this, with eyes shut," the philosophers assure them.—Rabindranath Tagore

I believe in a future and agree with the philosophers also. I am willing, however, to get my ears boxed by bureaucratic ghosts and to attempt to remove the conceptual organization cadavers we find stacked against the doors leading to the future. I suspect my outlook is closest to Nishida, that wise old Japanese philosopher, who had a movable synthesis of life where the lived past and the expected future constantly form the present. My topic will address itself to this forming present.

The state of flux in which we find ourselves today—no matter how isolated from it we sometimes hope we are—affects us both in the individual and social sense. Integration/disintegration, operationalism/intuitionism, capitalistic individualism/mass society, separation of power/centralization of power, youth/elderly, media/message, and so on. These are forces we must reckon with and more especially as they are distributed between different groups in our society and in such a way as to focus anxiety and aggression along singular structured lines of tension. We are not so much concerned now with flotsam and anomie as we are with the processes of individuation and socialization and upon polarizations which correspond rather closely to structured differentiations in society and which feed more fuel to the flame of collective narcissism. Our responses will depend very much on the world view we cultivate.1 It depends strongly on whether we view the world mechanistically as so many individual micro-units in competition and senseless conflict or whether we view the world organically as a giant micro of large sub-systems with a natural interdependence we prevent from functioning. Without this type of focus, I do not believe that we can understand behavior in organization and, because of that, I am purposefully taking a holistic perspective. This perspective, I believe, will help in treating more effectively the dysfunctional and dissenual aspects of organizations and will show how conflict can become purposeful rather than divisive, that competition using democratic sportsmanship is healthy and without it unhealthy, and that meaningful integration is essential for required differentiation in modern complex organizations.

I am going to discuss two significant forces which have affected behavior, structure, power equalization, participation, leadership styles, and so on. My time span will be the last two decades—the actors, the audience, the chorus, and the orchestration which have taken place. These past two decades have been the time womb for the birth and dynamic growth of twin forces which have profoundly dealt with the character of man and the character of his organizations. Until very recently these forces have operated as if they were born of separate parents. I intend to demonstrate how they must synthesize in the organizational setting be it university, federal, or industrial.

What are these forces? Group dynamics and project management. A brief historical synthesis of their genesis, growth, lack of interaction, and finally, a growing interaction can set our perspective.

Of course, as Reinhold Neibuhr maintained, "no new fact or event in history is an arbitrary

---

*This paper was first presented to the meeting of the Academy of Management, and appeared in The Academy of Management Proceedings, Dec. 1968, pp. 11-23. It is reprinted here by permission of the publishers.

novelty. It is always related to a previous event. But it is a great error to imagine that the relationship completely accounts for the new emergence...each new thing is only one of an infinite number of possibilities which might have emerged. So forgive me if I begin as though the forces emerge as if they are brand new without antecedents to the past.

Group dynamics was the outgrowth of collaborative efforts of Lewin, Lippitt, Benne, and Bradford. After its beginning in 1947, it was institutionalized in the National Training Laboratories and subsequently in other organizations such as the Western Behavioral Science Institute, Blake's Scientific Methods, Inc., the Esalen Institute, and others. Initially, it involved sensitivity training for small groups (12 to 15) in off-institutional settings for periods of one to three weeks.

With apologies to Cleland, Baumgartner, Ryan, and Steiner, since they cover the subject in depth, project management began in the 1950's with the Navy Polaris Program which was a complex program consisting of the development and production of the warhead, the missile launcher, and the submarine carrier through prime and subcontractor structure. I selected Polaris as the project management genesis because of the introduction and use of the Program Evaluation and Review Technique (PERT), the beginning of systems analysis and cybernetic information technology as tools in project management. Basically, PERT is the design of the man-machine interfaces from the design to operational use schedule of a weapons system. From this beginning, PERT and Project Management emerged as an integral management force throughout DOD, NASA, and the Aerospace industry. I think we can perceive immediately why the forces of group dynamics and the forces of project management tended to go their separate ways. Group dynamics tended to have an organismic orientation, and project management, a mechanistic orientation. While both involved interdisciplinary approaches, each was comprised of opposing interdisciplinary clusters as it were. Group dynamics clustered around the behavioral sciences—sociology, psychology, economics, and political science. Project management clustered around the physical sciences—mathematics, physics, and industrial, mechanical, and electrical engineering. So you see, from the beginning, there was a somewhat natural polarization.

From where would the catalytic force come to bring group dynamics and project management together? I think the impact of scientific democracy, my own term for describing the scientific as opposed to the engineering environment, provided the answer. Science must operate under rules which provide for meaningful dissent so that new innovations can replace worn-out truths and axioms. Science operates even better when there is democratic sportsmanship prevailing. This permits the best in human resources to draw together in mutually satisfactory efforts. Why?

In its early stages project management tended toward complete autonomy as if a project were a permanent fixture. This proved to be self-defeating to the institution housing it. For, in reality, projects are mortal and have a definitive closed system life span. An institution which "puts all its eggs in one basket," the project, dies when the project is completed. Unlike the project, the institution tends toward immortality, and to maintain its viability and the image of immortality, it must operate as an open system with the larger society and environment from which it gets its input and gives its changing project outputs. Consequently, institutions became wary, gave less autonomy to project structures, and revitalized their functional departments. Project Management, however, remained the focal point of adaptability for the institution and continued to assure that project goals were accomplished on time, within the budget, and met performance requirements. Functional departments, in the meantime, began incorporating newer disciplines to support the changing project objectives and to extend and maintain the institutional vitality. What was coming into being is best described as a matrix organization with its loose structural arrangements which permitted the institutional great leeway in adaptability. The matrix organizational arrangement is best depicted through linear responsibility charting or tableau design. Matrix could be simply described as a network arrangement for bringing together the man-to-man, man-to-machine interfaces throughout all the subsystems of an organization with those of the larger society. It is a network of the subsystems and, in complex organizations, is a series of matrices coupled to the hierarchy of work efforts.

At this point, I think it well to discuss the catalytic or, better put, synergistic forces which are drawing the systems sciences of project management and behavioral sciences of group dynamics into a total approach to organization. Bertalanffy, Boulding, Gerard, and Rapoport may be considered the founding fathers of general systems theory. General systems theory involves itself with holistic perception and holistic integration. The growing effect of specialization, division of labor, competitive autonomous functional areas—even though it was recognized that specialization was essential in modern complex institutions—led to an insistence by the general systems theorists that integration


530
and organizational conceptualization were indispensable. In other words, they were seeking to overcome a general view of the world as organized fragmented chaos by presenting it as one of a natural organized interdependence. Although general systems theory was seized as a panacea by cyberneticians, information systems theorists, decision and game theorists, and operations research analysts, it was not until the second generation came forth that more practical orientations of general systems theory were made. Chief among this "new breed" are Walter Buckley, Albert Biderman, Richard Ericson, Yehezkel Dror, Kenneth Berrien, Keith Caldwell, Robert Boguslaw, Aaron Wildavsky, Raymond Bauer, Bertram Gross, and Stanley Young. They seem to insist that although an organization is comprised of interrelated variables, it is an error to presuppose that these variables are of equal weight. Ericson points out in a recent paper that the systems sciences, through their capability of providing models for heuristic analogy, may help close the gap between the social responsibilities as perceived by an institution and the social values of the larger society. The point Ericson is making is that social change is so rapid and so dramatic that institutions don't have time to "institutionalize their peculiar hypocrisies" but must remain closer to the true social reality. With this brief discussion of general systems, we can begin seeing the institution in its organic setting of individual participants.

The micro individuation processes of the founding psychiatrists—Freud, Jung, Adler, Rank—finally found focus for macro individuation processes in Harry Stack Sullivan. R. D. Laing, Bion, and other members of the Tavistock Institute appeared to have made fruitful uses of Sullivan's holistic approach to interpersonal psychiatry before their American counterparts, although Goffman and Bales have done some excellent research. More recently Arieti has forged some links between general systems theory and intrapsychic forces. This is now culminating in transcultural developmental psychiatry and psychology being increasingly influenced by social psychology as represented in the efforts of Reza Arasteh, Byron Fox, Merton, Hozowitz, Homans, Sanford, Inkeles, Clau- sen, Brim, and Eleanor Maccoby.

In short, the processes of individuation and socialization, through the holistic processes of general systems theory, have, while expanding the boundaries, found realistic limits. It would appear that continuing research and application of general systems theory would be fruitful for the university, federal, and industrial institutions.

Nevitt Sanford states the case quite well: "Since we deal with a dynamic interaction between personality and social systems, it may be necessary to change certain personalities in order to change the social system. Individuals use their social roles for the expression of their personality needs; hence a change in organizational structure may be resisted by individuals in the same way they resist change in internal adaptive devices that have been found more or less satisfying. The practicing social scientist needs to be familiar with personality dynamics"; I would add social organizational dynamics.

Because we are all too prone to limit our thinking to the line/staff dichotomy, I would like to discuss briefly the trichotomy present in all complex organizations—the institutional, functional, and programmatic trichotomy. The diagram is a simple illustration of the power struggle inherent in any complex organization. By treating the trichometric power struggle, the matrix organizational arrangement attacks the heart of entrenched power clusters with their concomitant structural and stratification impasses. It goes beyond the autonomous project management organization, which did indeed overcome the impasse of lateral authority invested in functional departments, by treating the organization in an organic rather than organismic or mechanistic sense. Matrix organization may well be the first step toward industrial humanism.

While project management was tending toward less autonomy, group dynamics began emerging from its micro cocoon and entered the macro world. Complex organizations could not be changed by the micro changes. Macrosociology, general and open systems theory, and more informed knowledge of the character of complex organizations strongly influenced group dynamics practitioners to extend their boundaries. Second generation behavioral scientists were learning that, despite the dramatic changes in individual personalities perceived in the small group sensitivity sessions in the "castles" away from home, these changes were minimal in influencing organizational behavior back home. Participants who had developed higher levels of awareness, greater self-esteem, and new role expectations found the same climate, the same structure and stratification, and role expectations and role distance remained. In other words,
individual norms were changed while organizational norms were status quo. Subsequently, the individual restructures his role to conform with prevailing norms or becomes an outsider whose talents are lost to the organization. How was the individual/organizational conflict overcome?

Some institutional change agents working with outside consultants designed laboratory sessions to provide meaningful ways of influencing organizational change. These included diagonal slices (persons of different levels in different parts of the institution), instrumented labs, family labs, and corporate hideaway meetings. As examples, the Blake Managerial Grid, incorporating McGregor's Theory X/Theory Y, was implemented throughout institutions; Davis, of TRW, using a combination of inside change agents and outside consultants, designed family and instrumented labs and a matrix organization; and others used various combinations of approaches. The operational and intuitive approaches to group dynamics drew closer together. As a result, both personality and institutional variables were considered and the climate and character of institutions were more easily changed. Both the personality and the institutional structure benefited from a new awareness of what was involved. The ping-pong effect on observers and the polarization of participants could now be resolved.

Now I would like to turn my attention to another force factor which will bring group dynamics and project management processes into partnership—the force of rapid change. Until recently, organizational theorists and behavioral scientists concerned themselves with organizational dialectics and adjustment psychology or the processes of equilibrium rather than change. In other words, concern was directed toward adjustment and adaptation to environmental conditions, when in reality the constant socio-cultural evolution to higher forms of organizations begs a more dynamic approach. Many of you will recall Lewin's famous

unfreezing-change-refreezing or equilibrium model for changing individuals. Many of you will also recall Blau's masterful portrayal of the equilibrium processes involved in organizational dialectics. But neither Lewin nor Blau developed mobile steps for change to take care of emerging paradoxes. Alvin Gouldner's brilliant discussion of the norm of reciprocity in effect was another equilibrium process. I am not trying to infer that equilibrium or maintenance is not essential. What I am inferring is that existing organizational arrangement is temporary and must be adaptable to rapid and continuous change.

The rapidity of change introduced by technology makes it imperative that the management and behavioral scientists develop a technical literacy so as to understand the growing interdependence between social and technical systems. It also means to me that the universities should begin designing real-life research using interdisciplinary team efforts of both the technical and social science graduate students. (NASA has taken some steps in this direction with faculty, graduate students, and co-ops.) Not only would this tend to minimize the technical/nontechnical dichotomy on the job but serve as well to bring the best resources together for job accomplishment during temporary partnership in institutional environments. It seems to me that this is the best way to recognize and use human potentialities beyond the robot and the ape so much in vogue today if the nonfiction "top ten" book lists are accurate. It also seems to support a truth: experience is not an experience unless you experience it. The universities are being challenged to provide this experience.

I know that by now you detect Warren Bennis' influence. He observes that "because of the turbulent and uncertain environment and the changing markets and goals of organization, the social structure of the future will have some unique characteristics. The key word will be 'temporary'; there will be adaptive, rapidly changing temporary systems." Bennis believes that, "They will be organized around problems to be solved...by groups of relative strangers who represent a set of diverse professional skills...[operating] in essentially professional organizations...developed along organic rather mechanical lines, meaning that [groups] will evolve in response to a problem...
rather than be programmed without reference to specific tasks." Bennis states that "People will be differentiated not vertically according to rank but flexibly according to skill and professional training... Social structures will no longer be instruments of repression." In this respect, Bennis supports Fremont Shull's premise that matrix organization is indeed an evolution beyond bureaucracy, for Shull's matrix organization is responsive to the requirements set forth by Bennis for organic-adaptive structures.

Since I have made several references to differentiation, I assume that I should dwell briefly on differentiation/integration as an ongoing essential process within an institution and how it relates to individualization/socialization of the individual. I am convinced that the differentiation/integration process resulting from division of labor, specialization, and functionalization initially assumed a mechanistic posture in the bureaucratic setting. It was a differentiation/integration which tended to be instinctively fixed as in insect societies where, in many respects, the smoothness of social arrangements, lack of conflict, social problems, and almost a dearth of ambiguity far surpass human societies. For man, whose symbolism is ever evolving, this created real dilemmas upon his individualization/socialization process. Man is not a creature suited to perfection—he always seeks completion and establishes new boundaries in his visible evolution. He is something that Robert Frost said "that doesn't love a wall." Yet between his delirium for freedom and his desire for order, he creates his organizations and then becomes a neurotic victim of machines he forgets he made.

In describing differentiation/integration, I assume Lawrence and Lorsch's definition of organization "as a system of interrelated behaviors of people who are performing a task that has been differentiated into several distinct subsystems, each subsystem performing a portion of the task, and the efforts of each being integrated to achieve effective performance of the system." They treat structure as referring to "those aspects of behavior in organizations subject to pre-existing programs and controls." Pre-existing structural arrangements tend to create differences in perceiving informal operational and formal authority boundaries when new programs are initiated. This tends to support Bennis and Shull's conceptualizations and Burns and Stalker's field observations that organizations with a low degree of formal structure could more profitably cope with changing environments than those which have a higher degree of formal structure. I conclude that what has happened in mechanistic differentiation/integration is that increased specialization has caused a subsystems breakdown where subsystems are in competition with each other, and in the extreme, in competition with the institution itself. Subsystems tend toward autonomy or become faceless replicas of the larger body in mechanistic, overly formal organizations. I think this also explains that when the formal and informal structure are close, the organization is generally healthy and functional; where they are far apart, generally unhealthy and dysfunctional. Perhaps it explains the ecological patterns of development in older institutions with a low turnover in their upper hierarchy for here, indeed, one can see the strong sense of ingroup membership who base their definitions of members on the differences rather than the similarities between themselves and other groups.

Cleland, in his papers on Project Management, discusses the effect of pre-existing structural arrangement on newly formed structures. He maintains that there is a de jure (legal) and de facto (reality) authority structure. De jure authority is based on organizational charters, organizational position, position description, executive rank, policy documents, delegated power, and the hierarchical flow. De facto authority is based on technical knowledge, maintenance of rapport, negotiation, focal position, informal organization, deliberate conflict, and ability to resolve conflict. While I agree that Cleland has correctly assessed the reality of project authority in a bureaucratic organization, I support the positions of Bennis, Mee, and Shull primarily because a matrix structure provides a natural arena for conflict resolution and consensual feedback and is more adaptive to changing institutional needs. The de jure and de facto authority merge as they should through consensual resolution. The prime reason for this is that a matrix structure is more adaptable to the open system by which an institution must operate than the closed loop process of project management where there is a definitized input and a predicted

13. Gary Richetto, who has spent one year at a NASA Center, has found evidence of this in research for his dissertation (unpublished doctoral dissertation, Purdue University), p. 28.
output. The norms of exception are as important to the institution as predictable norms are to project management.

Since I am pursuing the uncommon, albeit critical, social psychology of organizations by capturing the gestalt of data through synthesizing my selection of critical variables to behavior in organization, let me concentrate briefly on the interpersonal dynamics, the group or subsystems norms. Mills suggests the following subsystems:

1. On the level of behavior, the subsystem is the interaction, which is the organization of overt actions among persons over time.

2. On the level of emotion, the subsystem is group emotion, which is the configuration of feelings among members and of their emotional responses to events that occur.

3. On the level of norms, the subsystem is the normative system, composed of the organized and largely shared ideas about what members should do and feel, and how these should be regulated, and about what sanctions should be applied when behavior does not coincide with norms.

4. On the level of goals, the subsystem is the technical system, which is the set of ideas about what the group should accomplish and the plans about how it is to be accomplished.

5. On the level of values, the subsystem is the executive system, which consists of the interpretations of what the group is, the ideas about what would be desirable for it to become, and the ideas about how it might so become.

Mills maintains that a clear understanding of the complex interdependence of these levels is highly desirable so that the interaction gestalt is recognized. I think Mills has brought social psychology closer to management science and the physical sciences by providing a subsystem approach to understanding behavior. As he points out, some people, like some organizations, operate at all five levels; but many people, like many organizations, operate at less than five. It is not essential that all people operate at each level, but it is essential for the organization to do so to enable its members to feel and perceive the process of completion. I stress this point because the modern tragedy is the overpotency attached to technical agencies and the sin of hubris committed by many of their leaders. They understand and appreciate the "engineering of musical notes" but not the orchestration to bring together the agent (actor) and the acted upon (audience) into a chorus. Without such orchestration, a society or institution becomes biologically dysfunctional either through alienation of its members or through the formation of fragmented collective narcissism. An orchestra blends a harmony, not by excluding discordant notes but by including them in the proper places. Too rigid orchestration would be shallow just as the exclusion of all deviant behavior would make an organization faceless.

Blau's pregnant statement, "To administer a social organization according to purely technical criteria of rationality is irrational, because it ignores the non-rational aspects of social conduct," captures this tragedy of modern technical hubris. The doer must consider the "done to." Unity within an institution is not determined by its constitutional breakdown but rather by its sets of relations—its processes and not its products. Therefore our attention should be directed not so much to structure but to the forces which produce the structure and maintain it. Ralph Turner, as interpreted by Walter Buekley, supports this premise. Turner maintains that the key to role-taking is the morphogenic propensity "to shape the phenomenal world into roles." Buckley points out that, "Formal organizational regulations that restrict this process are not to be taken as the prototype but rather 'a distorted instance' of the wider class of role-taking phenomena. To the extent that the bureaucratic setting blocks the role-making process, to that extent is organization maximal 'variety' or alternatives of action minimal, actors cogs in a rigid machine, and the morphogenic process frustrated." The rapidity of technical, social, economic, and political change makes organizational behavior a morphogenic and not merely a homeostatic or equilibrium process.

Although I have not satisfied Warren Bennis' requirement for changing institutions—that one must develop a theory of social change for use of practicing social scientists—hopefully, I have developed from research and my own observation and experience, a theory of change for practitioners to develop practices of changing. Hopefully, I have outlined a gestalt of holistic orientation which treats living systems functions as integral units without regard for size or complexity so that larger areas of the person and of society may be examined through theoretical models using modern systems techniques which make this possible.

Being in full agreement with Maslow's hierarchy of individual needs (the micro needs), I believe that the organization or institution should, in the macro context, fulfill these needs individually and collectively. The path from the valley of basic needs to the summit of self-actualizing individuals and institutions doesn't call for adjustment—it calls for growth and growth dynamics. It is morphogenic and holistic in its processes.

This supports Katz and Kahn's theorization

that "organizational properties are by definition systematic and their change calls for system change." Kibbutz management offers some fruitful results in its treatment of organizational properties. The vested interests inherent in functional compartmentalization, the impasses of lateral authority conflicts over boundary proprietorship, and the stagnation caused by immobility have been overcome, to a large degree, through Kibbutzim methods. In fact, Kibbutz management attacks what Parsons would describe as unhealthy institutionalization of property rights in nonproperty activities through a constitutional arrangement of rotating its managers. In a recent article, Golomb provides an overview of Kibbutz management. He maintains that "The Kibbutz system tries to solve built-in dilemmas by means of system adjustment and social mechanisms that work." Of course, modification of Kibbutz ideals, developed in an internally open but singular and of necessity an externally closed society, is essential before they can be made applicable to our pluralistic society. Too, they perhaps have more significance for governmental organizations than for industrial organizations.

In conclusion, I feel that social systems analysis with its emphasis upon morphogenic processes, its consideration of the gestalt of data, and its scientific technique using human variables will move us closer to industrial humanism. It will help to develop a living communion rather than empty intellectual dialogue. We will then be able to develop our awareness and enjoy our revelation in the process. It may even help us to fulfill Arasteh's plea to design fully integrated institutions in which fully integrated people can operate. The humanist, the scientist, the engineer, the ecologist, the professor, the urban planner, the rural relocater, the poet, the philosopher, the cybernetician, and the politician all share in helping this to come about.

I would like to leave you with the words of McLuhan, who, in referring to accepted cultural approaches to problems, maintains:

The entire approach to these problems in terms of uniformity and social homogenization is a final pressure of the mechanical and industrial technology. Without moralizing, it can be said that the electric age, by involving all men deeply in one another, will come to reject such mechanical solutions. It is more difficult to provide uniqueness and diversity than it is to impose the uniform patterns of mass education; but it is such uniqueness and diversity that can be fostered under electric (cybernetics) conditions as never before.

REFERENCES


TOWARDS A "TRULY" SCIENTIFIC MANAGEMENT:
THE CONCEPT OF ORGANIZATION HEALTH

Warren G. Bennis

"Muggeridge: Now, Charles, you, because you're a scientist... you have this idea, as I understand from your writings, that one of the failings of our sort of society, is that the people who exercise authority, we'll say Parliament and so on, are singularly unversed in scientific matters.

Snow: Yes, I think this is a terrible weakness of the whole of Western society, and one that we're not going to get out of without immense trouble and pain.

Muggeridge: Do you mean by that that, for instance, an M.P. would be a better M.P. if he knew a bit about science?

Snow: I think some M.P.'s ought to know a bit about science. They'd be better M.P.'s in the area where scientific insight becomes important. And there are quite a number of such areas." (52)

Extolling Science has become something of a national and international past-time which typically stops short of the truly radical reforms in social organization the scientific revolution implies. Knowing "a bit about science" is a familiar and increasingly popular exemplar of this which C. P. Snow treats in his Two Cultures. (69) But if culture is anything it is a way of life, the way real people live and grow, the way ideals and moral imperatives are transmitted and infused. Culture is more value than knowledge ("a bit of science.") Dr. Bronowski, who shares with Snow the view that "humanists" tend to be ignorant of and removed from science (they cannot discuss the Second Law of Thermodynamics) understands more than Snow seems to that a fundamental unification of cultural outlook is what is required. (18) The connective tissue required, than, is cultural, social, institutional - not grafted - on evening courses on Science.

In this connection, and closer to some of the general aims of this paper, Nevitt Sanford has said:

The ethical systems of other professions, such as business or the military, have become models for whole societies. Why should not the practice of science become such a model? After we have shown, as we can, that joy and beauty have their places in this system? At any rate, anyone who takes it upon himself to be a scientist, and succeeds in living up to its requirements, may be willing for his behavior to become a universal norm. (61).

This foreshadows the general theme of this essay: the recognition that the institution of science can and should provide a viable model for other institutions not solely concerned with developing knowledge. To demonstrate this proposition, this paper first discusses the criterion problem in relation to organizations. An attempt is made to show that the usual criteria for evaluating organizational effectiveness, "enhancement of satisfaction on the part of industry's participants and improvement of effectiveness of performance" (36, p. 238), are inadequate, incorrect, or both as valid indicators of organizational "health." (For the moment let us use the term "health" in the same vague way as "effectiveness." Organizational health is defined later in this paper.) Next it is suggested that an alternative set of criteria, extracted from the normative and value processes of science, provides a more realistic basis for evaluating organizational performance. These criteria are related to those of positive mental health, for it will be argued that there is a profound kinship between the mores of science and the criteria of health for an individual. From this confluence is fashioned a set of psychologically-based criteria for examining organizational health. Finally a discussion is presented of some of the consequences of these effectiveness criteria for organizational theory and practice.

THE SEARCH FOR EFFECTIVENESS CRITERIA

There is hardly a term in current psychological thought as vague, elusive, and ambiguous as the term 'mental health.' That it means many things to many people is bad enough. That many people use it without even attempting to specify the idiosyncratic meaning the term has for them makes the situation worse, ... for those who wish to introduce concern with mental health into systematic psychological theory and research. (34, p. 3)

... no one can say with any degree of certainty by what standards an executive ought to appraise the performance of his organization. And it is questionable whether the time will ever arrive when there will be any pattern answers to such a question---so much does the setting of an organization and its own goal orientation affect the whole process of appraisal. (56, p. 422).

Raising the problem of criteria, the standards for judging the 'goodness' of an organization seldom

1. For the purposes of this discussion, "organization" is defined as any institution from which one receives cash for services rendered. This paper deals with all such supra-individual entities, although reference is made mostly to industrial organizations.

539
fail to generate controversy and despair. Establishing criteria for an organization (or, for that matter, education, marriage, psychotherapy, etc.) accentuates questions of value, choice, and normality and all the hidden assumptions that are used to form judgments of operation. Often, as Jahoda has said in relation to mental health criteria, the problem "seems so difficult that one is almost tempted to claim the privilege of ignorance" (34, p. 77).

However, as tempting as ignorance can be, research on organizations--particularly industrial organization--has heroically struggled to identify and measure a number of dimensions associated with organizational effectiveness (74). Generally, these dimensions have been of two kinds: those dealing with some index of organizational performance, such as profit, cost, rates of productivity, individual output, etc. and those associated with the human resources, such as morals, motivation, mental health, job commitment, cohesiveness, attitudes toward employer or company, etc. In short, as Katzell pointed out in his 1957 review of industrial psychology, investigations in this area typically employ measures of satisfaction and performance (36). In fact, it is possible to construct a simple twofold table that adequately accounts for most of the research on organizations that has been undertaken to date, as shown in Table 1. On one axis is located the criteria variables: organizational efficiency (the ethic of work performance) and member satisfaction (the ethic of "health"). On the other axis is located the two main independent variables employed, human and rationalized procedures. 'In other words, it is possible to sum-

Table 1

<table>
<thead>
<tr>
<th>Criteria Variables</th>
<th>Organizational Efficiency</th>
<th>Satisfaction or Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>Management Science:</td>
<td>Human Engineering</td>
</tr>
<tr>
<td>(Rationalized</td>
<td>Systems Research,</td>
<td></td>
</tr>
<tr>
<td>Procedures)</td>
<td>Operations Research,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decision Processes,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>etc.</td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>Human Factors</td>
<td></td>
</tr>
</tbody>
</table>

Indeed today we are not clear about the relation of performance to satisfaction, or even whether there is any interdependence between them. Likert and his associates have found organizations with all the logical possibilities--high morale with low productivity, low productivity with low morale,
The concept of organization health, etc. Argyris' work (2, 4), with a popular assist from William H. Whyte, Jr. (76), clouds the picture even further by postulating the inevitability of conflict between human need-satisfaction and organizational performance (as formal organizations are presently conceived). This creates, as Mason Haire has recognized (31), a calculus of values: how much satisfaction or health is to be yielded for how many units of performance?

Generally speaking, then, this is the state of affairs: two criteria, crudely measured, ambiguous in meaning, questionable in utility, and fraught with value connotations (35). In view of these difficulties, a number of other, more promising, approaches have been suggested. The most notable of these are the criterion of multiple goals, the criterion of the situation, and the criterion of system characteristics.

The Criterion of Multiple Goals

This approach rests on the assumption that "...organizations have more than a single goal and that the interaction of goals will produce a different value framework in different organizations" (57, p. 42). Likert, who is a proponent of the multiple criterion approach, claims that very few organizations, if any, obtain measurements that clearly reflect the quality and capacity of the organization's human resources. This situation is due primarily to the shadow of traditional theory, which tends to overlook the human and motivational variables and the relatively new developments in social science that only now permit measurements of this type. Likert goes on to enumerate twelve criteria, covering such dimensions as loyalty and identification with the institution and its objectives, degree of confidence and trust, adequacy and efficiency of communication, amount and quality of teamwork, etc. (41). By and large, Likert's criteria are psychologically based and substantially enrich the impoverished state of effectiveness criteria.3

The Criterion of the Situation

This approach is based on the reasoning that organizations differ with respect to goals and that they can be analytically distinguished in terms of goal orientation. As Parsons pointed out: "As a formal analytical point of reference, primacy of orientation to the attainment of a specific goal is used as the defining characteristic of an organization which distinguishes it from other types of social systems." (55, p. 64).

In an earlier paper by Bennis (9) a framework was presented for characterizing four different types of organizations based on a specific criterion variable. These "pure" types are rarely observed empirically, but they serve to sharpen the difference among formally organized activities. Table 2 represents an example of developing effectiveness variables on the basis of organizational parameters.

### Table 2

<table>
<thead>
<tr>
<th>Type of organization</th>
<th>Major function</th>
<th>Examples</th>
<th>Effectiveness criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habit</td>
<td>Replicating standard and uniform products</td>
<td>Highly mechanized factories, etc.</td>
<td>No. of products</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>Creating new ideas</td>
<td>Research organizations; design and engineering divisions; consulting organizations, etc.</td>
<td>No. of ideas</td>
</tr>
<tr>
<td>Indoctrination</td>
<td>Changing people's habits, attitudes, intellect, behavior (physical and mental)</td>
<td>Universities, prisons, hospitals, etc.</td>
<td>No. of &quot;clients&quot; leaving</td>
</tr>
<tr>
<td>Service</td>
<td>Distributing services either directly to consumer or to above types</td>
<td>Military, government, advertising, taxi companies, etc.</td>
<td>Extent of services services performed</td>
</tr>
</tbody>
</table>

3. See also Kahn, Mann, and Seashore (35), Introduction, for other suggestions for criteria.


541
The Criterion of System Characteristics

This approach, most cogently advanced by sociologists, is based on a "structural-functional" analysis. Selznick, one of its chief proponents, characterizes the approach in the following way:

Structural-functional analysis relates contemporary and variable behavior to a presumptively stable system of needs and mechanisms. This means that a given empirical system is deemed to have basic needs, essentially related to self-maintenance; the system develops repetitive means of self-defense; and day-to-day activity is interpreted in terms of the function served by that activity for the maintenance and defense of the system. (62, p. 28).

Derivable from this system model are basic needs or institutional imperatives that have to be met if the organism is to survive and "grow." Selznick, for example, lists five:

1. The security of the organization as a whole in relations to social forces in its environment.
2. The stability of the lines of authority and communication.
3. The stability of informal relations within the organization.
4. The continuity of policy and of the sources of its determination.
5. A homogeneity of outlook with respect to the meaning and role of the organization. (62, pp. 29-30).

Caplow, starting from the fundamental postulate that organizations tend to maintain themselves in continuous operation, identifies three criteria of organizational success: the performance of objective functions, the minimization of spontaneous conflict, and the maximization of satisfaction for individuals (22). Obviously, with the exception of the second criterion, these resemble the old favorites, performance and satisfaction.

The preceding summaries do not do full justice to the nuances in these three approaches or the enormous creative effort that went into their development. Nor do they include the ideas of many thoughtful practitioners. Despite these limitations, the discussion of multiple criteria, situational parameters, and system characteristics represents the main attempts to solve the criterion problem.

INADEQUACY OF CRITERION VARIABLES
FOR THE MODERN ORGANIZATION

One thing that is new is the prevalence of newness, the changing scale and scope of change itself, so that the world alters as we walk in it, so that the years of man's life measure not some small growth or rearrangement or moderation of what he learned in childhood, but a great upheaval... To assail the changes that have unmoored us from the past is futile, and in a deep sense, I think it is wicked. We need to recognize the change and learn what resources we have.--Robert Oppenheimer (53, pp. 10-11).

The history of other animal species shows that the most successful in the struggle for survival have been those which were most adaptable to changes in their world.--H. Bronowski (17, p. 137).

The present ways of thinking about and measuring organizational effectiveness are seriously inadequate and often misleading. These criteria are insensitive to the important needs of the organization and out of joint with the emerging view of contemporary organization that is held by many organizational theorists and practitioners. The present techniques of evaluation provide static indicators of certain output characteristics (i.e. performance and satisfaction) without illuminating the processes by which the organization searches for, adapts to, and solves its changing goals. (56). However, it is these dynamic processes of problem-solving that provide the critical dimensions of organizational health, and without knowledge of them output measurements are woefully inadequate. 5

This rather severe charge is based upon the belief that the main challenge confronting the modern organization (and society) is that of coping with external stress and change. This point hardly needs elaboration or defense. Ecclesiastes glumly pointed out that men persist in disordering their settled ways and beliefs by seeking out many inventions. The recent work in the field of organizational behavior reflects this need and interest; it is virtually a catalogue of the problems in organizational change. In a 1961 monograph on managing major change in organizations, Mann and Neff stated the issue this way: "Among the most conspicuous values in American culture of the twentieth century are progress, efficiency, science and rationality, achievement and success. These values have helped to produce a highly dynamic society--a society in which the predominant characteristic is change." (45, p. 1). Kahn, Mann, and Seashore, when discussing a criterion variable, "the ability of the organization to change appropriately in response to some objective requirement for change," remarked: "Although we are convinced of the theoretical importance of this criterion, which we have called organizational flexibility, we have thus far been unable to solve the operational problems in...
THE CONCEPT OF ORGANIZATION HEALTH

involved in its use." (35, p. 4).

The basic flaw in the present effectiveness criteria is their inattention to the problem of adapting to change. To illuminate some of the consequences of this omission, let us turn to one rather simple example. The example is drawn from an area of research that started at the Massachusetts Institute of Technology about 1949 on the effects of certain organizational patterns (communication networks) on problem-solving by groups (38). Two of these networks, the Wheel and the Circle, are shown in Figure 1.

![Figure 1](image)

Two types of communication networks for problem-solving by a group of five persons.

The results of these experiments showed that an organization with a structure like the Wheel can solve simple tasks (e.g. identification of the color of a marble that is common to all five group members) more rapidly, more clearly, and more efficiently than an organization like the Circle. Thus the Wheel arrangement is plainly superior in terms of the usual criteria employed to evaluate effectiveness. However, if we consider two other criteria of organizational effectiveness that are relevant to the concern with change—flexibility and creativity, we discover two interesting phenomena. First, the rapid acceptance of a new idea is more likely in the Circle than in the Wheel. The man in the middle of the Wheel is apt to discard an idea on the grounds that he is too busy or the idea is impractical. Second, when the task is changed, for example by going from "pure" color marbles to unusual color marbles (such as ginger-ale color or Blue-green), the Circle organization is better able to adapt to this change by developing a new code (68). As Leavitt pointed out:

... by certain industrial engineering-type criteria (speed, clarity of organization and job descriptions, parsimonious use of paper, etc.), the highly structured, highly routinized, non-involving centralized net seems to work best. But if our criteria of effectiveness are more ephemeral, more general

(large fragment cut off, possibly discussing acceptance of creativity, flexibility in dealing with novel problems, generally high morale and loyalty), then the more egalitarian or decentralized type net seems to work better. (39, p. 22).

If we view organizations as adaptive, problem-solving, organic structures, then inferences about effectiveness have to be made, not from static measures of output, though these may be helpful, but on the basis of the processes through which the organization approaches problems. In other words, no single measurement of organizational efficiency or satisfaction—no single time-slice of organizational performance—can provide valid indicators of organizational health. An organization may be essentially healthy despite measurements that reveal that its performance and satisfaction measurements are lower than last month's; it can be unhealthy even if its performance and efficiency figures are higher than last month's. Unhealthy and healthy, that is, in relation to the ability to cope with change, with the future. Discussing the neurotic processes, Kubie makes the same point:

There is not a single thing which a human being can do or feel, or think, whether it is eating or sleeping or drinking or fighting or hating or loving or grieving or exulting or working or playing or painting or inventing, which cannot be either sick or well... The measure of health is flexibility, the freedom to learn through experience, the freedom to change with changing internal and external circumstances, to be influenced by reasonable argument, admonitions, exhortations, and the appeal to emotions; the freedom to respond appropriately to the stimulus of reward and punishment, and especially the freedom to cease when sated. The essence of normality is flexibility in all of these vital ways. (37, p. 20).

Any moment of behavior is unhealthy if the processes that set it in motion predetermine its automatic repetition, regardless of the environmental stimuli or consequences of the act. For example, it is plausible that lowering efficiency in order to adjust to some product change may be quite appropriate when market demands are considered. It is equally plausible that morale, or whatever measure is used to gauge the human factor, may similarly plummet during this period. In fact, maintaining the same level of efficiency and morale in new circumstances may be dysfunctional for the health of the organization.

Let us review the argument thus far. The main challenge confronting today's organization, whether it is a hospital or a business enterprise, is that of responding to changing conditions and adapting to external stress. The salience of change is forced on organizations because of the growing interdependence between their changing boundary conditions and society (a point that will be elaborated...
Effective organization is a function of the work to be done and the resources and techniques available to do it. Thus changes in methods of production bring about changes in the number of work roles, in the distribution of work between roles and in their relationship to one another. Failure to make explicit acknowledgement of this relationship between work and organization gives rise to non-valid assumptions, e.g. that optimum organization is a function of the personalities involved, that it is a matter connected with the personal style and arbitrary decision of the chief executive, that there are choices between centralized and decentralized types of organization, etc. Our observations lead us to accept that optimum organization must be derived from an analysis of the work to be done and the techniques and resources available. (19, pp. 18-19).

The work of Emery and Trist, which has influenced the thinking of Brown, stressed the "socio-technical system," based on Bertalanffy's "open system" theorizing. (13). They conclude that:

... the primary task of managing an enterprise as a whole is to relate the total system to its environment, and not internal regulation perse. (24, p. 10).

If management is to control internal growth and development it must in the first instance control the "boundary conditions"—the forms of exchange between the enterprise and its environment. ... The strategic objective should be to place the enterprise in a position in its environment where it has some assured conditions for growth—unlike war the best position is not necessarily that of unchallenged monopoly. Achieving this position would be the primary task or overriding mission of the enterprise. (24, p. 13).

In reference to management development, A. T. M. Wilson, Former Director of Tavistock Institute, pointed out:

8. Although not quoted here, a book by Salzick is also directly relevant. See (63).
9. See Bennis (9) for elaboration of this point.
10. Wilson lists six "areas of social activity: each of which contain a number of significant social institutions and social groups. These areas may be rather summarily labelled as: (1) Government, (ii) Consumers, (III) Shareholders, (IV) Competitors, (V) Raw material and power suppliers, and (vi) Groups within the firm." (77, p. 3). These represent some of the boundary conditions for the manager.

One general point of high relevance can be seen in these discussions of the firm as an institution. The tasks of the higher level managers center on problems in which there is a continuously high level of uncertainty; complex value decisions are inevitably involved; and this has a direct bearing on the requirements of personality for top level management ... (77, p. 13).

And H. J. Leavitt said on the same subject:

Management development programs need, I submit, to be oriented much more toward the future, toward change, toward differences from current forms of practice and behavior ... We ought to allocate more of the effort of our programs to making our student a more competent analyst. We ought, in other words, to try to teach them to think a little more like scientists, and indeed to know a good deal more about the culture and methods of scientists. (39, pp. 32-33).

What relevance have these quotations to the main theme of this essay? Note, first of all, that these theorists all view the organization (or institution) as an adaptive structure actively encountering many different environments, both internal and external, in their productive efforts. Note also the key terms: change, uncertainty, future, task, mission, work to be done, available resources, exchanges between the enterprise and environment. There is no dialogue here on the relation between "productivity" and "satisfaction," no fruitless arguments between the human relationists and scientific management advocates. Indeed, it seems that it is no longer adequate to perceive organization as an analogue to the machine as Max Weber indicated: "... (bureaucracy is like) a modern judge who is a vending machine into which the pleadings are inserted together with the fee and which then digorges the judgment together with its reasons mechanically derived from the code." (8, p. 421). Nor is it reasonable to view the organization solely in terms of the sociopsychological characteristics of the persons involved at work, a viewpoint that has been so fashionable of late. Rather, the approach that should be taken is that of these quoted writers: organizations are to be viewed as "open systems" defined by their primary task or mission and encountering boundary conditions that are rapidly changing their characteristics. Given this rough definition, we must locate some effectiveness criteria and the institutional prerequisites that provide the conditions for the attainment of this criteria.
THE SPIRIT OF INQUIRY AS A MODEL FOR ORGANIZATION

Findings are science's short-range benefits, but the method of inquiry is its long-range value. I have said that the invention of organization was Man's first most important achievement; I now add that the development of inquiry will be his second. Both of these inventions change the species and are necessary for its survival. But both must become a part of the nature of Man himself, not just given house room in certain groups. Organization is by now a part of every man, but inquiry is not. The significant product of science and education will be the incorporation within the human animal of the capability and habit of inquiry. --H. Tblen (70, p. 217).

Whether our work is art or science or the daily work of society, it is only the form in which we explore our experience which is different; the need to explore remains the same. This is why, at bottom, the society of scientists is more important than their discoveries. What science has to teach us here is not its techniques but its spirit: the irresistible need to explore. --J. Bronowski (18, p. 93).

It has been asserted throughout this paper that organizations must be viewed as adaptive, problem-solving structures operating and embedded in complicated and rapidly changing environments. If this view is valid, then it is fair to postulate that the methodological rules by which the organization approaches its task and "exchanges with its environments" are the critical determinants of organizational effectiveness. These methodological rules or operating procedures bear a close resemblance to the rules of inquiry, or scientific investigation. Therefore, the rules and norms of science may provide a valuable, possible necessary model for organizational behavior.

First, it should be stated what is meant and what is not meant by "science" in this context. It is not the findings of science, the vast array of data that scientists produce. Nor is it a barren operationalism --what some people refer to as "scientism"--or the gadgetry utilized for routine laboratory work. Rather it is what may be called the scientific "temper" or "spirit." It is this "spirit of inquiry," which stems from the value position of science, that such authors as Dewey, Bronowski, Geiger, and Sandford have emphasized must be considered if our world is to survive. This position says essentially that the roles of scientist and citizen cannot be sharply separated. As Waddington put it:

The true influence of science is an attitude of mind, a general method of thinking about and investigating problems. It can, and I think it will, spread gradually throughout the social consciousness without any very sharp break with the attitudes of the past. But the problems for which it is wanted face us already; and the sooner the scientific method of handling them becomes more generally understood and adopted, the better it will be. (72, p. xliii).

Now it is necessary to look a bit more closely at what is meant by this "scientific attitude." Relevant here are two important aspects of the scientific attitude: one having to do with the methodology of science and one related to the social organization of science. The former is a complex of human behavior and adjustment that has been summed up as the "spirit of inquiry." This complex includes many elements, only two of which are considered here. The first may be called the hypothetical spirit, the feeling for tentativeness and caution, the respect for probable error. As Geiger says: "...the hypothetical spirit is the unique contribution scientific method can offer to human culture; it certainly is the only prophylactic against the authoritarian mystique so symptomatic of modern nerve failure." (28, p. 11).

The second ingredient is experimentalism, the willingness to expose ideas to empirical testing, to procedures, to action. The hypothetical stance without experimentalism would soon develop into a rather arid scholasticism. Experimentalism without the corrective of the hypothetical imagination would bring about a radical, "dustbowl" empiricism lacking significant insight and underlying structures capable of generalization. These two features, plus the corrective of criticism, is what is meant by the methodological rules of science; it is the spirit of inquiry, a love of truth relentlessly pursued, that ultimately creates the objectivity and intelligent action associated with science.

The second important aspect of the scientific attitude is that concerning the social organization of science, the institutional imperatives of the scientific enterprise. A number of social scientists, inspired by the work of Parsons (54) and Merton (49), (50), have examined the society of scientific enterprise (5), (21), (47), (60). What they have said is important for the argument presented here. Only when the social conditions of science are realized can the scientific attitude exist. As Sanford pointed out:

Science flourishes under that type of democracy that accords freedom of opinion and dissent, and respect for the individual. It is against all forms of totalitarianism, of mechanization and regimentation. . . In the historical development of the ends that are treasured in Western societies there is reason to believe that science has had a determining role. Bronowski again: Men have asked for freedom, justice and respect precisely as science has spread among them. (61, p. 9).

or Parsons:

Science is intimately integrated with the whole social structure and cultural tradition. They mutually support one another--only in certain types of society can science flourish and conversely without a continuous and healthy development and application of science such a society cannot function properly. (5, p. 85).
What are the conditions that comprise the ethos of science? Barber identifies five that are appropriate to this discussion: Rationality, universalism, individualism, communality, and disinterestedness. (6) A brief word about each of these is in order. The goal of science is understanding, understanding in as abstract and general a fashion as possible. Universalism, as used here, means that all men have morally equal claims to discover and to understand. Individualism, according to Barber, expresses itself in science as anti-authoritarianism; no authority but the authority of science need be accepted or trusted. Communality is close to the utopian communist slogan: "From each according to his abilities, to each according to his needs." This simply means that all scientific peers have the right to share in existing knowledge; withholding knowledge and secrecy are cardinal sins. The last element, disinterestedness, is to be contrasted with the self-interest usually associated with organizational and economic life. Disinterestedness in science requires that role incumbents serve others and gain gratification from the pursuit of truth itself. These five conditions comprise the moral imperatives of the social organization of science. They are, of course, derived from an "ideal type" of system, an empirically imaginable possibility but a rare phenomenon. Nevertheless, insofar as they are imperatives, they do in fact determine significantly the behavior of scientific organization.

There are two points to be made in connection with this model of organization. The first was made earlier but may require reiteration: the spirit of inquiry can flourish only in an environment where there is a commitment toward the five institutional imperatives. The second point is that what is now called the "human relations school"(11) has been preoccupied primarily with the study of those factors that this paper has identified as the prerequisites of the science organization. In fact, only if we look at the human-relations approach with this perspective do we obtain a valid view of their work. For example, a great deal of work in human relations has focused on "communication" (43), "participation" (43), and "decision-making." Over generalizing a bit, we can say that most of the studies have been (from a moral point of view) predicated on and lean toward the social organization of science as has been outlined here. Note, for instance, that many studies have shown that increased participation, better communication (keeping worker "informed"), more "self-control," and decreased authoritarianism are desirable ends. Because of their emphasis on these factors, the researchers and theoreticians associated with human-relations research have sometimes been perceived as "soft-headed," unrealistic, too academic, and even utopian. In some cases, the social scientists themselves have invited these criticisms by being mainly interested in demonstrating that these participative beliefs would lead to heightened morale and, on occasion, to increased efficiency. So they have been accused by many writers as advocates of "happiness" or a moo-cow psychology. 12

These are invalid criticisms, mainly because the issue is being fought on the wrong grounds. The root of the trouble is that the social scientists have not been entirely aware or prescient enough to see the implications of their studies. Rather than debating the viability of socio-psychological variables in terms of the traditional effectiveness variables, which at this point is highly problematical, they should be saying that the only way in which organizations can develop a scientific attitude is by providing conditions where it can flourish. In short, the norms of science are both compatible and remarkably homogeneous with those of a liberal democracy. We argue, then, that the way in which organizations can master their dilemmas and solve their problems is by developing a spirit of inquiry. This can flourish only under the social conditions associated with the scientific enterprise, i.e. democratic ideals. Thus it is necessary to emphasize the "human side of enterprise," that is, institutional conditions of science, if organizations are expected to maintain mastery over their environment. 13

Now, assuming that the social conditions of science have been met, let us return to the designated task of identifying those organizational criteria that are associated with the scientific attitude.

THE CRITERIA OF SCIENCE AND MENTAL HEALTH APPLIED TO ORGANIZATIONS

Perhaps no other area of human functioning has more frequently been selected as a criterion for mental health than the individual's reality orientation and his efforts at mastering the environment.--M. Jahoda (34, p. 53).

I now propose that we gather the various kinds of behavior just mentioned, all of which have to do with effective interaction with the environment, under the general heading of competence.--Robert White (75, p. 317).

... all aspects of the enterprise must be subordinated to its primary task. It is not only industrial enterprises, however, which must remain loyal to their primary tasks. This is so of all human groups, for these are all compelled, in order to maintain themselves in existence, to undertake some form of appropriate action in relation to their environment... An organism, whether individual or social, must do work in order to keep itself related to its external environment, that is, to meet reality. --Eric Trist (19, p. xvi).
THE CONCEPT OF ORGANIZATION HEALTH

These quotations provide the framework for the following analysis. They express what has been the major concern throughout this paper: that, when organizations are considered as "open systems," adaptive structures coping with various environments, the most significant characteristic for understanding effectiveness is competence, mastery, or as the term has been used in this essay, problem-solving. It has been shown that competence can be gained only through certain adaptations of science: its attitude and social conditions. It is now possible to go a step further by underlining what the above quotations reveal, that the criteria of science bear a close kinship to the characteristics of what mental-health specialists and psychiatrists call "health."

There is an interesting historical parallel between the development of criteria for the evaluation of mental health and the evolution of standards for evaluating organizational health. Mastery, competence, and adaptive, problem-solving abilities are words-relatively new to both fields. In the area of organizational behavior these words are replacing the old terms "satisfaction" and "work competence." Similarly, an important change has taken place in the mental-health field, which has had some of the same problems in determining adequate criteria. Rather than viewing health exclusively in terms of some highly inferential intra-psychic reconstitutions, these specialists are stressing "adaptive mechanisms" and "conflict-free," relatively autonomous ego-functioning, independent of id energies. The studies of White (75), Rapoport (58), Erikson (25), Hartmann (33), and other so-called ego-psychologists all point in this direction. The main reason for the confluence of organizational behavior and mental health is at bottom quite simple. Both the norms of science and the methodology of psycho-therapeutic work have the same goal and methodology: to perceive reality, both internal and external; to examine unfailingly the positions of these realities in order to act intelligently. It is the belief here that what a patient takes away and can employ after treatment is the methodology of science, the ability to look facts in the face, to use the hypothetical and experimental methods--the spirit of inquiry--in understanding experience. Sanford has said in this connection:

...most notably in Freud's psychoanalytic method of investigation and treatment. (This method is in my view, Freud's greatest, and it will be his most lasting contribution.) By the method I mean the whole contractual arrangement according to which both the therapist and patient become investigators, and both objects of careful observation and study; in which the therapist can ask the patient to face the truth because he, the therapist, is willing to try to face it in himself; in which investigation and treatment are inseparable aspects of the same humanistic enterprise. (61, p. 12).

14. See Selznick (63), chap. 3, for similar emphasis.

and in Freud's words:

Finally, we must not forget that the relationship between analyst and patient is based on a love of truth, that is, on the acknowledgement of reality, and that it precludes any kind of sham or deception. (26, pp. 351-352).

It is now possible to postulate the criteria for organizational health. These are based on a definition by Marie Jahoda, according to which a healthy personality "...actively masters his environment, shows a certain unit of personality, and is able to perceive the world and himself correctly." (25, p. 51). Let us take each of these elements and extrapolate it into organizational criteria.

1. "Actively masters his environment": Adaptability. In the terms of this paper, this characteristic coincides with problem-solving ability, which in turn depends upon the organization's flexibility. Earlier it was pointed out that flexibility is the freedom to learn through experience, to change with changing internal and external circumstances. Another way of putting it, in terms of organizational functioning, is to say that it is "learning now to learn." This is equivalent to Bateson's notion of "deutero-learning," the progressive change in rate of simple learning (7).

2. "Certain unit of personality": The Problem of Identity. In order for an organization to develop adaptability, it needs to know who it is and what it is to do; that is, it has to have some clearly defined identity.14 The problem of identity, which is central to much of the contemporary literature in the mental-health field, can in organizations be examined in at least two ways: (a) determining to what extent the organizational goals are understood and accepted by the personnel, and (b) ascertaining to what extent the organization is perceived veridically by the personnel.

As to the problem of goals, Selznick pointed out:

The aims of large organizations are often very broad. A certain vagueness must be accepted because it is difficult to foresee whether more specific goals will be realistic or wise. This situation presents the leader with one of his most difficult but indispensable tasks. He must specify and recast the general aims of his organization so as to adapt them, without serious corruption, to the requirements of institutional survival. This is what we mean by the definition of institutional mission and role. (63, p. 66).

The same point is made by Simon, Smithburg, and Thompson: "No knowledge of administrative techniques, -then, can relieve the administrator from the task of moral choice--choice as to organizational goals and methods and choice as to the treatment of the other human beings in his organization." (67, p. 24).
In addition to the clear definition of mission, which is the responsibility of the leader to communicate, there also has to be a working consensus on the organization of work. Wilfred Brown's work is extremely useful in this connection. He enumerates four concepts of organization: the manifest organization, the one that is seen on the "organization chart" and is formally displayed; the assumed organization, the one that individuals perceive as the organization (were they asked to draw their phenomenological view of the way that things work); the extant organization, the situation as revealed through systematic investigation, say by a student of organizations; and the requisite organization, or the situation as it would have to be if it were "in accord with the real properties of the field in which it exists."

"The ideal situation," Brown goes on to say, "is that in which the manifest, the assumed, the extant, and the requisite are as closely as possible in line with each other." (19, p. 24). Wherever these four organizational concepts are in contradiction, we find a case of what Erikson calls "identity diffusion" (25). Certainly this phenomenon is a familiar one to students and executives of organizations. Indeed, the great attention paid to the "in accord with" criterion is used here for organizational health, within the limits of rationality, for successful mastery over the relevant environments. 15

In summary, then, I am saying that the basic features of organization rely on adequate methods for solving problems. These methods stem from the elements of what has been called the scientific attitude. From these ingredients have been fashioned three criteria or organizational mechanisms, which fulfill the prerequisites of health. These criteria are in accord with what mental-health specialists call health in the individual.

Undeniably, some qualifications have to be made. The mensuration problem has not been faced, nor have the concrete details for organizational practice been fully developed. Nonetheless, it has been asserted that the processes of problem-solving--of adaptability--stand out as the single most important determinant of organizational health and that this adaptability depends on a valid identity and valid reality-testing. 16

SOME IMPLICATIONS OF THE SCIENCE MODEL FOR ORGANIZATIONAL BEHAVIOR

There is one human characteristic which to-day can find a mode of expression in nationalism and war, and which, it may seem would have to be completely dented in a scientific society. That is the tendency to find some dogma to which can be attached complete belief, forthright and unquestioning. That men do experience a need for certainty of such a kind can scarcely be doubted. Is science, for all its logical consistency, in a position to satisfy this primary need of man?--C. H. Waddington (72, pp. 163–164).

We are not yet emotionally an adaptive society, though we try systematically to develop forces that tend to make us one. We encourage the search for
THE CONCEPT OF ORGANIZATION HEALTH

new inventions; we keep the mind stimulated, bright, and free to seek out fresh means of transport, communication, and energy; yet we remain, in part, appalled by the consequences of our ingenuity and too frequently, try to find security through the shoring up of ancient and irrelevant conventions, the extension of purely physical safeguards, or the delivery of decisions we ourselves should make into the keeping of superior authority like the state. These solutions are not necessarily unnatural or wrong, but historically they have not been enough, and I suspect they will never be enough to give us the serenity and competence we seek... we may find at least part of our salvation in identifying ourselves with the adaptive process and thus share... some of the joy, exuberance, satisfaction and security... to meet the changing time.—E. Morison (51, p. 11).

The use of the model of science as a form for the modern organization implies some profound reforms in current practice, reforms that may appear to some as too adventurous or utopian. This criticism is difficult to deny, particularly since not all the consequences can be clearly seen at this time. However, before necessity diminishes the desirability of using the science model, let us examine a few consequences that stand out rather sharply.

1. The problem of commitment and loyalty. Although the viewpoint does have its critics, such as William H. Whyte, Jr., most administrators desire to develop high commitment and loyalty to the organization. Can the scientific attitude, with its ascetic simplicity and acceptance of risk and uncertainty, substitute for loyalty to the organizations and its purpose? Can science, as Waddington wonders, provide the belief in an illusion that organizational loyalty is thought to provide? The answer to this is a tentative "yes and no." Substituting the scientific attitude for loyalty would be difficult for those people who are committed to the pursuit of knowledge, is both far too abstract and far too threatening. For some, the "escape from freedom" is a necessity, and the uncertain nature of the scientific attitude would be difficult to accept. However, it is likely that even these individuals would be influenced by the adoption of the science model by the organization. Loyalty to the organization per se would be transformed into loyalty and commitment directed to the spirit of inquiry. Hence, a higher rate of mobility is envisaged for organizations based on movement towards those environments in which the social conditions of science exist. Gouldner, in another context, has discussed this difference between individuals in terms of the split of organizational roles into "locals and cosmopolitans" (30). The cosmopolitan derives his rewards from inward standards of excellence, internalized and reinforced through professional (usually scientific) identification. On the other hand, the local (what Marvick calls the "bureaucratic orientation" (48) derives his rewards from manipulating power within the hierarchy. The former are considered to be better organization men than the latter. Loyalty within the scientific organizational conditions specified here, would be directed not to particular ends or products or to work groups but to identification with the adaptive process of the organization.

2. Recruitment and training for the spirit of inquiry. There are some indications that the problems of recruitment and training for the social organization of science are not as difficult as has been expected. For one thing, as Bruner has shown (20), today's school children are getting more and better science teaching. It is to be hoped that they will learn as much about the attitude of science as they will about its glamour and techniques. In addition, more and more research-trained individuals are entering organizations. As McGregor points out: "Creative intellectual effort by a wide range of professional specialists will be as essential to tomorrow's manager as instruments and an elaborate air traffic control system are to today's jet pilot." (44, p. 27). Individuals trained in scientific methodology can easily adapt to, in fact will probably demand, more and more freedom for intellectual inquiry. If McGregor's and Leavitt and Whisler's (40) prognostications are correct, as they presently seem to be, then there is practically no choice but to prepare a social milieu in which the adaptive, problem-solving processes can flourish.

As to training, only a brief word needs to be said. The training program of the National Training Laboratories (16) and the work of Blake (13), Blansfield (15), and Shepard (65) are based rather specifically on developing better diagnosticians of human behavior. It is apparent from such training studies that the organization of tomorrow, heavily influenced by the growth of science and technology and manned by an increasing number of professionals, appears to have the necessary requirements for constructing organizations based on inquiry.

3. Intergroup Competition. Blake and Mouton, guided partly by the work of the Sherifs (66), have disclosed for examination one of organization's most troublesome problems, intergroup conflict and collaboration. These perseverating conflicts, usually based on a corrupt practice of vested interests, probably dissipate more energy and money than any other single malady caused by humans. Intergroup conflict, with its "win-lose" orientation, its dysfunctional loyalty (to the group or product, not to the truth), its cognitive distortions of the outsider, and its inability to reach what has been called creative compromise, effectively disrupts the commitment to truth. By means of a laboratory approach Blake and Mouton have managed to break... the mental assumptions underlying win-lose conflict. Factually based mutual problem identification,
fluidity in initial stages of solution-proposing rather than fixed position taking, free and frequent interchange between representatives and their constituent groups and focusing on communality as well as differences as the basis for achieving agreement and so on, are but a few of the ways which have been experimentally demonstrated to increase the likelihood of arriving at mutually acceptable solutions under conditions of collaboration between groups (14).

What the authors do not explicitly say but only imply is that the structure of their experimental laboratory approach is based on the methods of inquiry that have been advocated in this paper. Theirs is an action-research model, in which the subjects are the inquirers who learn to collect, use and generalize from data in order to understand organizational conflict. Rational problem-solving is the only prophylaxis presently known to rid organizations of perseverating intergroup conflict.

Loyalty, recruitment and training, and intergroup hostility are by no means all the organizational consequences that this paper suggests. The distribution of power, the problems of group cohesive-

REFERENCES

18. It is suspected that group cohesiveness will decrease as the scientific attitude infuses organizational functioning. With the depersonalization of science, the rapid turnover, and some expected individualism, cohesiveness may not be functional or even possible.
THE CONCEPT OF ORGANIZATION HEALTH

45. OPPENHEIMER, R.: Prospects in the Arts and Sciences, Perspectives USA, vol. 11, pp. 5-14, Spring, 1955.
64. SHEPARD, H.: Superiors and Subordinates in Research, J. Busin., vol. 29, pp. 261-267, October, 1856.
Examples of Application of Systems Theory to Complex Social Systems

A. INTRODUCTION

Section I dealt with the General Systems approach as a means of analysis, a way of looking at complex interrelationships in the real world. Section II dealt with the nature of systems and positive and negative feedback. Section III focused on social organizations stressing their dynamics of adaptation, while Section IV concentrated on the complexities of social organizations and factors to be considered in the design and analysis of such systems. This final section is concerned with the application of General System techniques to specific systems along with ideas on management of these new systems.

Organization or system is a fundamental characteristic of all forms of life. Human life in modern society is organized around a large variety of complex social systems with behavioural, technical, and economic elements and which vary in size from small groups of persons to those that are international in scope. Intensive study is required to achieve an understanding of the nature, functioning, and management of complex social systems. This section presents selected studies dedicated to broadening such understanding and providing descriptions, models, analyses, and/or prescriptions concerning a wide variety of social systems.

The papers included in this section present discussions of topics ranging from component behaviour in large systems to applications of general system concepts at the world level. The papers may be grouped into six parts. In part (B) the two papers deal with human behaviour in large organizations and factors affecting the growth of systems. The three papers of part (C) concern applications of system theory to institutions. In part (D) three selected papers enlarge the scale of applications to the metropolitan and regional levels, while in part (E) the scale of applications is increased to the national political scene and total society levels by a group of five papers. The single paper of part (F) focuses upon a study at the level of empires, while in part (G) three papers deal with arms control, which may be construed as applications at the world level. The following paragraphs present more detailed comments on each of the papers.

In the first paper, in part (B) Argyris presents a portion of a long-range effort designed to eventually develop a systematic theory of human behaviour in organizations. General concepts of systems are used in the development of a series of hypotheses. Two industrial plants were selected to test the hypotheses empirically and to investigate stability, self-maintenance, and change in organizations. A model adapted from the physical sciences that includes input, output, and feedback mechanisms is applied to the examination of the functioning of social systems of the plants. The focus is on the response of employees to management pressure. It is concluded that, as pressure
from management is increased, employees will not tend to combat such pressure by forming cohesive groups and interpersonal rivalry among employees will increase.

Concepts of general systems theory are used by Goldsmith in a study to ascertain inherent factors that limit the growth of systems. System growth tends to cause a reduction of the degree of closeness in linkages between components and of order that set a natural limit to growth. Environmental selection is critical if an optimal horizontal organizational structure is to be maintained, otherwise growth is by multiplication rather than differentiation, structure will degrade, and growth will be limited. Relaxation of cultural constraints and increased permissiveness seem to accompany social and economic growth and to signal increased disorder. System growth is also constrained by breakdowns in control mechanisms, the loss of integrity of components, reduction of guaranteed resource availability, excessive production of wastes, and reduced variety due to simplification. Technological development is seen as producing positive feedback that can result in increased global instability.

The first paper in part (C) views corporations as elements of a larger social system, and corporate decisions affect the ecosystems in which corporations operate. In a study of future directions of the impacts of corporate decisions on the encompassing social system, Ericson proposes that the relationship between corporate values and general social system values can be clarified by means of computer-based feedback and dynamic process models. A review of research on value systems indicates that corporate decisions are typically made in view of multiple criteria or values, managers face value dilemmas and role conflicts, and little insight is provided into how multiple values are reconciled in the decision process. A summary of three examples of cybernetic system models provides the background for a discussion of the effects of cybernetic organization modeling and information technology on the values of managers and the relationship of organizations with the external environment. The study concludes that redefinitions of traditional values can be enhanced by more extensive use of information technology by organizations.

In an operations research study of the hospital as a system, Howland emphasizes the need for determining quantitative criteria for measuring the effectiveness of patient care. After attempting without success to establish a basis for interdisciplinary research through the identification of dependent variables common to the disciplines involved, the multidisciplinary team found that such a basis could be obtained by the application of cybernetic concepts, principally regulation and control. In order to apply the concepts, it was necessary to determine explicit measures of performance. The output of the hospital consists of the fulfillment of a hierarchy of tasks, categorized as selection, service, supply, and maintenance. The performance of the hospital system is a function of its intrinsic characteristics and its policies for utilizing primary resources consisting of staff, patients, plant, and equipment. A conceptual model presented in the paper includes relationships within and between hospital subsystems, interaction of the system with its external environment, and performance and resource measures.

Patient care is defined as the maintenance of a homeostatic balance for individuals by Howland and McDowell in a second paper on hospital systems. The measure of patient care is an indicator of overall system performance. A need for medical care is viewed as a deviation from the homeostatic balance of an individual, and patient care consists of therapy as regulation to
achieve well being, the desired balance. In a cybernetic model of the assessment and regulation of patient states, patient condition is measured as the difference between actual and desired states, the decision for treatment is based on the results of such measurement, and the regulatory function responds to variations in patient states, where patterns of deviations compared to specified limits indicate the level of treatment required. A model developed in the study reveals the critical need for obtaining accurate information about patient conditions. Recent developments in data processing and instrumentation offer great potential for the design of patient-centered systems that can help hospitals to be more effective in the regulation of patient states and the delivery of appropriate medical services.

In the first paper of part (D) the indiscriminate usage of prime land without adequate consideration of alternative future uses is the principal concern of a study by Hollingshead. A model is proposed that seeks to provide a holistic decision making format for the rational use of land and the reduction of urban sprawl. The paper identifies and discusses pertinent subsystems, system flows, feedback loops, relationships, and an extensive list of primary variables for the study of land allocation. The focus of the man/land study is on institutionalism rather than on organism because of the circularity in the behavioural dimension of and extensive feedback in the system. The study proposes that land use decisions be made by persons with appropriate authority and in view of specified goals, higher level governmental interests should have priority over those of local levels, innovations in land use be encouraged, and it is wise to conserve prime biotically productive land.

Metropolitan areas can provide residents with choices in greater variety and at lower costs than can more dispersed areas. Deutsch characterizes these choices as face-to-face transactions, where a selection is made between mutually exclusive alternatives, and each choice involves some kind of commitment. The useful size of a metropolis is a function of the capability of communication and transport facilities to provide alternatives to which residents can respond. The range of choices is a partial determinant of urban attractiveness. The attraction to the metropolis can be diminished by communication overloads that can be physical or behavioural in nature. To avoid the frustration that results from such overloads, residents tend to withdraw to suburbs at the expense of a reduction in the range of available choices. Several approaches for solving communication problems suggested for study include the analysis of the ratio of transport to shelter costs, communication to shelter costs, proportions of land use for various purposes, allocation of resources to communication and transport to city size, and several others. The study concludes that substantial investment in communication and transport facilities will be necessary to make the decentralization of metropolitan areas feasible, and the restoration of beauty in the metropolitan areas requires public control of land use.

Long term strategic planning and short term operational planning for regional development are given special emphasis in a study by Maitra that features the formulation and application of a planning model to an area in northwest India. Social, cultural, political, and economic factors considered in the formulation of the model and the dynamics of decision making are analyzed in terms of a feedback network that relates the inner and outer environments of the total system. The framework of the model integrates selected concepts from engineering control theory, a Forrester-type social system model, and the mathematical notation for complex information transfer adapted from the work of Hurwicz. The study presents a format for incorporating the institu-
tional characteristics and objectives of a regional development organization into a formalized multilevel framework as the basis for further empirical investigation.

Part (E) begins with a discussion of political activities that are intended to regulate systems in terms of internal relationships of components and external relationships with other systems. Regulation of political systems is seen by Vickers as being threatened by two principal factors. The first is an ecological trap produced by social and cultural processes that limit the transfer of skills, institutions, and ideas from one generation to another. The second factor is inadequate communication for defining problems, evaluating programs, and obtaining necessary cooperation and consensus. Changes in political systems necessitate continual restructuring of problems, modifications in the allocation of scarce resources among conflicting alternatives, and greater understanding of the role of communication in securing agreement on and compliance with system policies. The development of effective political communication is basically an art, and the critical question is whether such skill levels can be raised sufficiently to permit regulation and the development of ideologies. The criticism of ideologies, the principal function of communication in political systems, is an awesome but essential undertaking that can be facilitated through greater understanding of the nature and control of political systems.

The planning-programming-budgeting-system is viewed as a subset of general systems models by Hartley, and the presentation includes discussions of the principal characteristics, selected applications, and limitations of the system. PPBS is designed to provide a holistic framework for designing and translating objectives and resources into a comprehensive and coherent plan. Twenty principal characteristics that comprise the structural aspects, analytical procedures, and information system that can be provided by PPBS are described in the paper. The application of PPBS has been extensive in recent years, and some applications at federal, state, and local levels in the early years of the use of the system are illustrated. Ten limitations of the systems approach are presented to suggest areas for modification and improvement to enhance the effective application of PPBS to public policy.

In a study of social dynamics, Griesinger and McClintock consider conceptual and methodological approaches for analyzing the dimensions of poverty, means of setting objectives for social change, and anticipated social and psychological effects of anti-poverty programs. Poverty has multiple dimensions, and the quality of life, motivation of people, and social ills are not solely functions of economic factors. Cultural change is essential for solving problems of poverty. The decision process for effective poverty programs requires data on the status of physical aspects of neighborhoods, economic conditions of areas, and the social organization of communities; the definition of goals and priorities; measures of effectiveness; and the means for achieving optimum decisions through compromise. A model based on an analog from quantum physics is developed to relate social and ecological forces to human behavior, effort, and discontent. The energy required of persons to counteract social forces that tend to constrain them to norms that perpetuate poverty levels is related to income distributions in a model. An application of the model to the 1965 Detroit riot reveals that the tendency to riot is a function of a degree of nonventable discontent that is related to income and human energy expenditure levels. Other applications to housing and education described in the paper provide additional validation of the model.
The conceptual view of society as a complex system of interacting institutions is utilized by Rastogi in the development of a model with which to predict and measure total society systems. Inputs for the model are derived from primary societal institutions labeled economic, government, religious-ideological, educational, health, military, and family. Examples of the many performance or state variables of the throughput process of the model include population and economic growth rates, unemployment, ethnic tension, social expectations, and politico-military pressure. The viability or survival capacity of a performance variable is measured or scaled on the basis of an interpretation of available data. The weighted sum of these scalings provides an indicator for the viability of the total system. The information links between institutions are presented in a flow graph that depicts the dynamics of society. Model simulations of five societies produced generally close correspondence between calculated values of viability and governmental stability and between simulated and actual values of economic growth, price-rise, and unemployment.

In a note on a technique for the definition and tests of completeness of relationships of variables, Katzner suggests that an approach to modeling equilibrium conditions in economic analysis can provide a strategy for organizing research in political science. Considerations involving existence, uniqueness, stability, and optimality of political systems can be expressed in a structural manner to aid in ascertaining the logical completeness and usefulness of the equilibrium model to research in political science. The note concludes by posing a series of basic equations developed in the approach and suggesting that the direction of political change can be anticipated if it is possible to specify which parameters in a given model determine a particular political system.

Part (F) consists of a single paper in which a systems analysis of mercantile imperialism based on the British version in the mid-17th century is utilized by Merritt to characterize imperial systems as being comprised of dominate (metropolitan country) and dependent (colonial) subsystems. The relationships between subsystems are controlled by the direction of flows of policy decisions and communication transactions and the direction of loyalties and identification of individuals. The self-interest of the dominant country is fundamental and served by imposing restrictive trade, migration, production, and isolationist policies on colonies to insure their dependence. Dysfunctions in the imperial system arise from actions by dependent colonies to end their isolation, insufficiently rigid controls on colonial activities, emphasis on local affairs by colonies, migration of disaffected persons from dominant to dependent subsystems, and indigenous populations. Various actions to maintain dominance-dependency relationships and preserve stability of imperial systems are outlined, but colonies will remain in the system only so long as the benefits from such association are greater than those from independence. The case of the Anglo-American empire is presented in some detail to illustrate the general characteristics and disintegration of an imperialistic system. It is suggested that the value of the paradigm presented in the paper lies in its usefulness for understanding the structure and operation of imperial systems, explaining specific cases, and predicting system behaviour. To be effective, however, it is necessary to operationalize the model through quantitative measures of communication and identification within the system.

Part (G) begins with a presentation in which the adaptation of an economic model coupled with the incorporation of operational considerations are the basis for the extension of classical models of arms races by Intriligator. The ratio goal and Richardson models reflect the classical approach to depicting arms races in which a single variable, levels of weapons held, is the basis for
policy determination. Reaction curves, borrowed from duopoly theory, and the addition of strategic alternatives and limits on weapons stocks as added considerations provide the basis for a more comprehensive model of arms races. Deterring and depriving arms are strategies for imposing limits on weapons holdings that can lead to stable conditions or equilibrium points not forecast by classical models. Although the extended model is more comprehensive than the classical approach and it extends the analysis of arms races, there are limitations to the model. Additional research will be required to incorporate variety of weapons and policy dimensions for a more complete analysis.

An application of computer-based gaming to simulate arms control, described in a paper by Davis, Carpenter, and Missler, provides a laboratory setting for the study of a proposed disarmament agreement between two nations. Economic and military resources available for manipulation by high-level decision makers are defined in a computerized data base. The economic and military sections of the model are interdependent, and the outcomes of a series of decisions for the allocation of resources are reflected in terms of overall military capability and an index of aggregate economic production. Other factors considered by participants in the simulation include changes in the data base between simulated monthly periods, intelligence requests and reports, progress reports, and other data base summaries. Although the gaming model does not include all of the aspects of an actual conflict situation, the simulation of principal variables effectively portrays the dynamics of the interaction process and provides insights not available from other analytical methods.

Research on the causes and control of international conflict is the focus of an excellent contribution by Singer, the final paper of this section. Other studies of international conflict have generally not been sufficiently explicit longitudinally and have been limited to a single level of analysis. In contrast, a more comprehensive model presented in the paper was developed to fulfill a considerable range of epistemological, methodological, and substantive requirement. For example, five levels of analysis are outlined and considered in the design of the model. In addition, the study examines conflicting incentives, temptations, and constraints generated by national and international systems; the effects of limited capability of mechanisms on the control and escalation of conflict; and some feasible approaches for achieving short-term reductions in conflict levels. The proposed model features homeostatic mechanisms to control positive feedback arising from interactions between regime and opposition, mobilization, use of mass media, and redistribution of domestic power, all of which tend to escalate international conflict. The suggested control mechanisms include modifications to media output, constraining mobilization rates, and diplomatic negotiation by proxy. While the adverse cost-benefit ratio of war, declining colonialism, and decreasing nationalistic views tend to damp positive feedback, there remains a pressing need to modify political processes by utilizing self-reinforcing feedback mechanisms to avert and control escalation in international conflict.
B.

HUMAN BEHAVIOR IN LARGE ORGANIZATIONS
UNDERSTANDING ORGANIZATIONAL CHANGE*

Chris Argyris

The objective of this paper is to describe a study of organizational change. The research is a part of a long range study designed to develop ultimately a systematic theory of human behavior in organizations.

The first step in the series was a systematic review of the empirical literature from which ten interrelated propositions were derived.1 As a second step, several of these propositions were tested in a field study of high- and low-skill employees.2 In the third step we were especially interested in studying the underlying mechanisms of organizational stability and self maintenance. By what mechanisms does an organization maintain itself? 3

Equally as important as organizational stability is the property of organizational change. A systematic theory must eventually be able to shed light on the mechanisms of organizational change. It would be especially helpful if one could begin to make explicit the organizational mechanisms involved in the transition from organizational stability to organizational change. This is a central objective of the present (and fourth) step of the research program.

In order to make explicit the organizational mechanisms involved in the study of change, it was necessary to conduct the study of change in an organization whose mechanisms for stability were known. Having knowledge of how the organization maintains itself, we should then be able to make specific predictions as to how the organization would change from its original state to a new one.

An opportunity to achieve these conditions was offered by a large (for the industry it represents) corporation composed of a number of relatively small manufacturing units. A study was first made of one plant (hereafter called Plant X) to understand the mechanisms by which its social system maintained itself (step three above). We expected the next step to be a study of Plant X as it underwent changes designed by management (after the above research was completed) to increase pressure for more managerial controls in order to “tighten up” on costs. It soon became clear, however, that it would take at least six months to a year before the impact of these changes could be measured by our relatively crude research instruments.

Fortunately, another plant (hereafter known as Plant Y) with comparable technology, structure, leadership, and controls had been subjected to a similar change about a year earlier. It was our hope that on the basis of the model of Plant X it should be possible to derive a priori hypotheses and make specific predictions as to what the researcher would find in Plant Y. If, for example, we know that a “tightening up” process is going on in a particular plant (Y), and if the social system of Y is similar to X, then we should be able to predict the direction and, hopefully, the amount of change that will occur within the social system of Plant Y. More specifically we should predict, on the basis of the Plant X model, the content of the responses to our questions and the probable percentages that we would find in Plant Y for each variable or set of variables.

As soon as permission was received to study Plant Y, some refinements were introduced in the research design. It was felt that in addition to the objective of predicting a priori how Plant Y would differ from Plant X, it would be interesting if the research group could make a priori predictions about the differences that should be found to exist among specific departments within Plant Y. For example, it would be interesting if the researcher could state a priori predictions about which departments within Plant Y would tend to have a higher and which would tend to have a lower morale; higher absenteeism, lower absenteeism; higher interest in the quality of their work, lower interest in quality, etc. Such predictions were derived by the research group from the model of Plant X.

Two high-skill departments (sub-system A) and two low-skill departments (sub-system B) were selected to be studied. The management of Plant Y felt that the high-skill departments had “high morale” and the low-skill departments had “low morale”. All the managerial personnel (N=15 for each sub-system) were interviewed, and in addition some of these were observed by the use of non-participant observations. Fifty per cent of the employees in each department (N=15 for each subsystem) were interviewed. The total employees interviewed therefore was 60 and the total number of managers was 25.

In order to help make the empirical data more meaningful a few words are in order about the model to be used below. These comments are presented in order to make the writer’s assumptions and ideas more explicit. They are not meant to imply that this exhausts the types of models that may be useful.

The fundamental assumption is that the way organization has been and continues to be described in the traditional literature of scientific management, public administration, industrial engineering, and industrial economics, is not complete. The assumption is inferred from the numerous studies and from the countless everyday observations that make specific predictions as to what the researcher would find in Plant Y. If, for example, we know that a “tightening up” process is going on in a particular plant (Y), and if the social system of Y is similar to X, then we should be able to predict the direction and, hopefully, the amount of change that will occur within the social system of Plant Y. More specifically we should predict, on the basis of the Plant X model, the content of the responses to our questions and the probable percentages that we would find in Plant Y for each variable or set of variables.

As soon as permission was received to study Plant Y, some refinements were introduced in the research design. It was felt that in addition to the objective of predicting a priori how Plant Y would differ from Plant X, it would be interesting if the research group could make a priori predictions about the differences that should be found to exist among specific departments within Plant Y. For example, it would be interesting if the researcher could state a priori predictions about which departments within Plant Y would tend to have a higher and which would tend to have a lower morale; higher absenteeism, lower absenteeism; higher interest in the quality of their work, lower interest in quality, etc. Such predictions were derived by the research group from the model of Plant X.

Two high-skill departments (sub-system A) and two low-skill departments (sub-system B) were selected to be studied. The management of Plant Y felt that the high-skill departments had “high morale” and the low-skill departments had “low morale”. All the managerial personnel (N=15 for each sub-system) were interviewed, and in addition some of these were observed by the use of non-participant observations. Fifty per cent of the employees in each department (N=15 for each sub-system) were interviewed. The total employees interviewed therefore was 60 and the total number of managers was 25.

In order to help make the empirical data more meaningful a few words are in order about the model to be used below. These comments are presented in order to make the writer’s assumptions and ideas more explicit. They are not meant to imply that this exhausts the types of models that may be useful.

The fundamental assumption is that the way organization has been and continues to be described in the traditional literature of scientific management, public administration, industrial engineering, and industrial economics, is not complete. The assumption is inferred from the numerous studies and from the countless everyday observations that
an organization is something more than what is published on the formal organizational charts representing policies and practices.4

If the existing literature is not complete, where may one turn for new insights? In the writer's opinion the literature that exists on the basic nature of organization as seen by biologists, physicists, chemists, medical researchers, anthropologists, some economists and general system theorists is a very fruitful source that requires careful exploration. There is an increasing awareness that organization (or system) is a basic property of life, if not life itself. If one takes this literature seriously (and the writer does), then one arrives naturally at the assumption that the basic properties of human or social organizations are fundamentally similar to those of organizations existing on different levels of analysis.

If this assumption is valid, then one might obtain new clues as to how organizations may be conceived conceptually by turning to the literature that examines the properties of organization. This is the direction presently being taken by the writer.

As a result of this exploration, several preliminary propositions were formulated about the basic nature of organization, that were found to be helpful. Some examples are: An organization is characterized by an arrangement of parts forming a unified whole which feeds back to help maintain the parts.5 A "part" of an organization is actually an "organic" part in that it exists by virtue of its position in the pattern that is the whole.6 The whole, in turn, may be differentiated from the parts along two dimensions. First, the whole has a boundary different from that of any given part (or subset of parts).7 Second, the functional unity of the whole displays properties revealed only in the actual process of full operation of the whole.8

These propositions have led the writer to form his own (very tentative) conceptual definition of organization. An organization is a plurality of parts, each:

1. achieving specific objective(s);
2. maintaining itself through interrelatedness with the other parts;
3. simultaneously adapting to the external environment; and
4. maintaining thereby the interrelated state of the parts.

The reader may wonder how such theoretical notions can be of help to a researcher in constructing his model. Actually a number of specific suggestions are implied in these theoretical ideas.

The first implication is that the model should strive toward understanding the patterning of all the relevant parts as well as the nature of each part. If the properties and the output of the whole are revealed in the full operation of the whole, then one must focus the research to "capture" the wholeness of the unity under study.10 It is possible that a research strategy which assumes that a series of "part" studies eventually added together will provide insight into the whole, may lead to failure.11

For example, the writer learned after a series of failures that leadership behavior (which is one part of an organization) could not be separated by the subjects from such parts as the managerial controls and the organizational structure.12 The part (leadership) could not be studied separately from the whole.

The above definition of organization also emphasizes the organization's tendency toward stability or constancy. This implies that research should focus on the specific mechanisms by which the parts create the whole, which cannot be understood unless the "equilibrium" or "steady state" of the organization is first understood.13 If this guidepost is valid then researchers need not enter arguments as to the importance of the study of change versus study of the tendency toward constancy. Both are important but in a specific case the latter precedes the former.

The "parts" of an organization exist on many levels. For example, there are individuals, informal groups (large and small), departments, divisions, cultural norms, etc. Since all the parts that are relevant need to be studied, the model should cope with variables on different levels of analyses. It should be able functionally to relate personality, informal group and formal organization variables (for example) without much difficulty. To relate the variables functionally to one another is to arrange them in a meaningful pattern and to be able to show the processes (mechanisms) that maintain this pattern.

4. For example see the work of Arensberg, Bakke, Blair, Dubin, McGregor, Shartle, Whyte, etc.
11. Psychologists who know the history of "Structuralism" are intimately acquainted with the limitations of such logic. For an interesting discussion of similar problems in biological research see J. Z. Young, "The Evolution of Organization Within the Nervous System," Adv. of Sci., Vol. XIV, No. 54, Sept. 1957, pp. 48-57.
12. Argyris, Chris, Organizational Leadership, Conference on "Leadership" directed by Professor L. Petrullo and Professor Bernard Bass, and sponsored by the Office of Naval Research, March 1959.
Fig. 1. The Social System of Plant X.
For example, one might have to construct a pattern in which personality variables are connected to informal group variables, which in turn are connected to formal organization variables, which in turn are connected to the cultural variables, etc. In short, the model should take its cue from the nature of the phenomena under study and not necessarily from the traditional academic boundaries.

POSSIBLE MODEL FOR UNDERSTANDING ORGANIZATIONAL BEHAVIOR

Making the model subservient to the data leads to the development of a rather simple gross model which lacks the rigor toward which a scientist aspires. I admit this but believe that it is more respectable for a researcher to let known reality be his guide. Philip M. Morse makes a similar point when he notes that to understand certain problems, one must understand the pattern before the details. Slowly, as Kurt Lewin suggests, by successive approximations, a rigorous model will some day be evolved. Once one has "arrived" by this path, one will not have to "return" to try to do something about variables left out. Holding the above hypothesis, Karl W. Deutsch's definitions of a model seems to be a useful point of departure. Deutsch defines a model as "structure of symbols and operating rules which is supposed to match a set of relevant points in an existing structure of process." The model to be used is borrowed (more in spirit than in substance) from the physical sciences. It is the model of a system with an input, an output, and feedback mechanisms. None of the quantitative rigor found in the physical science models is emulated. Nevertheless, the model is useful as a device to help organize the complexity, lead to new insights and make a few predictions.

A CURSORY VIEW OF PLANT X'S SOCIAL SYSTEMS

We turn to our data. In order to understand the hypotheses concerned with the mechanisms for organizational change in Y, it is first necessary to know something about the mechanisms for stability in X (whose social system can be shown to be basically similar to Y's). The following outlines some of the major properties of Plant X's social systems.

The Input Process

All systems being considered are assumed to have "openings" where inputs can be brought into the system (see Figure 1). In the case of Plant X the intake activities are conducted in the personnel department (which is located away from the plant). Formal Requirements of the High-Skilled Sub-System

The individual passes through the intake process and enters one of two sub-systems; the high-skill (A) or the low-skill (B). If the individual enters the high-skill sub-system (A), he enters a world which according to the respondents has the following characteristics:

Challenging work. Eighty-three per cent of the employees in A report that they gain much personal satisfaction while working because they have challenging and creative work.

Fair wages, job security, and bonus. Ninety-nine per cent of the employees describe their wages as "usually excellent, although because of the recession, they are now a little slow" and could be better, but this is not a complaint, just a natural desire for more.

Fair incentive system. Almost all the employees studied are on individual piece rate system. The rates for the jobs are viewed by the employees as fair. Of all the employees responding to the question about the fairness of the rates only 6 per cent say the rates are not fair. The remainder describe them either as "They're OK, I wish they were higher but then who wouldn't? That's natural," or "Some are tough, some are easy, the overall average are OK."

The "kitty." In most cases all the employees interviewed are able to build up a "kitty." Some jobs are timed so that one can make quite a lot of money in one day if he produces to the utmost of his assumed capabilities. However, as is the case in many plants, the employees restrict their reported production. They do this by simply holding back.

16. For an interesting article related to this problem, see John L. Kennedy, "A 'Transition-Model' Laboratory for Research on Cultural Change," Human Organization, 14 (Fall 1955), 16-18.
18. See Deutsch's article cited above for an interesting discussion of the different functions a model may serve.
20. Because of space limitations we are omitting the evidence that Plant X and Y have similar social systems.
21. For a detailed discussion see ibid Chapter III.
"tickets" which they must turn in if they are to get paid for the work performed. The tickets state how much the man produced of a particular item, the piece rate, the order number, etc. Thus if a man produces more than he feels (and in many cases more than his foreman feels) is wise to turn in, he banks the tickets in his "kitty" and holds them until a day when he may be assigned a tough job or when he may not be feeling well or when his machine may break down.

Ninety-seven per cent of the employees report that the "kitty" is very important to them. It provides them a measure of control over their wage fluctuations and helps to guarantee them a steady take-home pay. The remaining three per cent feel the same but add they do not make enough.

Minimum pressure from management. The employees in sub-system A report that there is a minimum pressure from management. Thus, 74 per cent report, "Management in this place is excellent," "They're wonderful people," "They hardly ever bother us, and we hardly ever see them." One hundred per cent respond that the pressure from the leadership and the controls is at a minimum. One hundred per cent report they are their own boss on the job.

Formal Requirements of the Low-Skilled Sub-System

- Non-challenging work. Eighty-five per cent of the employees report they are performing work that requires minimal abilities and offers little or no challenge.
- Fair wages, job security and bonus. 23
- Fair incentive system.
- The "kitty."
- Minimum pressure from management.

The Organizational Predispositions

For the sake of consistency and simplicity any personality aspects upon which we focus are all categorized as "predispositions." A predisposition is defined as a tendency to act in a particular situation. The predispositions are inferred from the interview data. Every individual is assumed to have a number of predispositions that he wants to express while a member of the organization, each of which may vary in degree of importance (Lewin's potency). The potency may also vary within the same individual at different times in his life or different situations. The high-potency predispositions for both sub-systems are Togetherness, Wages, Job Security, Non-involvement, and Control. These are defined as follows:

Togetherness, the need for experience in relations to the other employees. The feeling that the employees like each other without knowing each other or experiencing close human relationships.

The emphasis is on skin-surface signs of friendliness.

Wages guaranteeing a fair standard of living, and a secure job. The level of wages should be comparable with those for similar work in the community. The degree of security desired is to have a permanent job during the most difficult depression periods.

Non-involvement in the formal activities of the organization. They need not feel responsible for anything concerned with the organization excepting one's own specific job. Non-involvement does not include the individual's attitude toward productivity. It includes such needs as not to be upwardly mobile, and not to worry about the "health" of the organization, etc.

Control over one's own immediate work environment. This includes the need to be left alone by the boss and not to be pressured by the managerial controls or by formal policies and practices.

These predispositions of "fit" the requirements of both sub-systems well. The management does not apply pressure on the employees. It does not require that the employees become highly involved in the formal controls. In fact, until recently, it had few formal controls such as budgets. Also, the management, through its intake process, emphasizes that it wants friendly employees who get along with others and who do not complain. Togetherness is congruent with this requirement.

The Informal Employee Culture

If the predispositions are, as we hypothesize, important to the individuals, then we may hypothesize that the employees will tend to try to guarantee their continued expression.24 From research on the relationship between culture and personality we may hypothesize that one way individuals can guarantee the expression of their predispositions is to create a culture that sanctions and approves them.

This is hypothesized to occur in Plant X. The employees in both sub-systems create an "informal employee culture" which sanctions those predispositions common to their sub-systems, thereby making them cultural norms.

Outputs of the informal employee culture. In an industrial organization, management clearly has power over the employees. The employees therefore live in a world where their opportunity for expression of their predispositions would be immediately cut off if the management decided to violate or even throw out the norms of the informal employee culture. However, the probability that management would take such an action is small. They realize that as long as the informal culture is maintained,
the satisfaction of the employees will tend to be high.

High self-expression, in turn, leads the employees to behave in ways desired by management. It is as if the employees say, "Since management is respecting our self-actualization, we will respect theirs." Consequently, the people in Plant X have a high production rate (in the eyes of management), are loyal employees, show low turnover rates, low grievance rates and low upward mobility.

Satisfied management. Data available show that most of the top management is highly satisfied with the outputs of the informal employee culture. For example, two of the top corporate officers interviewed, rate Plant X as "the best" and three rate it as "one of the two best." Turning to the top officials of the plant, we find that they believe the employees are the best producers in the corporation.

Passive Foreman Leadership

Let us now turn to the foremen, all of whom have "come up from the ranks." If our analysis above is valid it implies that the foremen's predispositions must have been influenced by the informal employee culture.

When made foremen, they are told by management that they will be considered successful to the extent that they maintain high production, low grievance rates, and low absenteeism. The foremen realize that the way to get the employees to behave in this manner is to maintain the informal employee culture and not to behave in a way that violates the culture's norms. Thus, 87 per cent of the foremen report that in order to be effective, they must strive hard (1) to keep everyone busy with work (2) to guarantee a fair take home pay, (3) to distribute the easy and tough jobs fairly and (4) to leave the employees alone as much as possible.

Outputs of a successful passive leader. There are at least two resultants from the passive or "understanding" foreman leadership styles. First is an employee-foreman relationship dominated by something we shall call the "psychological work contract." Second, a foreman who as a leader becomes "simplified."

a) Psychological Work Contract. Since the foremen realize the employees will produce best under passive leadership, and since the employees agree, a relationship is formalized between the employees and the foremen which might be called the "psychological work contract." The employees will maintain the high production, low grievances, etc., if the foremen guarantee to respect the norms of the employees' informal culture (i.e., let the employees alone, make certain they make adequate wages, and have secure jobs). This is precisely what the employees need. Thus, the foremen's leadership behavior feeds back to guarantee the perpetuation of the employees' predispositions and their informal employee culture.

b) Simplified Foremen. The impact of the psychological contract is to coerce the foremen to mold their leadership pattern in terms of its requirements. But the requirements are not particularly difficult. Certainly, they do not require complex human relations skills such as self-awareness, sensivity, etc. The psychological contract simplifies the employee-foreman relationship. It also simplifies the foreman's leadership pattern and may simplify him as a human being.

Management Dissatisfaction with Foremen. A (simplified) leadership pattern whose main characteristics are to interact minimally with, and not to pressure, subordinates is apparently not appreciated by the upper management. They wish that the foremen would manifest more active, striving, pressuring, characteristics toward the employees. The foremen find it difficult to convince management that their passive leadership is more effective. Consequently, they tend to become quite passive in their relationship with top management. If it is possible, they prefer (unknowingly at times) to get top management to make any decision that is important. The four top executives in the plant report that they perceive foremen to be of low caliber.

Resultants of a dissatisfied management.

a) Pressure on the Foremen. Management reacts to the passive leadership by applying pressure on the foremen to make them, as one top leader said, "more alive, go-getting, and hard-hitting." The foremen feel the pressure.

b) Employee Visits. Because the executives also do not trust the capabilities of the foremen they constantly make visits throughout the plant talking with the employees.

Foremen's dissatisfaction. Although the management visits act to increase the employee satisfaction, they tend to increase the foremen's dissatisfaction. Many feel they are undercut (77 per cent). They feel they are not truly trusted; that their job is menial with not much responsibility.

We may infer that the foremen's dissatisfaction is caused primarily by four factors. They are pressure from management, low status of their job, undercutting and their role as decision carriers rather than decision makers.

Resultants of the foremen's dissatisfaction are a desire for more money, higher status, and greater control; and reinforcement of the simplified leadership pattern.

Reinforcement of management dissatisfaction with the foremen. The resultants of the foremen's dissatisfaction by no means please the management. They feel that the foremen are not facing reality if they think that the best way to adapt to their (management's) pressure is to become even more "simplified," dependent, and submissive while at the same time asking for greater control. Thus, the foremen's adaptive activities reinforce the management's feelings of dissatisfaction, which only serves to increase the management's pressure, employees visits, etc. The "management-foremen" circuit is now closed, and we have a self-enhancing system. This leaves the management with the employees as their major source of satisfaction, which, in turn, feed back as another reinforcement of the particular hiring process now being used (which management views as highly responsible for the type of employee now working at Plant X). The feed-back completes the circuit to the input process, and we have a self-maintaining system.
This completes the outline of Plant X's social system. Armed with this model, the researcher turned to the task of stating a priori hypotheses about what he would find in Plant Y.

FORMATION OF HYPOTHESES AND FINDINGS

Formation of A Priori Hypotheses About Plant Y

The only information the writer had about Plant Y before conducting the research was that a general "tightening up" process had been instituted in the Plant about one year ago. The aims of the "tightening up" process were, and still are at the time of writing, (1) to cut costs, (2) to develop accurate cost standards, and (3) to develop accurate inventory control. Generally speaking, product costs were to be cut by reducing the costs of production work, wherever feasible. Errors, waste, and poor quality were especially to be reduced. With this knowledge, the writer set before him the analysis of Plant X and began to define any hypotheses that came to mind.

Some of the questions in the writer's mind about Plant Y that influenced the direction of his thinking were as follows. Where will management apply the pressure for a reduced costs? How will sub-systems A and B tend to react to the tightening up processes? What will be the impact of the pressure on the low- and high-skill employees, the foremen, the informal employee culture, the psychological contract, etc.? How will the foremen tend to react? How will the employees' and foremen's reactions tend to effect the other parts of the social system?

Fifteen a priori hypotheses were derived from this exploration. Ten were related to the impact of change upon the employees and five to the impact of change upon the foremen. These fifteen hypotheses were placed in an envelope and were not referred to until the time came to write the results (about one year later). The procedure raises a question as to the degree to which the knowledge of the hypotheses consciously (or unconsciously) influenced the writer's interviewing activities and the subsequent analyses of the interviews. This is a limitation of the project. Although it would have been better to have an independent researcher derive the hypotheses, other valid procedures can be used to partially correct such a situation, because of space limitations they cannot be discussed in detail.25

Suffice it to say that several of the a priori hypotheses that are confirmed, hypothesize relationships among data that are obtained independently of the interview situation and of the analysis made by the writer. For example, a priori hypotheses about turnover, absenteeism, quality of work, and errors in work are tested by data collected by the organization and not by the researcher.

Of the fifteen hypotheses defined before entering Plant Y and of the two defined after the data were gathered (but before the analysis began), ten will be presented below.

Hypotheses About Impact of Changes Upon Employees; Evidence

Hypothesis I: The major points of initial impact of the change upon the employees' world in Plant Y will be on (a) the kitty, (b) pressure from management, and (c) the piece rates. Since the management's concept of motivating employees was to provide them with high wages, high benefits, and high job security, and since top management had a policy of never changing the rate of a job, we predicted that the major points of contact for change in sub-systems A and B of Plant Y would be initially (1) the "kitty," (2) pressure from management, and (3) redefinition of jobs so that piece rates could be altered.

Evidence for Hypothesis I: Since the redefining of jobs could easily have been predicted from knowledge of industry practices, we will simply note that major changes were made in many jobs.

Turning to the kitty, we find abundant evidence that it is one of the first variables which management attempted to change in Plant Y. A policy was defined in the early stages of the change that no one could have more than one day's pay (or about $25) as a kitty. All kitties above that figure were to be turned in and the employees would be reimbursed financially. From that date on, all kitties were to be held by the foremen and wherever feasible an IBM system would be installed so that the "machine" would keep track of each man's daily production, including his kitty.

Hypothesis II: The organizational tightening up process will be resented by the employees in Plant Y. Since the kitty is related to the predispositions of Control, Wages, and Job Security, and since these are important ones for the Plant Y employees, we predicted that the reduction of the kitty and the elimination of "job security" by the employees would be resented by the employees.

Also, since we knew kitties tended to be higher in sub-system A of Plant X, we predicted for Plant Y that the resentment would be higher in sub-system A than in B.

Before these hypotheses could be tested, we needed to establish that the kitty was as well-liked in both sub-systems of Plant Y as it was in Plant X.

That kitties were equally important in sub-systems A and B (Plant Y) can be established by noting that 97 per cent in A and 88 per cent in B respectively described the kitty as a "wonderful thing" for the employee and the company. (The Plant X figures for the same questions are 97 per cent for employees in A and 88 per cent in B.)

"Well, that links up with the tension business that I was talking with you about before. You know there are jobs where we can make out well. The level of energy that you really have to give out to make a day's pay really isn't too much. But then there are jobs that you have to work your ass off just to even up to the day rate. If it isn't sometimes for the extra operations that they give me, 'I wouldn't be able to make a half a day's pay. You know, nobody can put a true rate, nobody can figure these things out right, maybe God can, but we ain't God around here so the kitty really helps both sides. It helps the employee and it helps the manage-

25. See Organizational Behavior, Chapter VI.
ment. It helps the management because if it’s not too accurate, it knows, what the hell, the men and women have a kitty, and they can make up for it. Technically speaking, the guy should turn in his kitty, at the end of every day, but if he doesn’t and he keeps it, then when a bad day comes along, a machine breaks down or something, he still can earn his money.”

Evidence for Hypothesis II: Fifty per cent of the employees in B and 100 per cent in A resented the reduction of the kitty and the elimination of their control (.0001). Some quality examples of reactions given by the respondents are:

“I think it’s a good thing to have. Sometimes you have a good week and you save up a few tickets and you know you can make a good amount of money. Now you can’t. Now we’ve got to turn them in right away. You have to keep track of our own to make sure that they’re not cheating us, and you know, that’s not easy to remember all the jobs you did and the piece rate and so on. I don’t think the fellows like it. They’re much more worried; of course, police the new change, management would increase its rate of initiation of action upon the employees, and since the employees could not respond with an increase in their rate of upward initiation of action, we predicted that a significant change would occur in Plant Y in the variable, “Minimum Pressure from Management.”

More specifically, we predicted the direction of the change to be a decrease in the number of employees in sub-systems A and B (Plant Y) who reported that (1) they were left alone by management, (2) they were their own boss, and (3) the pressure was at a minimum.

Evidence for Hypotheses III: Table I below summarizes the results. All the differences are in

Table I

Comparison of Scores Regarding
Minimum Pressure from Management In Plant X and Y (Expressed in Per Cent)

<table>
<thead>
<tr>
<th>Sub-System A</th>
<th>Sub-System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Management lets us alone</td>
<td>74</td>
</tr>
<tr>
<td>2. We are our own boss</td>
<td>100</td>
</tr>
<tr>
<td>3. Pressure is at a minimum</td>
<td>100</td>
</tr>
</tbody>
</table>

they don’t say much. They can’t say anything. What the hell can you say about it when they’ve done it—there is nothing you can say about it.”

Hypotheses III: The tightening up process will be perceived by the employees as pressure from management. The pressure will be greater in Plant Y than in Plant X. Since in order to institute and

Table II

Comparison of Scores Regarding
Maximum Pressure from Management in Plant X and Y (Expressed in Per Cent)

<table>
<thead>
<tr>
<th>Sub-System A</th>
<th>Sub-System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am never my own boss</td>
<td>0</td>
</tr>
<tr>
<td>2. The pressure is very high</td>
<td>0</td>
</tr>
</tbody>
</table>

but one are statistically significant.

Further supportive evidence was found in the analysis of the question, “The thing I like least about the Plant Y is ...” We found that only 10 per cent of the employees in sub-system B reported that they disliked the pressure most. However, 53 per cent of the employees in sub-system A reported that

566
they liked the pressure least of all (.0002). The differences are even more impressive if we compare the figures above with those of Plant X. None of the employees reported "pressure" as the thing they liked least about the company (.0001).

Finally, another indication that sub-system A employees (in Plant Y) felt more pressured than did those in sub-system B (Plant Y) is obtained when we note that in B, 43 per cent of the employees felt there was little or no pressure now (compared with the day they began to work at Plant Y), and 23 per cent reported more pressure.

On the other hand no employees in sub-system A reported less pressure while 77 per cent reported greater pressure (.0001). The next two hypotheses were not defined a priori but followed from the analysis above. Since we knew that the top management did not tend to apply pressure personally on the employees (see above), and since we knew the foremen were not perceived as pressure givers by the employees, we predicted that:

1) The middle management would be viewed as the culprit by the employees even though the employees clearly assigned the responsibility of tightening up process to the top management.

2) The degree of hostility toward the middle management would be greatest in sub-system A, since we had established that the pressure was greatest there.

The data confirm the hypotheses.

To test Hypothesis 1) above, we note that 92 per cent of the employees in A and B (Plant Y) reported that they did not see the top management often and that the top management was not pressuring the employees. Eight per cent placed the source of pressure on top management (.0001).

On the other hand, 52 per cent of the employees (in A and B) described the middle management executives with such statements as "too dominating," "not stable enough," "too damn easily upset," "confused," and "ready to give hell but not praise" (.0001). Only 4.9% reported them as being "fair" and not dominating.

To test Hypothesis 2), we note that 92 per cent of the employees in B perceived the middle management as pressuring them while 52 per cent of the employees in A reported the same. (Almost statistically significant.)

Hypothesis IV: Self-inflicted pressure will be greatest in sub-systems A and B (Plant Y). Since wages and job security were important for the employees, and since the kitty was reduced, we predicted that the employees would tend to compensate by striving to "make out" (i.e., make a day's pay) every day.

If this is valid, then we would expect a significant increase in "self-inflicted pressure" felt by the employees in sub-systems A and B (Plant Y) over that reported by the employees in Plant X, caused by their need to produce enough every day to make "a fair day's pay."

Evidence for Hypothesis IV: The data confirm the hypothesis. In Plant X the self-inflicted pressure to produce and "make out" every day was 39 per cent for A and 43 per cent for B. On the other hand, self-inflicted pressure in Plant Y was reported by 73 per cent in A (.0034) and 87 per cent in B (.0001).

Hypothesis V: Employees in Plant Y will decrease their emphasis on quality work. Since the employees in Plant Y placed greater emphasis on "making out" than did employees in Plant X, we predicted that a greater number of Plant Y employees would report that they could not achieve a high quality job because they felt pressured to produce quantity in order to make what they felt was a fair day's pay. Moreover, we predicted that a greater proportion of employees in A rather than in B (Plant Y) would report they could not achieve high quality.

Evidence for Hypothesis V: The data confirm the hypothesis. Seventy-one per cent in A (Plant X) and 90 per cent in A (Plant Y) reported that they could not achieve the quality they desired because of the pressure to make money (.0294). Ten per cent in B (Plant X) and 80 per cent in B (Plant Y) reported similar responses (.0001).

Some illustrative comments on quality are:
"If rates are good the man will give to the best of his ability. Now I don't give good quality all the time, Mister. I'll be the first to tell you, I try to give good quality. But I know past a certain point if I don't wise up I won't make a day's pay. If a man gets the feeling that he's a way behind, he's got nothing to get fussy about, because if he gets fussy, he may give good quality but his family will starve."

Hypothesis VI: Plant Y will not tend to feed back to management their negative feelings about the change. Since the employees tended to perceive the foremen as "ineffective" and "second-class citizens," and since they had established no patterned relationships with top management in which hostility was sanctioned, we predicted that the employees would not tend to communicate their negative feelings to (1) the foremen and (2) the members of upper management.

Evidence for Hypothesis VI:
Seventy-nine per cent of the employees reported their foremen's influence to be "not very much," "inadequate" and "none"; "Therefore, there isn't much sense in complaining to those guys because they're pushed as it is and they can't do anything anyway."

We have no direct evidence to test the hypothesis regarding communication to top management partially because the data exists on a more covert level. However, indirect evidence that confirms the hypothesis is available.

During the research project no employees were observed or reported stopping top management and questioning them on the increased pressure whenever the latter made one of their visits.

The managers were asked to predict which departments would have low morale and which would have high morale on the basis of their "visits." The results were interesting. The managers selected the two low-skill departments as having the lower morale because of "generally less work and lower
wages.” When the employees in these departments were asked to describe their own morale, 37 per cent reported “It is high” and 63 per cent reported “It is low.” When asked how they judged “low morale,” 67 per cent reported that low morale exists because of less work and lower wages. Thus, the management’s prediction was relatively accurate (although they misjudged nearly one third of the employees’ view of their morale).

However, in the two high-skill departments they chose as having high morale (again based on wages) 67 per cent of the employees reported they had low morale. The employees attributed their low morale to the increased management pressure and the reduction of the kitty. The point we are making is that management was relatively accurate in its prediction of low morale when this was due to low earnings because both the employees and management used wages as their criterion. They were not aware, however (nor are they informed by the employees), of the impact of their pressure. The management prediction that the employees in the “high morale” departments were satisfied with their wages is confirmed since 93 per cent of the employees reported their wages as “very good” or “satisfactory.” What management was not aware of was that the criterion for low morale was beginning to be changed and that the new criterion would not tend to be communicated to them.

Hypothesis VII: Employees will not tend to create cohesive informal groups. Since the predispositions Togetherness and Non-involvement were very important to the employees, we predicted that they would not tend to develop cohesive groups that could be used to combat management pressure.

Evidence for Hypothesis VII: Indirect evidence that confirms the hypothesis is found when one asks the employees (especially in sub-system A) why they do not communicate their complaints to top management. Fifty-seven per cent responded that no cohesive groups existed within the plant and that the employees did not trust one another. Forty-two per cent of the remainder stated that they could not communicate it themselves: “I wish I knew, funniest damn thing but employees here have never banded together.”

“Kitty I’d say is very very important. Everybody wanted it. Now when the fellows heard they weren’t going to have a kitty, they didn’t like it at all. They didn’t like it at all, believe me.”

Question: Could you help me to understand if the employees have done anything about it?

Answer: Nothing, what can they do about it? They tell you, and that’s it. Is there anything else you can do about it? You see most of the men here they really don’t stick together. It’s a dog eat dog world. Everybody’s trying to get the best jobs, the easiest jobs that pay the most money for the least amount of work. So they’re competing one against the other. Well, when you get that kind of a situation you don’t get people who’ll stick together and work. Another fear is that if they stuck together and management decided they were, you know, looking for a union, then they might be fired.

It is interesting to point out that the Noninvolvement, Apathy, and Togetherness that the employees sanctioned as major components of their culture could also act to prevent them from uniting to combat the further pressure that they were beginning to experience. We may hypothesize further that no hostile employee reaction against management will tend to appear until cohesive groups are formed. In order to develop cohesive groups, however, a major cultural change or changes will have to take place within the employee social system.

Hypothesis VIII: The more mature employees in Plant Y will tend to express a relatively high desire for unionization, (2) be absent. Since the theoretical scheme defines the informal behavior as a-adaptive behavior, and since the necessity of adaptation is greatest where the frustration is greatest, and since pressure will act to make employees more dependent, submissive, etc., we predicted that the more mature employees (in terms of our model) would be the most frustrated. Moreover, on the basis of the theoretical scheme we hypothesized that the high-skill employees would be more directed toward maturity than the low-skill employees.

Evidence for Hypothesis VIII: Let us begin with the last prediction first. If the prediction is valid, then the employees in sub-system A should aspire more toward the mature ends of the continuum than do the employees in sub-system B (Plant Y). Let us first examine the dimensions of individual aspiration. In Table III more of the employees in A express aspirations that are toward the mature end of the continuum than do employees from B. Conversely, more employees in B express aspirations that are toward the immature end of the continuum than do employees in A.

Having established that more employees in sub-system A than B aspired toward the mature ends of the continuum, we hypothesized on the basis of our

---

27. See Personality and Organization, Chapter II, for a definition of the concept "maturity."
theoretical propositions that if, as the framework predicts, the employees in sub-system A are more frustrated than the employees in sub-system B, then:

1) The absenteeism in Plant Y will tend to be higher in sub-system A than in B.
2) The turnover will tend to be higher in sub-system A than in B.
3) The quality of production will be relatively lower in sub-system A than in B.
4) The need for unionization will be higher in sub-system A than B.
5) The expression of "pent up tension influencing outside activities" will be higher in sub-system A than B.
6) All of the predicted changes will be greater in Plant Y than in Plant X.

Let us see if the data fit the hypotheses.

1. The personnel records of Plant Y show that for the period of the research the absenteeism figures were in the direction predicted for the two departments. In sub-system A the absenteeism was 3.6 per cent and 3.9 per cent. In B the figures for the two departments were 3 per cent and 3.2 per cent.

2. The problem was more complex when we considered turnover. The turnover figures (for the same period) for the two departments, in sub-system A were 7.3 per cent and 21.7 per cent. The turnover figures for the two departments in sub-system B were 15.7 per cent and 15.4 per cent. Thus we found that in one department of A the figures were higher as predicted. This was not the case for the other departments. Why?

We can state some hypotheses whose confirmation would require further research to be tested. If we may assume that leaving the organization is a more drastic adaptive mechanism than being absent, then some possible reasons become evident. In both departments (sub-system A) the wages are high. The employee may earn high wages if he "makes out" in piece work. In order to "make out" the employee must learn the skills that are necessary. An analysis of the turnover in the department having the turnover figure of 21.7 per cent suggests that all the men who left were men who, after a few months of apprenticeship, decided that they would never be able to "make out" and asked to be transferred to another department or left the factory. Thus the turnover figure (although in the predicted direction) does not really support the hypothesis. In fact, among the regular employees in A, turnover is less than in B. It is our hypothesis that since the employees in sub-system A are highly skilled and highly paid, they will not find it convenient to leave the plant, but rather try to adapt to the pressure either by greater absenteeism (which is the case) and/or by increased thoughts of unionization and/or an increase in quality of the work. In other words, we hypothesize that since wages are perceived by the employees as high, they will tend to explore other adaptive behavior than leaving the plant. Let us see if our hypothesis are confirmed.

3. Looking at quality, we again found that the data were not as neat as we would have liked. Because of technical reasons one of the departments in B never has quality problems. Their product is either acceptable or scrapped. Unfortunately, no scrap records for each department are kept. Also, in one of the departments in A the technology is such that they can correct errors easily without their being officially recorded.

Consequently, we had only one department in each sub-system that could be compared. But even this comparison raised a problem since A and B have radically different technologies. It is not quite fair to compare the quality figures of low- and high-skill departments.

In order to resolve this problem we compared each department's expectation of what it planned to achieve in terms of quality (during a six month period), with its actual achievement. Thus we compared each department against its own performance, and predicted that the high-skill department would do worse in achieving its own level of aspiration than the low-skill department.

The measure available to judge the quality is called "back work," which simply means work that had to be returned because of poor quality. The figures support the hypothesis.

Table IV

<table>
<thead>
<tr>
<th>Department</th>
<th>Budgeted</th>
<th>Actual</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4 qt 1958</td>
<td>4 qt 1958</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1 qt 1959</td>
<td>1 qt 1959</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>4 qt 1958</td>
<td>4 qt 1958</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1 qt 1959</td>
<td>1 qt 1959</td>
<td></td>
</tr>
</tbody>
</table>

4. Next, we turn to the hypothesis about unionization. Here we found that the most numerous and intense comments on unionization were clearly in sub-system A. Whereas only 10 per cent in B spoke of the necessity for a union, 27 per cent in A reported the need for a union (.0465).

"Well, I sometimes wonder what would happen if there was an organization that would really represent the workers and the management where the people could really sit down together to talk about many of the problems that are bothering both sides. And I think it's a very good thing for the workers to have some sort of an organization where they can get their point of view up to the top. You can get your point of view to the foremen, but you know that doesn't mean any difference because of as I told you, he doesn't have any authority."

"Yes, I'd say there is a hell of a lot of pressure, and a lot of men are getting rundown. A lot of them go home, they're tired, dead tired. You know it's only the guy who really worries about quality that feels the pressure. The other guys, they don't give a damn so they don't care what they give out. I'd say if they're not care-
ful, they're going to have a union on their hands someday."

5. As to the degree to which employees mention that the tension they experience within the plant influences their outside activities, we again found that the hypothesis was confirmed. Only one employee in B (3.3 per cent) mentioned that he was dead tired from the tension whereas 37 per cent of the employees reported that they slept when they went home because of the great amount of tension that they experienced (.0006).

"I'm dead tired when I go home and I sleep. I don't have any desire to do anything. My wife complains about that once in a while, but boy, after a day here with the pressure you have, you're really dead tired."

6. Finally, we turn to the hypothesis that absenteeism, turnover, quality, and unionization should be greater problems in Plant Y than in X. The reader will recall that because Plant X's turnover and absenteeism are so low no figures are kept. We may only conjecture that since the need exists in Plant Y to keep the figures, the absenteeism and turnover are higher.

The figures for quality in each plant are clearly in line with the hypothesis, since all of the appropriate figures for Plant X are in the plus category, whereas none are in the plus category for Plant Y. Perhaps the best indication is that the upper management agree that Plant Y has greater quality problems than does Plant X.

As to unionization in the two plants, the figures again support the hypothesis. Less than 1 per cent of the employees in Plant X reported the need for a union, while 18 per cent (N=60) in Plant Y reported that they desired a union (.0001).

Hypothesis IX: The competition and rivalry will be greater, and the friendship less in sub-system A than in sub-system B (Plant Y). We have shown that as the pressure is increased by management to decrease costs, errors, back work, and as the opportunity to have a kitty is decreased, the employees tend to place greater emphasis on "making a fair day's pay" every day. This follows because without the kitty they do not have anything to fall back on in case they are given "tough" jobs to produce (i.e. jobs whose piece rates, in the eyes of the employees, are low). It is important to emphasize that every time a piece-rate employee comes to work he is anxious about being able to make a "day's pay" during that day. We recall that 82 per cent of the employees reported that their rates were "low" or "some low, some high". This gives us some idea of the degree of uncertainty that the employee experiences. With the kitty his anxiety about uncertainty is greatly diminished. Without the kitty, it is increased.

It followed that there should be increased competition among the employees for the "easy" jobs (which increase the probability of "making out"), and that this competition should lead to interpersonal rivalry and hostility.

If the above was valid, then the increase in interpersonal rivalry and hostility should be greater in sub-system A than in B (Plant Y) because in A the kitties have been almost completely eliminated and the IBM system has taken control of whatever small kitty is permitted. Another reason why the hostility was expected to be greater in sub-system A than B is that the kitty (as has already been pointed out) was lower in sub-system B. Therefore, employees in sub-system B had less to lose from the elimination of the kitty.

Evidence for Hypothesis IX: The data confirm the hypothesis. Seventy-one per cent of employees in sub-system B reported that people were friendly whereas only twenty per cent in A reported the same (.0008). The responses to the corollary question were also in the same direction. Whereas only twenty per cent of the employees in B reported "unfriendly" employees, seventy per cent of the employees in A reported the same (.0001).

Moreover, when we questioned the employees who reported unfriendliness, the overwhelming majority described it in terms of rivalry induced by piece work.

"Oh, I think it has changed quite a bit. In the old plant there was much more friendliness. Now the prices are getting higher and the fellows are squabbling among each other. Also in the old plant there were more ways of skinning a cat and making a buck. Things are tightening up now and you can't make a buck as easily. They are kind of putting the squeeze on, I guess—something maybe they have to do to keep the company going."

Table V

<table>
<thead>
<tr>
<th>Comparison of Friendship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scores in Sub-Systems A and B, Plant X and Y (Expressed in per cent)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sub-System A</th>
<th></th>
<th>Sub-System B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>People are friendly</td>
<td>(N 34)</td>
<td>(N 30)</td>
<td>.0001</td>
<td>(N 90)</td>
</tr>
<tr>
<td>People are unfriendly</td>
<td>90</td>
<td>30</td>
<td>.0001</td>
<td>90</td>
</tr>
</tbody>
</table>

570
Hypothesis X: Plant X will have less employee rivalry than Plant Y. Since we have shown that Plant Y reported more pressure than Plant X, and since we have demonstrated (Hypothesis IX) that the degree of friendship is related to the pressure, we predicted that the reported friendship should be higher in sub-systems A and B in Plant X than in sub-systems A and B in Plant Y.

Moreover, we predicted that the differences in reported friendship should be greater between the sub-systems A in Plant X and Y. Since the degree of friendship is related to the degree of pressure, it follows that the reported friendship should be higher in sub-systems A and B in Plant X than in sub-systems A and B in Plant Y, and that the differences should be greater between the sub-systems A in Plant X and Y.

Evidence for Hypothesis X: The data are presented in Table V. They confirm the hypotheses. Sub-systems A and B in Plant X reported a greater degree of friendliness than did sub-systems A and B in Plant Y. Moreover, the differences are greater between sub-systems A in Plant X and Y.

SUMMARY

To summarize, we predict that as the pressure from management increases, the employees will not tend to combat the pressure as a cohesive group, because their culture rewards apathy, noninvolvement, and togetherness, none of which are norms upon which cohesive groups can be based. Moreover, we predict further that one reaction of the employees will be to place greater emphasis on "making a fair day's pay" every day. This new emphasis will increase competition for the easy jobs and consequently increase the interpersonal rivalry among the employees. This, in turn, will act to further weaken the probability that cohesive groups will arise to combat management pressure.
THE LIMITS OF GROWTH IN NATURAL SYSTEMS*

Edward Goldsmith

The first stage in the study of any aspect of the world we live in is to amass all the relevant material. During this stage, researchers are invariably struck by the amazing diversity of nature. During the second stage, this material is organized; again, researchers are equally struck by the surprising similarities underlying much of this diversity. It is during this latter stage that science really begins. Though the process is at present well advanced within specific disciplines, there has been little attempt until recently to discover what there is really in common between things which at present fall within the domain of different disciplines, such as molecules, cells, biological organisms, societies, and business enterprises. Many scientists reared on Empiricist philosophy would in fact refuse to admit that anything more than a vague analogy could possibly obtain between things that appear so very different. This is not the view, however, of those involved in the new field of "general systems" who prefer to regard these things as different specialized instances of a basic organization which they call a system. A system is defined as something made up of parts in dynamic inter-relationship with each other. I prefer to regard it as an autonomous unit of behaviour—which, by its very nature, must be made up of such parts. Systems, however different they may appear, have a basic structure in common. All are bound by the same set of laws, which they must obey as rigorously as the law of gravity and the laws of thermodynamics. The great value of general systems theory is that, by determining what it is that systems have in common, it becomes possible to develop a general theory of behaviour or a unified science which enables one to examine our biosphere as a whole and determine, in this way, the total effect on it of any local change. It is only when this is possible that science can serve the true interests of mankind. In this article, I shall make use of general systems theory to determine what are the principles governing growth in systems. I shall attempt to show that growth, like all other aspects of systems, is bound by rigorous laws from which there is no escape, and that these laws must apply equally well to the growth of complex systems, such as human societies and business enterprises, as to that of the much simpler systems studied by chemists and biologists.

A System as a Negentropy Machine

The second law of thermodynamics states that our world is running down or moving towards disorder or "entropy." In spite of this tendency, order, or negentropy (negative entropy), has been increasing over the last few thousand million years, during which time complex and highly ordered systems have developed, which we call living things. It is convenient to regard a system as an organization specifically designed to increase order, or negentropy. A system receives an input of low order resources and transforms them into a high order output. During this transformation, waste is generated. However, so long as the reduction of order of that amount of input that has been transformed into waste is less than the increase of order of that amount of input that has been transformed into output, then the system's order, or negentropy, will have increased. Let us consider the conditions in which this process can occur.

Vertical Structure

One of the basic conditions of order is that the parts of a system should be closely linked together. In every different type of system a different set of bonds assures this linkage. Thus, while the bonds holding together an atom are very different from those holding together a cell or a human family, they also have much in common; their function is the same, and also they have limited extendability. This means that they cannot stretch to include things that are too distant from each other. This sets a limit to the size of a system, and when this limit is reached, it can only grow by associating with others to form a new type of system held together by a new set of bonds. When this occurs, a new 'level of organization' is said to have been attained. It is at this point that atoms join to form a molecule, that molecules join to form a cell and that cells join to form a biological organism. There is every reason to suppose that the same principle applies to more complex systems such as human societies. Individuals can be joined together to form a family and families can be linked together to form a small community. However, it is as impossible to create a society out of a whole lot of individuals who are not organized into families and communities as it is to form a biological organism out of atoms that

*This article will appear in a volume being edited by the author, entitled Can Britain Survive?, published by Tom Stacey Ltd., and is reprinted here by permission.
are not organized into molecules and cells; i.e., a system must display its correct structure.

Cancer is an example of the growth of tissue which no longer displays its correct structure. A modern city is an example of the same principle at the level of a society. Demographic and economic growth tends to destroy the essential structure of a society, by tending towards ever increasing urbanization and the development of ever larger social and economic units. The fact that, with growing economic units, people tend to live ever further from their work also means that the bonds linking them with their neighbours become minimal: housing estates can never make sound and stable communities. As we advance to more complex systems, so the bonds become more sophisticated and appear to take longer to develop. Those holding together a family are an obvious example, a fortiori those that hold together a community.

Economic growth invariably means increased mobility. People are no longer treated as members of a community, or of a specific culture, but simply as units of labour that can be shifted around in accordance with the demands of industry. Therefore people are prevented from living together long enough for the necessary bonds to develop. The social disorder—and its various manifestations measured in terms of crime, delinquency, alcoholism, drug addiction, mental disease, etc.—resulting from the erosion of these bonds must set a limit to the desirability and feasibility of growth.

Horizontal Structures

A system must also have an optimum horizontal structure, i.e., a correct ratio must be maintained between the differentiated parts of a biological organism as between the different skills required in a business enterprise. Assuming that all the parts of the system can be quantified, we can then formulate the essential principle of all systems, which we can refer to as the law of optimum value: there must be for every part of the system an optimum value which is determined by that of the other parts. To allow one of these values to increase without reference to the others is to destroy the essential structure of the system and bring about its breakdown. So, if we regard the U. K. as a system, there is an optimum population at any given moment. There is also an optimum number of houses, an optimum number of cars, an optimum standard of living, an optimum differential between the wages paid to different people; there is an optimum longevity and even an optimum amount of social deviation. It must follow that there is no conceivable variable whose value can be increased or decreased indefinitely without bringing about the breakdown of the system.

Economic growth is no exception to this rule. It cannot possibly be regarded as desirable per se, but only in accordance with its effect on the other variables in terms of which we describe our social and ecological system.

Selection: How the Structure is Maintained

The growth of each part of a system, in other words, must satisfy the requirements of the system as a whole. This is ensured by the circumstance that all parts of a system are closely connected by feed-back loops, and that it is the system as a whole that must trigger off the sub-system's behavioural responses. In fact, the environment can be regarded as selecting the required response from among all responses which the system is capable of providing. Thus, within a biological organism, a cell comes into being with a full complement of hereditary material rendering it capable of performing any function within the organism. Gradually, however, it will become specialized in fulfilling a specific function, as a part of the liver or the intestine, for instance. It is in response to the requirements of its environment that a particular cell has developed to fulfill a very small range of the functions it was initially capable of. Its development, in fact, can be regarded as being selected by the environment from among all potentialities it originally possessed. The behaviour of populations obeys the same principle. The environment selects those genetic features of the population that are adaptive to it, to the exclusion of those that are not. In fact, natural selection, rather than being a unique principle, is but a specialized instance of a very general one. In a polymorphous ant colony the same is also true. Only two types of egg are produced; male and female. The female is capable of giving rise to different types of ants, and the actual distribution of those types is determined by environmental selection. In a human society, it is clear how, in ideal conditions, environmental requirements will determine the number of people that should be trained in the different trades and professions.

Environmental selection is clearly essential in a system whose correct horizontal structure is to be respected. It is only in this way that its output remains differentiated so as to correspond qualitatively and quantitively to the requirements of the larger system of which it is part. When growth develops too quickly, however, the parts are no longer selected by the environment. They no longer develop to fulfill specific environmental requirements. They are therefore no longer differentiated. Growth proceeds by multiplication rather than by differentiation, and the system's essential horizontal structure must break down. An example is to be found in the field of education. If selection were allowed to occur normally, the correct ratio would be maintained between the different specialists made available by the educational system, corresponding to the economic, social, and ecological demand for them. As it is, we are producing a vast quantity of young people with specialized knowledge in obscure
branches of learning for whom there is unlikely to be any demand. In this way we are methodically creating unintegrated parts that must of necessity rebel against a system in which they have no place. As economic growth proceeds, so the educative process; i.e., the process whereby members of our society are differentiated culturally, to fulfill specific functions within it, is becoming ever more chaotic. Information is ever less transmitted by the family and the small community and ever more by the State. In this way, it becomes ever less designed to enable people to fulfill their essential differentiated functions as members of their family and of their small community and ever more to fulfill functions as members of a technologically based economy—functions for which there may, in fact, be no real demand. Thus education is foisted on people in a totally indiscriminate way without regard for their intellectual capacity, nor for the sort of society that they are to inhabit. For instance, people living in agricultural or pastoral tribal societies are being provided with an education designed to enable them to fulfill functions within an economy totally alien to their own, and one which, simply on the basis of the approaching world shortage of raw materials, they cannot conceivably aspire to achieve. The basic mechanism of environmental selection is breaking down more and more, and the ever growing chaos associated with the uncontrolled proliferation of culturally undifferentiated people must set a further limit to economic growth.

The Destruction of Cultural Constraint

Social disintegration can be looked at from another angle: the destruction of cultural constraints. Order can be defined as the influence of the whole over the parts. It is also defined as limitation of choice, for the greater the influence of the whole over the parts, the greater must be the constraints imposed on them to ensure that they behave in a way that will further the interests of the whole.

Every system owes its existence to the operation of a specific set of constraints. As it increases order so as to increase its ability to face a given challenge, there is an increase in the constraints applied, and hence a reduction in the range of choices open to the parts of the system. As the system develops and achieves new levels of organization (e.g., as molecules join together to form cells, or as families join together to form small communities, and small communities to form larger ones), new constraints are imposed.

Each system possesses an organization of information, which we have referred to as a "cyberism," which constitutes a model of the environment and at the same time provides the system with a goal-structure and its corresponding constraints. A culture, i.e., that set of beliefs cherished by an ordered society, constitutes its "cyberism" in terms of which it interprets environmental data and mediates responses to them. We can best understand a culture as a control-mechanism that applies the constraints which will ensure that each member of the society behaves as a differentiated part of it. Once these constraints are no longer operative, the society will disintegrate.

One implication of this principle which we might not be too happy to accept is that permissiveness can only be regarded as another word for disorder—as the inevitable sign of social disintegration. Permissiveness appears to be a natural concomitant of social and economic growth. It is the result of the breakdown of those ethical codes associated with the culture of small, relatively isolated communities, and it is inextricably linked with the other symptoms of social disorder.

Conservation of the Environment

Social disintegration can be looked at from yet another angle. It is essential to realize that a system must provide the ideal environment for its parts, since the only reason the parts were developed was to fulfill specific differentiated functions within it. People are part of a family system, which is, in turn, part of a social system, which in its turn is part of a vaster ecological one. One must therefore regard people as having been developed (phylogenetically and culturally) to fulfill specific functions within their family, society, and ecosystem, and one must assume that it is in fulfilling these functions that people obtain maximum satisfaction. To do this, they must be provided with the appropriate information, and also with the correct environment, or rather, one whose characteristics are maintained within the required parameters. When these conditions obtain, people will behave in that way which favours their survival or the increased homeostasis of the system they belong to. However, if the system does not provide the optimum environment, or if the information is not appropriate, then they will behave in a way which will tend to lead to the system’s disintegration. It is for this reason that a stable society requires little or no government, while the more unstable it becomes, the greater the need for autocracy in all its forms.

Unfortunately, what constitutes a satisfactory environment for human beings has never been properly determined. One would have thought that this would be one of the principle goals of sociologists. However, they have not yet got around to thinking in these terms. Clearly, the external environment must have certain basic features, such as the availability of food, water and air, but the presence of the physical necessities of life alone does not suffice to create a satisfactory human environment. Man also needs a satisfactory social environment, which involves the maintenance of the correct social structure. He needs a family, a
small community, probably a larger one, and he certainly needs enemies. If he is not provided with them he tends to invent them. In this way, when the Comanche Indians were put on reservations, they simply invented a host of evil spirits to replace the enemies of which they had been deprived.

Man has other requirements which his environment must also cater to. He has a sense of aesthetics. He cannot adapt readily to living in the grey, monotonous surroundings of our urban conglomerations. He cannot work up any enthusiasm for conserving an environment made up of chaotic complexes of concrete blocks, or bleak fields mutilated by pylons, factories, and housing estates. These, however, are the inevitable concomitants of economic growth. It must be true that to create such an environment specifically for the purpose of increasing society's "standard of living" is to sacrifice its long-term stability in the interest of acquiring dubious short-term benefits.

Self-Regulation

We have noted how the environment selects the response that it requires. This selection process will establish an equilibrium situation between the sub-system and the system. However, this equilibrium can be established with varying degrees of homeostasis. The mechanism that ensures that response satisfies the requirements of the sub-system as well as the system, through tending towards their maximum homeostasis, is referred to as a control mechanism. It is basically the same in all systems. It involves detecting data from the system, transducing them into the correct informational medium made use of by the sub-system's cybernism (brain, gene-pool, etc.), and interpreting them within the model of the system that this cybernism constitutes. The corresponding response must be the only one reconcilable with the particular interpretation of the situation. This mechanism of self-regulation must be a feature of all stable systems. It is the only way to ensure that a system's response to its environment, i.e., to the larger system of which it is a part, corresponds to the requirements of both the former and the latter—requirements which it learns to satisfy ever better as its control-mechanism becomes more perfected. It is in fact the only way to ensure that a system conserves its environment.

The relationships between the individuals making up a human family, and the family itself, provide an excellent illustration of how this mechanism of self-regulation achieves the requisite stability of the systems involved. A woman satisfies her basic biological and psychological requirements best by fulfilling her functions as a wife and a mother within her family unit. A man satisfies his basic requirements best by fulfilling his functions as a husband and a father. In other words, behaviour that satisfies the requirements of the sub-system is also that required to ensure the stability of the system—in this case the family. The system is thus totally self-regulating.

There are many ways in which the family can break down, as we have already seen. When this occurs, its members automatically seek substitute satisfactions. The husband may take a mistress, for instance. This may provide him with satisfaction, but it will not contribute toward the stability of his family unit. The self-regulating mechanism will have been impaired. He may also take to drink and to drugs, undoubtedly also as substitute satisfactions. These, however, will tend to reduce his capacity for fulfilling his essential functions within the family unit, and the self-regulatory mechanisms will be further impaired. It is only when these mechanisms have broken down and the family ceases to be capable of self-regulation that outside forces, or asystemic controls, are required to look after its members. In such a situation only help from the community can permit the family, or what is left of it, to survive at all, for it has ceased to be self-regulating. This help, though it may appear to alleviate the suffering of the members of the disintegrating family, will in fact favour its further disintegration by reducing the need for responsible behaviour on the part of the father. This must be true of all asystemic or external controls, including those introduced by modern technology, for by destroying natural self-regulating mechanisms, they must further contribute to the system's disintegration, thereby increasing the need for further asystemic controls.

Integrity

A system, as we have seen, has an optimum structure, no surplus capacity, and the parts are all differential. It is an integral whole, and the destruction of any of its parts can lead to total breakdown. This is a point which has rarely been taken into account at a cultural level. Colonialist powers have constantly interfered in the most irresponsible way with the cultures of the societies they controlled. Missionaries and colonial administrators have tampered with the delicately adjusted cultural systems of highly stable and ecologically sound societies, which they regarded as "primitive" or "barbarous," and often brought about their breakdown. The consequences for the inhabitants of these societies has been disastrous. They usually become rootless members of a depressed proletariat in the shanty-towns we are thereby methodically creating. The consequences for the ecosystem as a whole have been equally disastrous. By reducing order as well as cultural variety or complexity, we have seriously reduced the stability or homeostasis of the world's human population.

Multi-Ethnic Societies

It is not surprising that systems which are sufficiently differentiated, such as biological
organisms and societies, will tend to develop mechanisms enabling them to exclude foreign bodies likely to menace their integrity. At the biological level such devices are known as rejection mechanisms. Experience with organ transplants has revealed that to suppress these mechanisms is to increase one hundredfold the patient’s susceptibility to cancer, i.e., to the proliferation of cells. Mechanisms of this kind are essential at all levels of organization. Of the 3,000 simple societies so far examined by anthropologists, all appear to have laws of exogamy and endogamy. Marriage is forbidden within a restricted family circle but also outside the cultural group, the object being to avoid cultural hybridization and hence the production of sub-systems that are differentiated parts neither of one system nor of another. What is today regarded as prejudice against people of different ethnic groups is a normal and necessary feature of human cultural behaviour, and is absent only among members of a cultural system already far along the road to disintegration. The notion of the universal brotherhood of man is therefore totally incompatible with the systemic approach to human cultural systems. It is as absurd as the notion that the cells making up a vast number of different biological organisms can be shuffled and still give rise to viable biological systems. Economic growth is leading to increased mobility. Industrial countries tend to develop labour shortages and to import labour from elsewhere. In this way quite large ethnic minorities are being built up in many countries. In addition, economic development is tending towards the development of ever larger political units, which often embrace ethnic groups with little in common with each other. All this is creating a very unstable situation, one which can only lead to civil wars and to the massacre of minorities singled out as scapegoats when inevitable economic and social crises occur.

Resources

A system cannot be regarded as stable unless there is a guarantee that the resources on which it depends will always be available. The only way it can ensure this is to live off the interest and not the capital of available resources. In the highly stable societies of our hunter-gatherer ancestors, this was undoubtedly the case. The Plains Indians provide an excellent example. They lived off the vast herds of buffalo and pronghorns without causing any reduction in their numbers. Their consumption of non-renewable resources such as metals was minimal, and timber provided them with all the fuel they required. To satisfy this essential requirement, it is clear that populations must be kept low as must their standard of living measured in terms of their consumption of non-renewable resources. Once we start living off our capital, it is but a matter of time before it is exhausted and our economic system grinds to a halt. The greater the economic growth and our corresponding dependence on these resources, the more dramatic this eventual collapse must be. Conservation is made possible by a philosophy that holds man to be part of nature rather than above it. Primitive people do not regard the possession of a soul, for instance, as a prerogative of man distinguishing him from all other creatures. All have a soul, and often the primitive hunter will pray to the soul of the animal he is about to kill, explaining the necessity for the crime he is about to commit. Seldom, too, will he kill more than he strictly requires. Indeed, it is said in Southern Africa that the bees do not sting the bushman because they know he will take only the amount of honey he requires, never more.

Disposal of Waste

A system can only remain stable if the larger system of which it is but a differentiated part is capable of absorbing its waste products at the rate at which they are being produced. If it grows too quickly and becomes too big, then it will produce more than the larger system is capable of absorbing. It will then be steadily reducing the order of the larger system by replacing its highly differentiated parts with waste or random parts. By destroying the environment in this way, a system is simply spelling its own doom as it cannot survive in an environment made up of random parts, i.e., displaying total entropy. Needless to say, our society is moving in just this direction. Contrary to what many people think, technology does not provide any means of getting rid of waste, only of shifting it from one place to another, and it cannot be long before the global problem of waste disposal presents an insuperable barrier to further economic growth.

Pollution

Toxic substances can be generated during the development of any system, mainly but not entirely in the form of waste products. These must not be produced faster than they can be absorbed by the environment, otherwise the latter will be modified in such a way that the system can no longer adapt to it. Let us not forget that a system is developed phylogenetically (and culturally in the case of society) as an adaptive response to a specific environment. If the latter is changed sufficiently radically, i.e., if those qualitative parameters within which a system is designed to function are no longer respected, then it can no longer survive. Industrialization is bringing about a radical transformation of our environment at an ever increasing rate. It is but a question of time before the accumulation of insecticides, detergents, radioactive wastes, and carbon-dioxide and heat from the burning of fossil fuels, so transforms our world that it ceases to provide a suitable environment for complex forms of life.
Complexity

If there is a tendency for systems to become more and more complex, it is because complexity renders them more stable. Another way of looking at complexity is in terms of variety, assuming that the variants do not occur at random but together constitute an integrated system—though, in the case of a population or gene-pool, the degree of integration is not very high. The greater the variety, the greater the system’s ability to deal with improbable changes. Serious disruption of its basic structure also becomes less likely. A reduction of variety, or a simplification of a system, will thus lead to a reduction in stability.

It is worth noting that the destruction of the numerous cultures of primitive people throughout the world, and the absorption of their cultures, has produced a radical and dangerous simplification at the cultural level of organization, reducing our stability and rendering our species vulnerable to changes or accidents that would normally affect only a small section of it. In agriculture, monoculture is a drastic simplification of plant life. The use of antibiotics and insecticides causes drastic simplifications in that it involves replacing complex controls that normally keep insects in check by crude and indiscriminate killers. Technological processes, when used to replace natural ones, are further simplifications. In all these cases, stability is being reduced and vulnerability increased.

We are forced to accept the unpleasant fact that practically all man’s efforts today are tending towards the simplification of the total ecosystem and leading to our ever greater vulnerability to environmental changes. As this drastic simplification of our biosphere proceeds, so there must be a corresponding probability of the occurrence of plagues and epidemics of all sorts. This sets a further limit to economic growth, for a time must eventually come when they become ecologically and socially intolerable. To understand what is wrong with asystemic control, one must understand that, during the normal process of feed-back development, each individual response is selected not to satisfy a single environmental requirement but a whole set of environmental requirements—to maintain, in fact, the balanced structure of the larger system of which the system concerned is but a differentiated part. Vis-à-vis each individual requirement, the response must therefore be a compromise. For instance, nature does not aim at breeding strains of wheat with maximum yields, as we do. A high yield is only one of the system’s countless requirements. As a result, the strains of wheat devised by nature are adaptive while those designed by us are not. Whereas self-regulating systems tend towards increased stability, ours on the contrary tends towards increased instability, because every time we "solve" one problem, we are, by the same token, creating others. Take the following example.

Due to a faulty diet, people in Britain have developed a new pattern of diseases. Among these is tooth decay, which has got so bad in Britain that there are now some 17 million people with no teeth at all. The correct remedy is clearly to return to a healthier diet. This, however, would mean closing down a large number of food factories and returning to a sounder, non-industrial agriculture. It would mean putting a lot of people out of their jobs, cutting down profits, increasing the price of food, reducing the GNP and the standard of living. Nevertheless, it is clearly what should be done. However, since we are not willing to do it, we must find some gimmick for getting rid of some of the more obvious symptoms of the diseases we have created. This gimmick is fluoridation of drinking water. This may reduce tooth decay, but it does not prevent other side-effects of a faulty diet, such as diabetes, heart and kidney diseases, etc. In fact, all that has happened is that our diet has been rendered that much more tolerable, which means that we are less likely than ever to do anything about changing it. This is true of most technological devices: By helping to render more tolerable the symptoms of the pathological situation we have brought about, modern technology serves but to perpetuate it.

The technological devices developed to control pollution are no exception to this rule. If we succeed in developing filters that prevent all air pollution from the exhaust of motor cars, we will simply have eradicated one of the many symptoms of the pathological situation created by the preposterous proliferation of these machines. By rendering them that much more tolerable, we would then be better capable of supporting the noise they make, the destruction of our towns and countryside with the roads required to accommodate them, and the highly undesirable level of economic centralization that they render possible, etc. Technology is in fact introducing positive feedback into the ever more unstable system that our world is becoming. It thereby renders it that much more unstable and increases in this way the severity of the final catastrophes that must inevitably restore social and ecological equilibrium.

Asystemic controls have the further disadvantage of rendering the functioning of our environment dependent on human efforts and ingenuity. A so-called controlled environment, the technologist’s dream, is one in which all the self-regulating mechanisms of our biosphere have been replaced by the externally regulated asystemic mechanisms of our technosphere. In such conditions, we would have to depend for our drinking water on desalination.

plants and sewage works; for our food supply on factory farms and processing factories, and for the very air we breathe, on vast plants to filter noxious gases out of the atmosphere. Have our technocrats ever bothered to think of the almost unbelievable vulnerability of such a society? An industrial dispute, an act of sabotage, a technical failure, a shortage of some key resource, or any one out of countless such incidents would deprive us of the basic necessities of life. The recent sewage strike in Britain revealed the extreme vulnerability of a society depending too heavily on technology for its survival. Clearly, there must be a limit to the extent to which external controls can be allowed to replace self-regulating ones, hence to the extent to which the technosphere can be allowed to replace the biosphere.

Conclusion

By using general systems theory, it is possible to point to a large number of limits to the desirability and feasibility of growth in all systems. Such limits must apply equally well to growth in technologically based human social systems. To suppose that technology can permit permanent growth is to refuse to face basic scientific facts. It can indeed enable it to proceed temporarily beyond these limits, but only at the cost of increasing the instability of the social and ecological system of which we are part, and of increasing thereby the likelihood and the severity of the cataclysm that must eventually restore a stable situation.
c.

APPLICATION OF SYSTEMS THEORY TO INSTITUTIONS
THE IMPACT OF CYBERNETIC INFORMATION TECHNOLOGY 
ON MANAGEMENT VALUE SYSTEMS*

Richard F. Ericson

INTRODUCTION

This analysis is essentially a further development and an updating of one aspect of a previous paper where it was indicated that, as I saw it, a "universally viable philosophy of management" must be built with reference to at least five major constructs: (1) a basic attitude of tentativeness and contingency; (2) an ultimate reliance on reason; (3) an emphasis upon holism; (4) a transcultural identification; and (5) a transcendent concern for human values. It is this last philosophic dimension with which the present paper is concerned.

I. THE HISTORICAL BACKGROUND: SOCIOECONOMIC VALUE POLARIZATION AND THE MANAGER'S ROLE DILEMMA

From the days of the "Robber Barons" to those of the recent judgment that General Motors' welfare and the country's are correspondent, there has been more or less explicit struggle to harmonize managerial with social values. Indeed, some have argued that there is a basic socioeconomic ethical dilemma deeply rooted in American culture. On the one hand, the Calvinistic tradition extols acquisitiveness and competition. On the other, the Judeo-Christian tradition, philosophically crystallized in the Golden Rule, highly values sharing and cooperation. In political-economic terms, the contrast is between the Conservative and the Liberal. The 1930's debates between Pigou and Keynes are paralleled to a degree in those between, say, Milton Friedman and Walter Heller or John Kenneth Galbraith.

To put it more generally and more bluntly, there seems to be a "Gospel Mission"/"Madison Avenue" schizophrenic dimension in our national character. And this, of course, manifests itself in the ideologies of corporate management. In Selekman's justly noted epigram, corporate managers are constantly confronted by the "technical must versus the ethical ought." Out of such stuff are managerial ulcers made.

Still it seems a clear trend since World War II that concern with the "social responsibilities" of management has been the dominant theme. The triumph of Keynesianism is reflected in the so-called "New Economics." Whether because of, despite, or merely coincident with this fact, our affluent society has doubtless been an important factor in providing a basis for managers to attend to the longer-run economic and the social impacts of their decision making.

The past two decades are themselves divisible in two rather distinct developmental phases. During the period following World War II and into the fifties, managerial concerns in this regard were primarily directed inward. The "human relations" approach to management process was ascendant. But in recent years greater attention has focused on managerial evaluation of the impact of corporate policies upon the wider world within which the corporation pursues its goals. The infamous reputation achieved by Defense Secretary-designate Charles Wilson's professedly innocent epigram reflects this shift in focus.

Thus, our concern in this paper is with the emerging value milieu in which corporate policy formulation occurs. This emergent concern was characterized in 1963 at the 13th International Congress of Management as follows:

But as the congress drew to its end, another mood became apparent. Unspoken but real, it lay behind the words of many of the speakers, it was present in conversations in corridors and hotel rooms. It was common search for philosophical bearings in an era that is putting businessmen more and more on the defensive, even as it places greater demands on their managerial skills.

"Technique, efficiency, management, results ... But what's the whole blooming thing for?" asked Lebanon's Dr. Charles H. Malik, former president of the United Nations General Assembly. His question struck to the heart of a question familiar to many U.S. executives—but newer to their counterparts from abroad.

Recent social history of the United States has brought the dilemma into clearer view. Slowly the

*Based on a paper presented at the 15th International Meeting of the Institute of Management Sciences, Cleveland, Ohio, September 12, 1968. This version is printed by permission of the Institute. The complete paper appeared in Management Science Application, October 1969.

socioeconomic policy microscope has swung to put an examination of corporate decision behavior into sharper focus. Two examples: The effluent of an affluent society is being traced back to its polluting origins, and corporations are often found at the source. (Incidentally, so also are governments.) And now that deeply infected complex of social wounds referred to as the "urban problem" is being scrutinized, in part at least, to assess the contributions to the crisis made by typical corporate policy. The question is: to what degree has corporate behavior increased our "social pollution"?

Yet even as corporate managers acknowledge the role enterprise institutions must play in social renovation, they are confronted with the Calvinistic/Judeo-Christian dilemma again. The rationale of the primacy of the stockholders' claims still permeates. More than two years ago one observer found that "In practice, fear of controversy and the brand of the 'do-gooder' are probably the biggest obstacles to the spread of the new philosophy.... Indeed, some businessmen go out of their way to find economic reasons for their good works." We shall return to this question later in this paper; however, it might be well even at this point to note that as "business wrestles with its corporate conscience"—to use a phrase titling a contemporary review of these problems—it is possible even to identify that "lunatic fringe" of corporate managers who believe that the "rationale for committing corporate resources to the cities is basically uneconomic...business owes an obligation to society in these critical years." Here-sy indeed.

II. MANAGEMENT VALUES, CYBERNETIC TECHNOLOGY, AND ORGANIZATIONAL ECOSYSTEMS

The central concern of this paper is with the future direction of the impacts of corporate decisions upon the larger social system of which they are a part. Over the past two decades management and organization literature abounds with specifications of the nature of the "social responsibilities of management," and with concern for "corporate accountability."

But even among those managers that have honestly made an effort to relate more effectively to the social system and its needs, there have been at least two kinds of difficulties: (1) they have been hard put to find really satisfying philosophic rationale for such "extra-curricular" behavior, in view of the traditional rationalizations of the corporate enterprise; and (2) management experience has almost inevitably generated an uncomfortable feeling that the corporate/societal interface is such a "no man's land" that there was great difficulty in evaluating the impacts of such effort, much less demonstrating its worth. Only in such legitimized cases as philanthropy to colleges and universities has some measure of satisfaction and response been generated. And even here there was a history of shareholders' suits to be experienced.

The argument this paper explores is that we seem suddenly to have entered a "new era." The recent advent of institutions such as the World Future Society and the Institute for the Future emphasizes the possibility that man has entered a time when he can do more than react and adjust to "external" events. Now we hypothesize that man has the ability to design his future, on the basis of application of rational criteria for selecting and implementing an optimal model. In this spirit we observe that corporate managers are increasingly striving to translate pious words, uttered over the past two decades, into action—in the context of the "urban crisis" at least.

We conjecture that the interface between corporate subsystem values and general social system values is now more readily comprehensible in terms of computerized feedback and dynamic process models. Thus cybernetic information technology has created new potentials for the more effective articulation of management values and social needs. It has made possible the construction of operationally viable holistic systems models of ever-widening scope and complexity.

This prelude to our analysis must also take notice of the fact that the corporate system itself has been evolving in such a way as to give real urgency to our concerns here. The advent of the "conglomerate" is once again raising the question of whether the size of our major "private enterprise" institutions violates optimal relation to the whole. While in years past the "social responsibilities of business" theme was argued primarily on moral and obligational grounds (corporations should be "good citizens"), now holistic systems analysis is demonstrating that organizational success—at the level of the large-scale corporation at least—is as much a product of effective management of the organization societal interface as anything else.

Evidence supporting the hypotheses that managements are ecosystem oriented is available. To quote the essence of three articles which have appeared during the past two years, from reputable non-academic sources:

(1) Whenever an organization isolates itself from the objective values of the outside environment, whether it be stable or changing, it has taken the first step toward its deterioration. A large
organization should not be reluctant to change its methods, goals, or viewpoints. For only when a system is responsive to outside influences does it prevent a conflict of goals leading to decay.6

(2) Each management system has a rationale inseparable from other systems and inseparable from the context of business activity. This need for congruency extends down to the least subsystem of management... A management system provides a tie between internal company activity and the external environment... Promoting congruency between the company and the economic world is the rationale underlying management systems. The circle of company activity has to match in every possible phase the alignment of the world outside. The closer this correspondence, the greater the purchase the company has on survival and, possibly, on growth. Where congruency exists between the substantive demands of the external world and the response of the individual business enterprise, competitive excellence is normally assured.7

(3) Just as a living organism survives by adapting to stimuli—both from within and without its environment—a corporate organization survives by adapting its purpose, content and structure to changes taking place within itself and its ecosystem....

As the selective agent of change, the manager is the catalyst, the mutation selector... He can become the "Maxwell Demon" of adaptive mutations and increase the probability that adaptive response will succeed.

This means that the manager is like the ecologist, that specialist in the relatively new interdisciplinary science devoted to understanding the interrelationships between living organisms and their environment.... For the species or community to exist symbiotically in its ecosystem, it must be able to respond creatively and adapt to the forces generated by itself, by all other species, and by the physical environment of its ecosystem. (Underscoring added.)

The manager's role... requires understanding of the source of creative ideas and the importance of communications among specialists, plus knowledge of the needs, limitations and purpose of the whole...8

III. VALUE THEORY AND CONTEMPORARY VALUE SYSTEMS

A. Value Theory

Review of the ethics and values literature suggests that about the only commonly-held definitional proposition is that somehow the concern is with normative rather than existential descriptions of behavior. One of the most lucid modern scholars in this field, Samuel Pepper, initially identifies ethics as "the study of the criteria of good and bad conduct," later modified to read "the study of the structure and operation of selective systems bearing on human activity and the lines of legislation running through them."9

This is in contrast, for example, with the viewpoint of another modern value theorist, Robert Hartman, who follows G. E. Moore's classical dictum "Good is good, and this is the end of the matter." Building upon this a priori premise, Hartman's "axiological science" rests on the proposition that "a thing is good when it fulfills the definition of its concept."10

We shall not allow ourselves to become mired in the ethics and values theory controversy epitomized by Pepper versus Hartman. Rather, in the spirit of pragmatism so typical of managerial behavior, definitions of value such as the following seem most appropriate for our purposes:

Values are normative standards by which human beings are influenced in their choice among the alternative courses of action which they perceive.11

A value is a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable which influences the selection of available modes, means, and ends of action... value may be defined as that aspect of motivation which is referable to standards, personal and cultural, that do not arise solely out of immediate tensions or immediate situations.12

9. Pepper reaches this modified definition on the basis of his examination of the greater ethical schools: the five Empirical types (comprising Cultural Relativity, hedonism, the social situation theory, the self-realization theory, and the theory of evolutionary ethics); the Intuitive and Formal theories; and the Linguistic Theories. After dismissing the latter two, Pepper develops a "Social Adjustment" theory, which is based upon deriving "lines of legislation" among the various "natural norms" developed by each of the empirical theories. On this basis, he contrasts the dynamics of "purposive drives" which characterize the "Functional Authoritarian" society, with the dynamics of "evolutionary process," which he argues characterize the "Individualistic Democratic" society. Thus, decreases in social pressure in the five selective systems associated with the five empirical theory schools-cultural pattern, social situation, personality structure, personal situation, and purposive structure-increase the tendency to realize Individualist Democratic values: open society; freedom; initiative; enjoyment; and pleasure. Correspondingly, increases in social pressure with respect to the selective systems leads to Functional Authoritarianism's values: functional society; security; discipline; achievement, and success.10
12. Clyde Kluckhohn, quoted idem., p. 10.
... values (are) conceptions of desirable states of affairs that are utilized in selective conduct as criteria for preference or choice or as justifications for proposed or actual behavior. ... Values are closely related, conceptually and empirically, to social norms.13

More precisely, as we approach the application of value analysis to managerial behavior, we will be drawn toward the ethical problem orientation of the modern operational philosopher, which Anatol Rapoport argues is characterized by "(1) the principle of parsimony [Occam’s Razor: "concepts are not to be multiplied beyond necessity"]; (2) freedom of choice among means in the pursuit of ends; (3) a knowledge of consequences of one’s choices; and (4) a knowledge of one’s motivations in making the choices one makes."14 Our approach here will utilize a "pragmatic logical positivism" or "operational philosophy," defined as:

... the philosophy of action-directed goals. It starts with logical analysis. But logical analysis is not conducted in a social vacuum by electronic computers. It is conducted by men and women with beliefs, prejudices, and convictions. The results of logical analysis must inevitably produce an impact on these beliefs and convictions and induce to further action. ...15

B. Evidence of Contemporary General Value Orientations

A recent insight into the value milieu in the United States is provided by the election-year research conducted a few months ago by Milton Rokeach.16 He compared eighteen "terminal values" aspired to by each of the seven groups of respondents (one group for each of the then candidates: Kennedy, McCarthy, Johnson, Rockefeller, Nixon, Reagan, and Wallace), as well as for a like number of "instrumental values" favored by each group. The most significant general finding reported was that

The seven American groups compared here are on the whole remarkably alike in their systems of values. Sharper differences would have been obtained, for example, had we compared the more diverse French political groupings.

Since Mr. Humphrey was not explicitly included among the seven candidates listed by Rokeach, we use Mr. Nixon's devotees as a more or less representative indication of present ranking of general social values:

<table>
<thead>
<tr>
<th>Terminal</th>
<th>Instrumental</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Family security</td>
<td>(1) Honest</td>
</tr>
<tr>
<td>(2) A world at peace</td>
<td>(2) Responsible</td>
</tr>
<tr>
<td>(3) Freedom</td>
<td>(3) Ambitious</td>
</tr>
<tr>
<td>(4) Wisdom</td>
<td>(4) Forgiving</td>
</tr>
<tr>
<td>(5) Salvation</td>
<td>(5) Courageous</td>
</tr>
<tr>
<td>(6) Happiness</td>
<td>(6) Broadminded</td>
</tr>
<tr>
<td>(7) Self-respect</td>
<td>(7) Helpful</td>
</tr>
<tr>
<td>(8) Sense of accomplishment</td>
<td>(8) Loving</td>
</tr>
<tr>
<td>(9) National security</td>
<td>(9) Capable</td>
</tr>
<tr>
<td>(10) A comfortable life</td>
<td>(10) Clean</td>
</tr>
<tr>
<td>(11) Equality</td>
<td>(11) Self-controlled</td>
</tr>
<tr>
<td>(12) Inner harmony</td>
<td>(12) Cheerful</td>
</tr>
<tr>
<td>(13) True friendship</td>
<td>(13) Independent</td>
</tr>
<tr>
<td>(14) Mature love</td>
<td>(14) Polite</td>
</tr>
<tr>
<td>(15) A world of beauty</td>
<td>(15) Intellectual</td>
</tr>
<tr>
<td>(16) Social recognition</td>
<td>(16) Logical</td>
</tr>
<tr>
<td>(17) Pleasure</td>
<td>(17) Obedient</td>
</tr>
<tr>
<td>(18) An exciting life)</td>
<td>(18) Imaginative</td>
</tr>
</tbody>
</table>

In commenting on the comparative outcomes as between the seven candidates' groups of supporters, Rokeach said in that article

The major differences observed among the seven American groups are primarily differences in the judged importance of relatively few values, particularly equality and salvation among the terminal values and broadminded and clean among the instrumental values. (Underscoring supplied.)

Thus, while for Nixon's followers "salvation" was the fifth most important terminal value, it was twelfth on the list generated by McCarthy followers, placed thirteenth by Rockefeller's, and was fourth on the list of Wallace's.

As we turn to managerial values per se, Rokeach's conclusion from this research into basic values of the electorate may have some meaning for the increasing attention given to management values and the social impacts of the historical profit-oriented policies of our corporate sector:

The appeals that various presidential candidates have for the American public arise primarily from the different images projected, involving a relatively small number of values.

C. Management Value Systems

Managerial value systems are, of course, essentially subsets of the general value system within which the manager has been acculturated and/or operates. Empirical evidence for this general assumption has recently been provided by the comprehensive cross-cultural analysis of management in the work of Haire, Ghiselli, and Porter.17 They used the following four vectors in their "attempt to cover four distinct areas of disagreement

15. Ibid., p. 222.
between the traditional-distinctive and the democratic-participative approaches": (1) capacity for leadership and initiative; (2) sharing information and objectives; (3) participation; and (4) internal control. An indication of the tensions generally inherent in managerial value systems in contemporary cultures is found in their major conclusion that

Despite the wide variety of cultures represented by the fourteen countries in our sample, there was considerable similarity among the managers from these various countries on the major finding from this part of our questionnaire: The tendency to agree with the belief that the average individual has a capacity for initiative and leadership, and, at the same time, a tendency to agree that the best methods of leadership are the democratic-participative methods (p. 21).

... Indeed, in the United States, though it is virtually "un-American" to declare oneself in favor of centralization, and though a public declaration in favor of unilaterality is anathema to most managers, the fact remains that the tremendous majority of American businesses are still run on a tight rein—highly centralized and closely supervised (p. 172).

This leads Haire, Ghiselli, and Porter to conclude that "there is ample evidence here that management everywhere needs to examine implicit assumptions underlying policies and practices espoused" (p. 173).

The authors of Managerial Thinking conjecture (p. 21) that this managerial value ambivalence may reflect a "sort of partial digestion of the exhortations of group-oriented consultants and professors of management during the past decade." Certainly much of the "human relations approach" to management referred to above has hortatory. For example, Robert Golembiewski has exorted that "behavioral conditions associated with high output consistent with Judeo-Christian values which should guide man-to-man relations in organizations," are that:18

1. work must be psychologically acceptable, generally non-threatening;
2. work must allow man to develop his faculties;
3. the task must allow the individual room for self-determination;
4. the worker must influence the environment within which he works; and
5. the formal organization must not be the sole and final arbiter of behavior.

Golembiewski's exposition clearly derives from the analysis of Chris Argyris, as developed in his mid-1950's volume Personality and Organization.19 There Argyris decried the "fundamental lack of congruency between the needs of healthy individuals and the demands of the formal organization." Argyris has recently tried to provide firmer underpinnings for his thesis by utilizing Hartman's axiological approach to value prescription. We shall consider this later in the paper. But our immediate need is for an indication of contemporary management value sets, rather than for prescriptive admonitions. Let us therefore consider first an analysis growing out of the experience of a thoughtful executive, and then turn to a recently reported study of the personal value systems of American managers.

Abram Collier has sought to illuminate the value conflicts with which modern executives are confronted by skillfully devising a series of scenarios in which the dilemmas are confronted by corporate managers. He concludes his book by indicating five value clusters "which are above, beyond, or before profits":20

1. the "A-values": self-teaching; hard work; self-realization; personal responsibility, i.e., "vocation." Looking at things from our own internal subjective point of view, we develop certain goals; how we go about resolving them, the degree of resolution we display, all contribute to this cluster of values.
2. the "B-values": organizational skills; administrative genius; communicative powers; integration of mental and physical health. ... If the "A-values" deal with ends, the "B-values" deal with means ... bodies, both individual and corporate, are but means to other ends.
3. the "C-values": facts of the market; professional training: legal realism; historical objectivity; looking at things impersonally and objectively ... metaphor of the ivory tower as the place from which we can see what is going on in order to get the "true" picture.
4. the "D-values": pupil-centered teaching; customer-centered selling; participative management; self-transcendence ... these values represent the viewpoint that looks at others impersonally but from their subjective situation; and
5. the "Omega-values": ultimate values. ... The capacity to adapt to changed circumstances; the ability to integrate different and opposing points of view; the power to rise above the limits of any one of the four value structures (above).

The question is, how does this exposition, based upon managerial experience, square with value orientations found in actual management research?

George England has studied the value systems of over a thousand American managers, and has developed a theoretical model for analyzing

---

the impact of values on behavior. He views a personal value system as "a relatively permanent perceptual framework which shapes and influences the general nature of an individual's behavior." Using five clusters of value concepts—(1) goals of business organization; (2) personal goals of individuals; and ideas associated (3) with people, (4) with groups of people, and (5) with general topics—England uses four scales to represent four modes of valuation. The power mode ("important-unimportant" scale) is primary, with the other three modes—pragmatic ("successful" scale); ethical-moral ("right" scale); and affect or feeling mode ("pleasant" scale)—employed to ascertain why managers thought certain value laden concepts were important or unimportant.

England's work serves to justify the general value position taken at the outset of this paper, for he found that "as a total group, managers' primary orientations are pragmatic; that is, when managers view some concept as important they also tend to view it as successful." Generally speaking, managers' secondary value orientations are moralistic and ethical. Specifically, England's analysis produces the following categories of value clusters:

(1) Operative Values: those rated as "highly important and viewed as successful";

(2) Adopted Values (Situationally Induced Values): those that have been observed as being successful in the manager's organizational experience but which he finds difficult to internalize and view as being of high importance;

(3) Intended Values (Socio-culturally Induced Values): those that have been considered as highly important by the manager throughout most of his life but do not fit his organizational experience; and

(4) Low Behavioral Relevance Values: those that would not be expected to influence a manager's behavior to any large extent since they are not considered important and do not fit the pragmatic orientation of managers.

On the basis of his findings, we may associate each of these value concepts with their relevance to managerial behavior as in Matrix I.

The array of values found in the matrix cells confirm many of our intuitive understandings of the value orientations of American managers, e.g., that managers are "action oriented" and "pragmatic" in their behavior. But the analysis also highlights some rather disturbing data, e.g., many of those value concepts in the "Intended" and "Low Relevance" columns. They emphasize the value dilemma and role conflict experienced by American managers. Note that such concepts as employee welfare, government, trust, loyalty, honor, dignity, and individuality are included as values which "have been considered as highly important by the manager throughout most of his life but do not fit his organizational experience." Here is independent research support for Argyris' dismal description of the personality/organization conflict.

Since this paper is primarily concerned with management values relating specific systems—namely, corporate systems—to their ecosystems, we refer finally to Bernthal's Value Hierarchy decision model. He recognizes four levels of values, in terms of which decision criteria should be applied. In ascending order—i.e., from the immediate to the remote environment, from the decision maker's standpoint—they are (1) the business firm, seeking ownership welfare (profits, survival, growth); (2) the economic system, seeking consumer welfare (allocation of resources; production and distribution of goods and services); (3) society, seeking social welfare ("the good life"; culture; civilization; order; justice), and (4) the individual, seeking individual welfare (freedom, opportunity, self-realization and human dignity).

But the tough question is, operationally speaking: Just how are these various levels of values to be reconciled in specific decision processes? To that query, this literature provides precious little guidance.

IV. CYBERNETIC SYSTEM MODELS, MANAGEMENT ECOSYSTEMS, AND INFORMATION TECHNOLOGY

Having reviewed major recent contributions to our understanding of social value systems, and the managerial subclass, it remains to consider the other prong of our concerns—cybernetic information system theory and technology—before evaluating their actual and latent impacts upon managerial value systems.

We shall consider three examples of cybernetic system models. We first turn to Jay Forrester's quasi-closed-system Industrial Dynamics, not because his model is necessarily the most promising from a management ecosystem standpoint, but rather because (1) it is an operationally oriented, computerized (quantified) model; (2) it is now into its second decade of development; (3) it is widely emulated as a management instructional tool; and (4) it enjoys considerable success "in the field."

To most readers the essence of the industrial dynamics model is clear. It has been adequately expounded in the literature over the past decade, and was the subject of a major portion of

THE IMPACT OF CYBERNETIC INFORMATION TECHNOLOGY

MATRIX I


<table>
<thead>
<tr>
<th>CATEGORIES OF VALUE CONCEPTS</th>
<th>OPERATIVE</th>
<th>ADOPTED</th>
<th>INTENDED</th>
<th>LOW RELEVANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Goals of Business Organizations</td>
<td>High productivity, Industrial leadership, Organizational stability, Profit maximization, Organizational efficiency, Organizational growth</td>
<td></td>
<td>Employee Welfare</td>
<td></td>
</tr>
<tr>
<td>(2) Groups of People</td>
<td>Employees, Customers, My Co-workers, Craftsmen, My Boss, Managers, Owners, My Subordinates, My Company, Stockholders, Technical Employees, Me, White Collar Employees</td>
<td>Labor Unions</td>
<td>Government</td>
<td>Social Workers, Laborers, Blue Collar Workers</td>
</tr>
<tr>
<td>(3) Ideas Associated with People</td>
<td>Ambition, Ability, Skill, Cooperation</td>
<td>Aggressiveness, Prejudice</td>
<td>Trust, Loyalty, Honor</td>
<td>Obedience, Compassion, Tolerance, Conformity</td>
</tr>
<tr>
<td>(4) Personal Goals of Individuals</td>
<td>Achievement, Job Satisfaction, Creativity, Success</td>
<td>Influence, Power</td>
<td>Dignity, Individuality</td>
<td>Leisure, Autonomy, Money, Security, Prestige</td>
</tr>
<tr>
<td>(5) Ideas About General Topics</td>
<td>Change, Competition</td>
<td>Conflict, Risk, Force</td>
<td>Property, Rationality, Religion</td>
<td>Authority, Caution, Conservatism, Equality, Liberalism, Emotions</td>
</tr>
</tbody>
</table>

the March 1968 issue of Management Science. A few general comments will focus the model in terms of our interest here.

In that above-mentioned publication Forrester describes industrial dynamics as having systems implications far beyond those implied by its name: as "the science of feedback behaviour in social systems," as the "interrelation and the extension of feedback system concepts to apply in the multiple-loop, nonlinear systems to which the social processes belong," and, again, as a "philosophy of structure in systems." This sounds promising for our purposes, but there is a difficulty. In Forrester's model, the feedback loop is considered the "basic system component." He writes (p. 407, Mgmt. Sci. March 1968)

The feedback loop is fundamentally a closed process in which a decision, acting through time delay and distortion, influences the state of the system which, after further time delay and distortion, is detected as the observed state of the system. . . . There are no closed loops of significance to the particular study going from inside the system to outside of the boundary and returning. (Underlining supplied.)

Of course the choice of system boundaries is a judgmental decision, and one could be liberal in his notion of the system's range, if he would study system/environment interfaces; for after all, at any given level, these interfaces are but subsystem boundaries of the larger system of which they are constituents. But Forrester's admonition to the model-builder is to select system boundaries according to what might be viewed as the application of the "law of parsimony" (cf. Rapoport, page 90 above). Forrester says (p. 408)

. . . For an industrial system model, the boundary should include those aspects of the company, the market, the competitors, and the environment which are just sufficient to produce the behavior being investigated. . . . The concept of the closed boundary seems elementary yet it is apparently hard to
Yet industrial dynamics does have promise for use in ecosystem study, for Forrester goes on to remark (p. 410)

... We believe that the structure (of industrial dynamics) will come to be recognized as having simple elegance, universality, and a fundamental character common to a very broad range of systems running from physical devices through medicine and psychology to social and ecological systems.

The Homomorphic Model of Stafford Beer is a more straightforward general systems formulation than is Forrester’s Industrial Dynamics. For example, in describing the nature of a scientific model, Beer indicates the following: Insight leads an observer to compare a managerial with a scientific situation; a perception leads to analogy of the conceptual models which can be constructed from each situation; a homomorphism (a many/one transform; a reduction in variety) produces a rigorous formulation of each conceptual model, which enables the investigator to study the isomorphies of the two formulations; and from this a common scientific model is derived, through the process of generalization. Or, in Beer’s words: “A scientific model is a homomorphism on to which two different situations are mapped, and which actually defines the extent to which they are structurally identical.”

Schematic I is Beer’s “sketch for a cybernetic factory.” It demonstrates how a company, interacting with its environment, can be described by the operations researcher in cybernetic modeling terms. Clearly, it represents the kind of concern with environmental interfaces that preoccupies us here, as we consider impacts upon management values.

Finally, we turn to the Living Systems approach to general systems theory, as expounded by James Miller. Of his organismic system modeling Miller says (p. 234)

My analysis of living systems uses concepts of thermodynamics, information theory, cybernetics, and systems engineering, as well as the classical concepts appropriate to each level. The purpose is to produce a description of living structure and process in terms of input and output, flows through systems, steady states, and feedbacks, which will clarify and unify the facts of life.

The fact that Miller’s is explicitly an open systems model clearly increases its relevance for our analysis. Suffice it to remark that he identifies three critical subsystems of living systems (p. 338): matter-energy processing subsystems,
THE IMPACT OF CYBERNETIC INFORMATION TECHNOLOGY

(3) Systems flexibility for new applications will be vastly increased, and costs greatly reduced, through a broad range of new peripheral equipment developments;

(4) Significant cost reduction and vastly expanded use of random-access files and memory will permit the drawing together on an integrated basis of the data needed to manage and operate the company, and provide instantaneous and flexible access to it;

(5) A totally new data storage and processing capability—graphic storage and processing—will become economical and commonly available;

(6) Information storage and retrieval of technical, management, and general data will become an increasingly important aspect of information systems; and

(7) There will be significant improvements in the means of communicating with the system—the so-called man-machine interface.

The forecast development of the information technology, when coupled with the progress that we have sampled in the evolution of cybernetic system models, indicates the strong likelihood that we may be entering a new phase in our ability effectively to interface the organization with its environment. We are now in a position to address the primary concern of this paper: an assessment of the potential impacts of cybernetic organization modeling and the information technology upon management values, particularly as regards external environmental relations.

V. THE IMPACT OF ORGANIZATIONAL CYBERNETICS ON MANAGEMENT VALUE SYSTEMS

A. General Comments

In developing the concept of the organization as an open system with cybernetic loops, we have been moving toward the view that information technology and cybernetic system modeling are making it increasingly possible—therefore necessary—for managements more effectively to deal with the organization/environment interface. Herbert Simon's concept of "bounded rationality" as an explanation of management's tendency toward "satisficing" behavior remains a useful way of dealing with the uncertainties with which any decision-maker will always be plagued. But our argument is that (1) the rationality boundaries are being rapidly expanded by cybernetic technology, and (2) the vagaries in dealing with organizational ecosystems will lessen. Thus, management values, to the extent that they are still oriented to closed system—"ceteris paribus"—reasoning, are due to undergo substantial change in the new ethos.

In passing it is relevant to note that Petit indicates that the institutional manager's viewpoint is philosophical; his technique is to use opportunistic surveillance to assess environmental risks, and then to develop strategies to decrease them; his time horizon is the long run; and his decision-making strategy is judgmental. Thus his values are a vital factor in his performance. Hence our interest in him here.

B. Some Specific Impacts of Organizational Ecosystem Management

1. Increased Responsibility for Outcomes. Such developments as on-line, real-time data processing, remote retrieval systems, large-scale data banks, homomorphic modeling, industrial dynamics simulation, heuristic computer programming, and stochastic processes for dealing with uncertainty, create new possibilities for men and organizations to design their futures. Thus it will increasingly be possible to identify areas of success or failure more precisely.

2. Greater Potential for Learning. This will lead to far greater potential for individual and organizational learning than we have ever known. Performance evaluation will be a more certain art; perhaps even a "science."

3. More Need for Skill of Abstraction. Dealing with systems of ever-increasing complexity, where inter-related probability functions must be assessed, will increase the use of schematics and models. Therefore ability to deal with abstractions in interaction will increasingly be sought and valued.


Thomas Petit has recently utilized Parsons' tri-level theory of organization to identify three types of managers found in any sufficiently complex system: technical, organizational, and institutional.

We are, in this paper, primarily concerned with institutional managers, whose major task is, in Petit's words, to "cope with the uncertainty produced by uncontrollable and unpredictable elements in the firm's environment." Petit writes

Uncertainty is greatest at the firm's boundary, where the institutional level comes into contact with the environment. The closed-system logic emphasizing rationality is clearly inappropriate here because the firm is subject to the influence of external elements over which it has little or no control. The open-system logic, which permits the intrusion of environmental forces into the operation of the firm, is called for. Therefore, the institutional level specializes in coping with uncertainty. Its particular function is to avoid as much uncertainty as possible and deal with that which cannot be avoided.


889
Richard F. Ericsson
depicts an emerging condition in our society. As policy decisions are made and implemented in large-scale organizations, there will be increasing need to assess ecosystem impacts, if organizations wish to influence the decision options available to them at a later date. The essential point is that our super-corporations and conglomerates increasingly will find their previous policy implementations boomeranging upon them, to the extent that today's ecosystem is a product of the organization's past decisions. Thus the schematic also suggests the necessity for organizations to create feedback loops with their ecosystems, so that adjustments in goal realization paths will take cognizance not only of internal change but of autonomous and induced external change as well.

5. Greater Ability to Plan. Long-range planning will tend increasingly to be, in essence, a design for the organization's future. Fewer uncertainties will prevail, in the sense of the organization being blown by the "winds of fortune." Even adventitious events will be more capable of probability assessment. Serendipity will likely be less often available to rescue poor managements from bad competitive decisions.

6. More Opportunity for Holistic Integration. Rapoport has summarized^14 the argument of holistic philosophy in these words:

There are different levels of organization in the occurrence of events. You cannot explain the events of one level in terms of the events of another. For example, you cannot explain life in terms of mechanical concepts, nor society in terms of individual psychology. Analysis can only take you down the scale of organization. It cannot reveal the working of things on a higher level. . . . The proponents of holistic methods, therefore, advise scientists to try to contemplate wholes rather than parts.

Not only will the opportunity exist, but, as is indicated in Matrix II, there seems ample ground to project that computerized organizational ecosystem decision matrices will make possible higher levels of holistic integration. Algorithms or, more likely, heuristics of the projected impact of decisions upon interacting organizational participant groups will be possible. But more than this, assessment of these ramifications can be integrated with extra-organizational, inter-institutional, and even "quality of life" considerations, as the matrix suggests. (It may be noted in passing that there are presently before the Congress bills which would provide for basic data helpful in these respects, namely, Senator Walter Mondale's proposal for a President's Council of Social Advisers, and Senator Fred Harris' proposal for a National Foundation for the Social Sciences.)

7. More "Game Theory" Type Decisions. The outlook is for an increase in recognized interdependence among large-scale organizational units of all kinds in our society. Thus conditions for
### Matrix II

**A Schematic Organizational Ecosystem Decision Matrix for Management**

<table>
<thead>
<tr>
<th>AREAS OF COGNIZANT DECISION IMPACTS</th>
<th>TYPICAL DEPARTMENTATION MODES</th>
</tr>
</thead>
</table>
| **I. "QUALITY OF LIFE" AREAS**      | PRODUCTION | FINANCE | MARKETING | PERSONNEL | R & D | ACCOUNTING | ... Etc. ...
| Material (Economic)                 |            |        |           |           |       |            |       |
| Spiritual (Religious)               |            |        |           |           |       |            |       |
| Humanities (Arts)                   |            |        |           |           |       |            |       |
| Values (Philosophy)                 |            |        |           |           |       |            |       |
| Education (Schools, etc.)           |            |        |           |           |       |            |       |
| Physical Environment                |            |        |           |           |       |            |       |
| Choice-Enlarging                    |            |        |           |           |       |            |       |
| Etc.                                |            |        |           |           |       |            |       |
| **II. "INTER-INSTITUTIONAL AREAS"  |            |        |           |           |       |            |       |
| Financial                           |            |        |           |           |       |            |       |
| Elymosnary                          |            |        |           |           |       |            |       |
| Military                            |            |        |           |           |       |            |       |
| "invisible colleges"                |            |        |           |           |       |            |       |
| Schools and Universities            |            |        |           |           |       |            |       |
| Hospitals                           |            |        |           |           |       |            |       |
| Churches                            |            |        |           |           |       |            |       |
| Governmental                        |            |        |           |           |       |            |       |
| Unions                              |            |        |           |           |       |            |       |
| Etc.                                |            |        |           |           |       |            |       |
| **III. "ORGANIZATIONAL PARTICIPANTS" |      |        |           |           |       |            |       |
| (1) Technical                       |            |        |           |           |       |            |       |
| (2) Organizational Managers         |            |        |           |           |       |            |       |
| (3) Institutional                   |            |        |           |           |       |            |       |
| Employees                           |            |        |           |           |       |            |       |
| Stockholders                        |            |        |           |           |       |            |       |
| Vendors                             |            |        |           |           |       |            |       |
| Customers                           |            |        |           |           |       |            |       |
| Etc.                                |            |        |           |           |       |            |       |
| **IV. "EXTRA ORGANIZATIONAL"       |            |        |           |           |       |            |       |
| Trade Associations                  |            |        |           |           |       |            |       |
| Lobbies                             |            |        |           |           |       |            |       |
| Regional Associations               |            |        |           |           |       |            |       |
| Social Service Groups               |            |        |           |           |       |            |       |
| Etc.                                |            |        |           |           |       |            |       |

Decision strategies on the basis of rivalrous interaction will be increasingly common.

In a recent article dealing with "The Evolution of Organizational Environments," 31 Shirley Terreberry has hypothesized that organizational change is increasingly externally induced; and organizational adaptability is a function of ability to learn and to perform according to changes in the environment. Building upon Emery and Trist's identification of four ideal types of organizational environments—the "placid, randomized," the "placid, clustered," the "disturbed-reactive," and the "turbulent field" (where dynamic processes "arise from the field itself")—Terreberry argues that (1) organizational environments are increasingly turbulent; (2) organizations are increasingly less autonomous; and (3) other formal organizations are increasingly important components of organizational environments. She concludes (pp. 606-610):

The random walk which suffices in a placid-randomized environment must be replaced by stochastic processes under placid-clustered conditions, and by cybernetic processes in disturbed-reactive fields. There is some question, of course, as to whether man actually has the capacity to cope with the turbulence that he has introduced into the environment.

Obviously, it is the assumption of this paper that man is indeed moving toward that capability.

8. Management Will Operate in Increasingly Demanding Environments. The outlook for management, insofar as its own "quality of life" is concerned, is bright, in terms of challenge, opportunity for creativity and innovation, and, doubtless, rewards. But as the man/machine interface develops and we are able to build more responsive and sensitive organizations, greater levels of managerial tension and stress will also doubtless be experienced.

C. Some Managerial Value Impacts: Implicit and Explicit

Throughout the foregoing discussion there are clearly implicit value ramifications attending the impacts of organizational cybernetics upon management systems. For example, the discussion of the increased opportunities for learning and the greater need for the intellectual capacity for abstraction clearly suggests that planned educational experiences will be highly valued. But we want now to deal more explicitly with impacts upon moral values and ethical conduct.

1. Moral Sensitivity. We can project that increasingly the human quality of sensitivity to the moral aspects of organizational decision making will be sought. There will, for instance, be much greater awareness of the possible pollution of the sociosphere, and attention to minimizing the socially dysfunctional consequences of managerial action. One can anticipate that research and literature, with a focus such as is attempted here, will find increasing place on managerial bookshelves.

2. Service. A decade ago, when Forrester first published his work in industrial dynamics, he expressed the belief that here was the path to real managerial professionalism. Whatever might have been this contribution, the fact is that management is indeed an increasingly professionally-oriented activity. As Forrester predicted, both the science and the art of management have moved upward in knowledge and skill requirements.

A significant dimension of professionalism is the concept of service to others, above extrinsic reward to self. As organizations are more closely related to their socioeconomic environments, we may expect managers' performances will tend to be evaluated with service higher on the rating scales than heretofore.

3. Loyalty (Commitment). Cybernetic information technology will likely accelerate a trend toward a shift in the locus of loyalty such as we have observed in recent times. That is, managers with increasing expertise and technical knowledge are likely to be far more loyal to their professional standards and organizations than to their superiors, or even to the organizations of which they are a part. And this "extra-organizational" loyalty is likely to be quite strong indeed. The organizational cement that relates people and produces a cooperative system will likely then have to be of a different sort than the traditional fear of superior authority. An important ingredient will likely be a creation of challenge. This must have been partly what was in Morton's thinking when he expressed the view that "the manager's job is to innovate innovation."32

4. Tentativeness; Tolerance. There has been considerable speculation in the literature that the corporation of the future will be far more "free form" than we have known in the past. Intuitively, there would seem to be ample grounds for this projection. Organismic, open-system organizational models, utilizing cybernetic feedback stochastic processes and heuristics, suggest greater fluidity than does the traditional hierarchical, standard-operating-procedure, bureaucratic model. Managers will therefore value a quality of tentativeness, as the organization pursues a "balance of forces" path to its objectives. Thus more "ad hoc" and "task force" subgroups will appear, and there will be more tolerance for "corporate wild men," i.e., "far out" intellectuals.33

5. Democracy. Warren Bennis has probably gone farthest—in print, at least—in forecasting the "coming death of bureaucracy" and the "inevitability of (organizational) democracy."34 Without pushing the argument to these extremes, most of the evidence provided by cybernetic information technology does suggest that freedom of thought and action, and participation in decisions that affect one's concerns—professional or personal, will be increasingly highly regarded values.

Perhaps this is the best point at which to underscore the fallacy of the popular notion that cybernetic arrangements inevitably result in transforming humans into "cogs in the machine" or robots. Pictures of the restraints imposed by yesterday's efficiency experts in the era of scientific management are sometimes thought to pale by comparison as today's automated factories and cybernated offices are contemplated. Nothing could be farther from the reality which cybernetic control

makes possible. Cybernetic control is self-control; that is its essence. Thus the result of automated decision and control processes is that man is unchained, not chained. Norbert Wiener’s dream of "the human use of human beings" is at long last on the horizon.

6. Compassion. In a recent article, William Scott has painted a bleak picture of the failure of Industrial Humanism. Even more than failure, he argues that the techniques it has developed to further the realization of humanist values in organizations—democracy, freedom, greater individual diversity, social pluralism—have in fact been perverted to the service of La Technique (a la Elull).

I have deliberately chosen "compassion" as the value to provide a vehicle for discussion of many of these humanistic values because, it will be recalled, this was one found by England to have "low behavioral relevance" for the managers he studied. I will support the opposite prognosis. It is possible that man has unleashed, in cybernetic control systems, a Frankenstein’s monster that will enslave and ultimately destroy him. It has been persuasively argued that La Technique has a moral imperative all its own, and is already out of man’s control. Perhaps so. But not inevitably so, I think.

My case rests on the fact that cybernation is essentially different from older techniques man has invented to deal with nature. In the new man/machine interface, the machine provides the man with increasing capability of understanding his new creation and the new inter-relationships generated. Besides, there seems to be ample evidence that socio-cultural counter-forces are in fact being generated. Albeit a modest effort yet, the newly-created National Endowment for the Humanities exemplifies the hypothesis that concern will lead to action in this area. An even more striking example of the melding of management science technique and humanistic values is provided by Stafford Beer in his previously cited work. We may expect their increase. Thus I will venture that values such as compassion will increasingly be weighted in management decision processes as cybernetic technology comes of age.

7. Organizational Stability. At first thought it may seem that cybernetic technology, with its emphasis upon homeostasis as the norm, provides the answer to a manager’s prayer for organizational stability. And so it does; but this is not the whole story. Again, reference to Schematic I will show what is in mind. The concept of "heterostasis" is introduced to emphasize the notion that cybernation makes it possible for a system to undertake larger degrees of risk with greater confidence than heretofore. That is, organizations may now deliberately seek unsettling but potentially very rewarding experiences, with the assurance that the new gyroscopic control subsystems provided by organizational cybernetics will make possible system control in relatively turbulent environments. Thus stability in the old sense is likely to be less an organizational value in the future. Perhaps a new concept of "optimum instability" will really make operational the old notion of "creative conflict."

8. Rationality. There seems little evidence to contradict the common opinion that logico-deductive and empiric-inductive decision modes will increasingly dominate managerial thinking and behavior. Not that "intuition" and "creative impulse" will not have their place; but I suspect that even these will gradually submit to the rationality of operational philosophy and computer simulation. Thus human creativity is likely more and more to be in the form of effective interaction with that machine of pure logic, the computer.

On the other side of the coin is that, with more timely, better interpreted, more relevant, and harder data at hand, managements will be able to bring to participants a greater degree of professional satisfaction than has ever before been possible. Besides, cybernation will go a long way toward creating the "collegial" environment which we gradually see emerging in our more advanced corporate systems. The ability to pursue one’s interests, as he himself relates them to the corporation’s needs, will increasingly be possible. And this surely is of the essence of "self-actualization."

VI. CONCLUSION: CORPORATE VALUES AND THE ORGANIZATIONAL ECOSPHERE

It will be recalled that early in this discussion note was taken of the fact that Chris Argyris had appealed to Hartman’s axiological approach to a science of value. In these terms, Argyris has described the evolution from "axiologically not-good" to "axiologically good" organization as...
movement as follows:36

(1) From a state of affairs wherein a part (or a subset of parts) directs the organizational "core activities" (achieving the objectives, maintaining the internal system, and adapting to the environment) to the state of affairs wherein these core activities are directed through interrelationships of all the parts;

(2) From a state of awareness of the organization as a plurality of parts to a state of awareness of the organization as a pattern of parts;

(3) From a state in which the organization is unable to influence its internally oriented core activities (achieving the objectives, maintaining the internal system) to a state in which it can influence these activities as the organization desires;

(4) From a state in which the organization is ineffective in problem solving to a state in which it is optimally effective in problem solving;

(5) From a state in which the organization is unable to influence its externally oriented activities to a state in which it can influence these activities as the organization desires; and

(6) From a state in which the nature of the core activities is largely determined by the present to a state in which the present core activities are continually influenced by considerations including the past history, the present, and the anticipated future of the organization.

My perception is that cybernetic technology has the potential to increase the possibility that our corporate organizations will move in the desired directions. Moreover, if the new technology proceeds as I have indicated seems likely, then the following favorable normative value consequences posited by Argyris should also tend toward realization:

(1) the axiologically good organization is consonant with individual mental health;

(2) individual mental health is positively associated with the capacity or potentiality (not the actuality) of the individual to produce, to be committed, creative and flexible; and

(3) the existence of these factors will tend to enhance the organization's capacity to achieve its objectives, survive, and perhaps develop.

As we are primarily concerned with corporation ecosystems, we have examined some of the major potentials of cybernation to improve the corporate/societal interface, and will conclude by assessing the needs and the prospects.

Some years ago that trenchant social critic and economic philosopher Robert Heilbronner was concerned with "the seriousness of the future."38 After reviewing the extent and the depth of our social problems, he concluded

... traditional agencies will not solve our problems. They have been unable to clear our slums, discipline our enormous corporations or unions, guide our technology, stimulate our economy, or elevate or educate our people. Instead they have given us, and they will continue to give us, what we have got—a Great Paralysis.

Can this Paralysis be overcome? To do so will require more than a mere scolding, more than a few palliative measures. What is needed is a fundamental change, both of institutions and of outlook, a new balance between private prerogative and public right: in short, a reformation of that social and economic order we call American capitalism. (Underscoring supplied.)

Very recently another economist—one who just now is President of the American Economic Association—suggests that the reformation Heilbronner urged may now be at hand. He writes:39

Perhaps the most critical question of the next generation will be the relative legitimacy of private versus governmental organizations. In the last hundred years, on the whole, private organizations have been defending themselves against rising tides of legitimacy of public organizations. . . . It may be that this long tide is now turning.

Certainly the National Alliance for Business and the Urban Coalition are evidences of this. And the public words of businessmen have, to me at least, a greater ring of sincerity than was true with almost all those pious pronouncements regarding the "social responsibilities of business" of yesterday. For example:40

. . . looking ahead 25 years. . . . this country's major concern will be to improve the quality of American life. . . . I believe that the U.S. business and industrial machine is one of man's greatest creations. . . . But its future—its continued smooth running—is, in my opinion, no longer dependent on the factors that have made it what it is today; it depends, rather, on the resolution of some of the larger issues that are eating away at the heart and soul of American society. . . . and I predict that the American business community will in the decade ahead involve itself deeply with these issues. . . .

Also a recent Gallup Poll more than suggests that the corporate community will indeed have its chance insofar as the public is concerned. In August 1968 some 46% of the respondents thought that "big government" was the "biggest threat to the
country in the future." Only 12% thought "big business" had that dubious distinction.

In his recent book dealing with the Young Radicals, Kenneth Keniston indicates that one way of explaining the rejection of contemporary social values by today's activist youth lies in the fact that, in this time of unprecedented change, there is an insufficient opportunity for the usual "institutionalization of hypocrisy." That is, ordinarily there is time for value conflict to be rationalized in the social pattern; today there is not. So the visible gap, between what is said to be believed and what is practiced, is unconscionably large—for youth at least. There is glaring hypocrisy.

Thus we close with one last reference. It provides some evidence that the gap may really be closing between corporate values and social values. There may be less need to "institutionalize the hypocrisy" as was necessary because of the gulf between what corporations have professed are their "social responsibilities," and their actual behavior. Recently Henry Ford II provided Ford shareholders with the following rationalization of Ford's participation in the National Alliance for Business:

... improving the quality of society—investing in better employees and customers for tomorrow—is nothing more than another step in the evolutionary process of taking a more farsighted view of return on corporate investment. . . . Prudent and constructive company efforts to help overcome the urban crisis are demanded not only by your company's obligations as a corporate citizen, but by your management's duty to safeguard your investment.

The Fortune commentary on his words declares this to be almost a redefinition of "profits" as that term has traditionally been understood in the modern corporation. What we have been saying is that cybernetic information technology creates new possibilities effectively to implement such redefinitions of traditional values.

In a previous paper, several approaches to system problems were summarized (1). Since none of these approaches provided a patient-centered basis for hospital system planning, our research objective became one of developing a workable alternative.

The observation that the methods engineering studies had relatively little impact on the hospital led to the second stage of our thinking. It appeared that higher level management problems would have to be attacked if our efforts were to affect operations, particularly operating costs. Accordingly, an operations research study of the hospital was initiated. The objective of this study was to develop a procedure which would enable any hospital to measure the level of patient care it was providing and suggest areas for corrective action. As our work progressed, it became increasingly evident that a quantitative measure of patient care was essential for all levels of hospital planning, from bedside to community levels, wherever resource allocation decisions were made.

Initially, it was assumed that patient care, like profit in the business enterprise, could be measured and optimized to some criterion. Given the requisite measure of patient care, we planned to identify and measure the social, psychological, physiological, and physical factors contributing to it, determine their relative weights by questioning qualified experts, and develop an instrument which any hospital could then use to evaluate its own level of performance. This formulation of the problem, suggested by Gomberg's analysis of time standards, is shown below (2).

\[
\text{Patient Care} = A_1X_1 + A_2X_2 + A_3X_3 + A_4X_4
\]

where

- \(X_1\) = Sociological factors
- \(X_2\) = Psychological factors
- \(X_3\) = Physiological factors
- \(X_4\) = Physical factors, and

\(A_1 \ldots A_4\) = Respective factor weight, determined by expert opinion.

Since no single agreed upon measure of patient care could be found, and the concept of value theory as used in operations research offered little hope of providing one, two types of activities were engaged in simultaneously: (1) a series of studies was initiated by a multidisciplinary group comprised of engineers, physicians, psychologists, a nurse, and sociologists, and (2) the research team focused its attention on ways of organizing the disciplinary studies into an integrated view of the hospital.

Within each disciplinary group, research was conducted in the traditional area of concern to the discipline. Engineers, for example, applied operations research models to inventory and hospital location problems, with emphasis on the blood bank inventory (3,4). Sociologists investigated the relationship between status commitment and selected medical performance measures (5,6,7,8,9,10 & 11). The predicate calculus was utilized by a philosopher-mathematician team to develop a logical hospital model based on the assumptions of the sociologists. Psychologists studied the contribution patients could make to their own care, concentrating on the problems of information-handling between diabetic patients and their physicians (12). An investigation of the hospital labor market was initiated by an economist.

All members of the research team recognized that the kind of integrated interdisciplinary approach originally visualized was not developing, and problems of inter- and multidisciplinary research were examined (13). It was decided that, although the on-going research could be viewed as multidisciplinary, with each discipline investigating traditional problems using traditional methods, interdisciplinary research would be possible if, and only if, all disciplines focused on common dependent variables (14). An attempt was then made to select variables which would be of interest to all disciplines. It became clear that the conventional variables of interest to the disciplines represented on the study team would not provide the required common basis for interdisciplinary research, and a common dependent variable was sought which would satisfy our requirement for interdisciplinary research. Initially, there was some doubt about the existence of such common dependent variables. It was found, however, that they were provided by cybernetics (15).

--

*Reprinted by permission from Nursing Research, Vol. 12, Fall 1963. This investigation was supported in part by a Public Health Service Research Grant (GN-7797) from the Division of Nursing Resources, Public Health Service. It is part of a long range program in systems research conducted by the Systems Research Group, Engineering Experiment Station, The Ohio State University, Columbus, Ohio.

The author is indebted to Wanda McDowell, Ph.D., for her substantive contributions to this paper.
The concepts of information and control are central to cybernetics, defined by Wiener (1948) as "control and communication in the animal and the machine" (16). Cybernetics, according to Ashby "...is a 'theory of machines,' but it treats, not things but ways of behaving. It does not ask 'what is this thing?' but 'what does it do?' It is thus essentially functional and behavioristic" (17). Cybernetic concepts are familiar to the disciplines traditionally concerned with the variables which must be considered in measuring patient care: psychology and physiology, economics, engineering, sociology, and medicine (18,19,20,21,22,23).

Although an extended treatment of the concepts of cybernetics is beyond the scope of this paper, the terms regulation and control must be understood in this context. The concept of regulation is defined as the ability of a system to maintain an organism in specific survival states, and control is defined as the specification of the survival states (24).

Using these concepts, the hospital may be viewed as a goal-directed system which provides patient care by the regulation and control of patient condition defined in terms of patient states. The first requirement of a hospital system model, therefore, is to provide a means of measuring the consequences of planning and management decisions from community to bedside on the ability of the system to regulate patient condition, i.e., maintain patients in specified states. Given such a hospital system model, hospital planning and operating decisions can be based on explicit quantitative predictions of expected hospital system performance as a consequence of hospital resource utilization in different socio-economic environments, rather than on implicit, qualitative estimates.

We assume that societal goals of providing patient care are met by the performance of medical and supporting tasks. The relationship of medical to supporting tasks is indicated symbolically below. Although the use of symbols may be unfamiliar, they provide a convenient shorthand to express ideas which would be difficult to discuss otherwise. In terms of this symbology,

\[ E = F(M_1 \ldots M_m) \]

where

- \( E \) = Global goal of the hospital system in our society. The achievement of a general level of wellness (25).
- \( M_1 \ldots M_m \) = Measures of the performance of the primary hospital system mission: the practice of medicine.
- \( F \) = Societal decisions about levels of performance of \( M_1 \ldots M_m \) to meet goal \( E \).

The specification of \( E \), the goal of the health system to provide a requisite level of wellness, and the decisions regarding \( M_1 \ldots M_m \) represented by \( F \), are value judgments beyond the scope of research. That is to say, \( E \) cannot be manipulated experimentally. Decisions at this level must always be based on value judgments. Science can only assist in making these value judgments by providing organized information for decision-makers.

The performance of the primary mission of a health system, the practice of medicine, depends on the performance of a number of lower level hospital support tasks. Therefore, the next step in our analysis of the hospital system is to consider these support tasks which may be grouped into four major categories: selection, service, supply, and regulation. The relationship between tasks and missions may be represented symbolically:

\[ M_1 \ldots M_m = f_1 \ldots f_4(t_1,t_2,t_3,t_4) \]

where

- \( f_1 \ldots f_4 \) = Functional relationship between task and mission performance.
- \( t_1 \ldots t_4 \) = Measures of hospital task performance, where:
  - \( t_1 \) = Measures of performance of SELECTION subsystem.
  - \( t_2 \) = Measure of performance of SERVIC subsystem.
  - \( t_3 \) = Measure of performance of SUPPLY subsystem.
  - \( t_4 \) = Measure of performance of REGULATORY subsystem.

The selection task \( (t_1) \) is one of categorizing patients and assigning them to medical services such as medicine, surgery, obstetrics, and pediatrics.

Once a patient—by definition some one unable to maintain his own homeostatic state unaided—is selected or admitted to the hospital system, it becomes necessary to provide supporting services for the practice of medicine. The service task \( (t_2) \) includes housekeeping, administration, and other functions not directly associated with the maintenance of patient homeostasis, but necessary in support of the directly related functions of the physician and nurse.

A third major task category is supply \( (t_3) \). If the hospital as an organization is to persist in time, problems of acquisition and replacement of supplies, equipment, and personnel must be solved. Supply or logistic problems include acquisition and replacement of items such as food, drugs, and personnel. The development and maintenance of the hospital staff also poses logistics problems in that physicians, nurses, and paramedical workers must be employed and replaced if they leave the system.

Regulation \( (t_4) \) is defined as the task of minimizing the difference between actual and desired
A HOSPITAL SYSTEM MODEL

System performance, measured in terms of the system's ability to maintain patient states. Regulation is characteristic of the functioning of goal-directed systems, and is performed by a subsystem which is an information-processing analog of the over-all system. The selection, service, and supply tasks are symbolically represented in the regulatory subsystem by information about them. The regulatory subsystem provides management with information for determining how closely actual system performance approximates desired system performance and suggests action to reduce the difference between the two.

To perform the four tasks described above, several classes of resources are utilized. Resources, like tasks, can be broken down into categories for study. The major resource categories are staff, patient, plant, and equipment. The performance of system tasks at all levels depends on the characteristics of these resources and policies for using them. These relationships may be summarized as follows:

\[ t_1 \ldots t_4 = q_1 \ldots q_4(R_1, R_2, R_3) \]

\[ t_1 \ldots t_4 = \text{Measures of subsystem task performance} \]

\[ q_1 \ldots q_4 = \text{Relation of subtasks to resources} \]

\[ R_1 = \text{Physical characteristics of hospital plant and equipment, such as number of beds} \]

\[ R_2 = \text{Psychological, physiological, sociological, and economic characteristics of staff, such as their skills, attitudes, and specific competencies} \]

\[ R_3 = \text{Physiological, psychological, sociological, and economic characteristics of patients, such as their physical condition, expectations, understanding, and attitudes toward the hospital system} \]

In order to design and operate a hospital system, it is necessary to relate levels of task performance, resource characteristics, and resource utilization policy. We take the provision of a general level of wellness as the over-all hospital system goal (26). The provision of a general level of wellness may be measured in terms of the regulation of patient condition. It is then necessary to specify tasks and missions which must be performed to achieve this goal. The next step in the design and operation of a hospital system is to relate missions and tasks to resource characteristics necessary for their accomplishment. For example, if we consider the vacuum system used by the surgeon to drain an operative field, we find a resource, a piece of equipment made available by economic resources, which is performing a necessary support function for the surgeon. The ability of the hospital to support the surgeon by maintaining the patient's operative field in the desired state—that is, free of blood so that he can work—depends upon the reliable operation of this piece of equipment. Thus, we find a hospital system mission, the performance of a surgical procedure, depending on a mechanical resource, the vacuum system, which is performing a service task. Other examples of equipment which extend the capabilities of the surgical team might include measuring devices for determining patient condition such as the electrocardiograph and electroencephalograph, and equipment such as the heart-lung machine used for open heart surgery.

Similar examples could be cited for the staff. A particular procedure, such as open heart surgery, for example, requires special skills as well as equipment. A service may or may not exist in a given hospital by virtue of the fact that a staff trained in a medical specialty simply is not available.

Thus it can be seen that task and mission performance depend on the availability of resources and policies for using them. These relationships are summarized as follows:

<table>
<thead>
<tr>
<th>Area of Judgment</th>
<th>Area of Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E \sim F )</td>
<td>( (M_1 \ldots M_m) )</td>
</tr>
</tbody>
</table>

\[ E = \text{Global criterion of system performance; provision of a general level of wellness} \]

\[ F = \text{Relation of mission performance to } E \]

\[ M_1 \ldots M_m = \text{Mission performance measures} \]

\[ t_1 \ldots t_4 = f_1 \ldots f_4(t_1 \ldots t_4), \text{ where} \]

\[ f_1 \ldots f_4 \]

\[ t_1 \ldots t_4 = q_1 \ldots q_4(R_1, \ldots, R_3), \text{ where} \]

\[ q_1 \ldots q_4 = \text{Relation of task performance to resource utilization} \]

\[ R_1 \ldots R_3 = \text{Man-machine resource measures. Then} \]

\[ E = G(R_1 \ldots R_3), \text{ or the general level of wellness, depends ultimately on} \]

\[ \text{the resources allocated to the hospital system to achieve this goal.} \]

The research objective is to replace the symbols of the model with performance and resource measures and, subsequently, formal functional relationships. This means replacing the task and mission symbols \( (t_1 \ldots t_4) \) and \( (M_1 \ldots M_m) \) with quantitative task and mission performance measures; the functional symbols \( (f_1 \ldots f_4) \) and \( (q_1 \ldots q_4) \) with explicit functions, and the plant, staff, and patient symbols, \( R_1, R_2, \) and \( R_3, \) with measures of resource characteristics.

A first approximation of such a model is shown in Figure 1. This figure summarizes the mission–task–resource relationships described above. The multidimensional nature of tasks and resource is suggested by their profiles. Vertical
lines in the profiles represent measurable mission, task, and resource dimensions. The functions \((g_1 \ldots g_4)\) symbolize relationships between plant, patient, staff resources, and levels of task and mission performance.

The first step in replacing the general statement of relationships with explicit measures and functions is to define the system by "listing the variables that are to be taken into account" (27). The variables selected are then measured and related in a model using procedures of formal and empirical research (28). Thus, the system model evolves from a qualitative list of variables to quantitative statements of relationships between measured variables.

This description of the hospital system provides a general conceptualization to serve as a guide to research. In order to plan and conduct specific studies, a further breakdown is necessary. The over-all system has therefore been partitioned into community, hospital, and nurse-patient-physician triad levels as shown in Figure 2. Partitioning the system in this way provides the basis for a study of the interactions between the major segments of the system. This is important since the interaction between segments of the system may contribute more to over-all system performance than does within-segment performance.

The community provides the environment for the hospital and the hospital in turn provides the environment for the nurse-patient-physician triad. The environment in each case imposes constraints on the resources, and consequently, affects the performance of the subsystem below it. Constraints may be categorized according to the ultimate effect they have on system performance. Economic constraints, for example, limit the acquisition of staff and equipment. Limitations of space impose a physical constraint on the number of beds, operating rooms, and laboratories that a hospital can provide. At the hospital level, staff, patients, and plants are the general resource categories.
In contrast, at the triad level, we are concerned with individual patient performance.

Although our research is presently focused on the nurse-patient-physician triad, the constraints imposed on the triad by the larger hospital and community systems are of concern. As we move into the hospital and community area, it will be with the triad as the focus of our attention. As suggested in Figure 2, our ultimate objective is to relate community level resource allocation decisions to their consequences for patient regulation at the triad level.

This conceptualization of a hospital system provides a focus for research both within and between the principal segments of the system. It suggests the kinds of information needed for understanding the over-all hospital system, and provides a conceptual framework for the integration of our own research and the work of others at each of the three levels of a total health system—triad, hospital, and community.

In a subsequent paper, these concepts will be expanded, and their application to the measurement of patient care will be illustrated.

REFERENCES

3. SONNENDECKER, J. P., and MILLARD, D. W. Industrial Engineering Analysis of a Hospital Blood Laboratory. Parts I and II. (Engineering Experiment Station Bulletin No. 180) Columbus, Ohio, College of Engineering, Ohio State University, 1960.
6. EVANS, J. W. and ROGERS, L. E. Patient Characteristics and Patient Care in Ten Ohio Tuberculosis Hospitals. Columbus, Ohio, Systems Research Group, Ohio State University, 1961. (Multilith)
7. OXAAL, IVAR. Social Stratification and Personnel Turnover in the Hospital. (Engineering Experiment Station Monograph No. 3) Columbus, Ohio, College of Engineering, Ohio State University, 1960.
8. ROGERS, L. EDNA. Measurement of Status Relations in a Hospital. (Engineering Experiment Station Bulletin No. 175) Columbus, Ohio, College of Engineering, Ohio State University, 1959.
12. THRUST, R. S., and LANESE, R. R. Hospital-Patient Communication: Parts I and II. (Engineering Experiment Station Bulletin No. 182) Columbus, Ohio, College of Engineering, Ohio State University, 1960.
24. ASHBY, op. cit., p. 197.
26. Ibid.
27. ASHBY, op. cit., p. 40.
The growing size, complexity, and cost of health facilities for a growing population are matters of universal concern. They must be provided within constraints imposed by limitations of real and monetary resources. These factors impose stringent planning requirements, since the consequences of hospital system design and operating decisions should be made with respect to their effects on patient care. Because of the complexity of this problem, the model approach is indicated (1).

The problem of developing a measure of patient care is not new. Numerous attempts have been made to solve the problem by focusing on such activities as "nursing care," "hospital care," and "medical care" (2,3,4). Investigation of the ratio of hours of professional nursing care to numbers of patients, studies of patient opinion, infection and autopsy rates, and audit of medical records are examples of approaches to the problem. Nevertheless, explicit procedures for predicting the consequences of hospital design and operating decisions for patients have not resulted.

Since a patient care measure is vital for present operation and future planning, this research program has been concerned with the development of a measure which would be useful for hospital design and operation. As a result of the research, a conceptual framework has been developed, and will be described.

The principle consideration in developing our measure was that it be useful for the solution of daily operating and long range hospital planning problems. Basically, both long range and immediate problems involve trade-offs between resources as they affect hospital system performance. For example, decisions must be made regarding the purchase of new equipment, modification and expansion of the physical plant, and allocation of resources to the staff. The basic question is: What is the effect of these decisions on patients?

Given this planning requirement, a patient care measurement, based on Cannon's concept of homeostasis, was developed (5). In recent years, this concept has been extended and incorporated in a larger body of knowledge defined by Wiener as "cybernetics" or "Control and communication in the animal and the machine" (6). Cybernetic concepts have proved useful as an interdisciplinary language and are familiar in such diverse disciplines as sociology, economics, and engineering (7,8,9,10). More recently, attention has been focused on cybernetics as a way of integrating information, regulation, and control for the development of models of large, complex systems (11,12).

A consideration of patient care in cybernetic terms provides us with definitions of patients, hospitals, and patient care which are useful at hospital design and operating decisions (13). We are defining patients as members of our society who are unable to maintain desired homeostatic balance unaided with respect to a range of variables. Hospitals are organizations provided by society to maintain homeostatic states for its members, patients, who are unable to maintain required levels for themselves. Patient care, therefore, is measured in terms of the homeostatic regulatory capacity of the hospital. If homeostatic disequilibrium occurs with respect to a range of variables, the hospital must provide the resources, such as staff, equipment, and supplies to restore it. Patient care then, is a system performance measure, rather than a measure of any component.

To expand on these definitions, we think of people as organisms maintaining dynamic equilibrium in the face of continual disturbances from both external and internal sources. Their homeostatic mechanisms allow most people in most situations to adjust to these disturbances. Multi-dimensional survival states are thus maintained by the individual organism. Illness represents a departure from these states and therapy is the process of regulation and control to reduce the discrepancy between the actual state (illness) and what is possible (wellness). Regulation is the action taken to maintain a system in some specific survival state following a disturbance which tends to move the system to a nonsurvival state.

If disequilibrium occurs with respect to certain classes of variables, for example, the patient may be cared for in a mental hospital. If it occurs in other classes of variables, regulation may be achieved in a general hospital or in some other specialized type of institution. The possibilities for the control of the patient's condition, i.e., the level of patient care which can be provided by a
hospital, depends on the state of the patient and provisions for regulation provided by the hospital. These provisions depend, in turn on medical science, and the willingness of society to allocate its resources to insure a high "level of wellness" (14).

Having arrived at these definitions of patients, hospitals, and patient care, the next problem is to develop a quantitative set of relationships which will allow the prediction of the probable consequences of various resource allocation policies for the provision of patient care. In order to provide a way of relating the consequences of hospital decisions to patients, we are defining the basic subsystem of the hospital which is responsible for maintaining patient states, i.e., for providing direct patient care, as the nurse-patient-physician TRIAD. In this frame of reference, the patient may be conceptualized as a process generating signals to a monitor located in an information feedback loop, as shown in Figure 1. The monitor, observing a time-varying series of signals, may respond in one of three ways. These responses have been identified as three levels of monitoring where the levels refer to the function performed by the monitor (15,16). At the first level, information about patient condition, i.e., patient "output" information, is recorded by the monitor and/or passed directly to the comparator. The function of the comparator is to measure the difference between actual and desired patient states. The comparator transmits error difference information to the regulator who makes a decision regarding the need for corrective action. Dependent upon the characteristics of the error, the regulator specifies the action required to reduce the difference between the actual and desired state of the patient.

At the second level, the monitor assumes the comparator function and transmits difference information to the regulator. At the third level, the monitor serves as comparator and regulator, and takes necessary action to maintain patient condition within specified limits. If, for example, we consider the postoperative patient in the recovery room, desired lower and upper limits for the systolic blood pressure are specified for the individual patient. If the actual systolic blood pressure is outside these limits, i.e., a disturbance has threatened or occurred, regulatory activities to return the patient to the desired state are instigated by the regulator. The nurse may take appropriate action to regulate the blood pressure by changing the patient's position. In this situation, she is serving as a third level monitor since she detects actual blood pressure, compares it with desired, and takes appropriate action on the basis of difference information. If the blood pressure should exceed the limits within which the nurse has authority to act, responsibility for the specification of regulatory action is assumed by the physician, acting as both controller and regulator. This situation is shown in Figure 3. The physician has a wider range of regulatory strategies, and might, for example, prescribe a medication to regulate the blood pressure.

Although this illustration refers to a physiological variable which is of primary importance in the care of patients in the operating and recovery rooms, other variables must be considered here as well as in other settings. If we consider a patient on a surgical service, for example, we may find him in a number of settings as shown in Figure 3. In the nursing unit, physiological, psychological, sociological, and economic variables must all be considered. During surgery, variables of concern are primarily physiological and biochemical. The postoperative patient in the recovery room must be supported not only physiologically but also psychologically as he emerges from the anesthesia. The essential variables of concern to the triad, as well as to the hospital designer manager, will depend on the setting.

When the patient is returned to the nursing unit from the recovery room, physiological problems may decrease in importance with time, while it is likely that sociological, psychological, and economic problems will increase. In these situations, we hypothesize that good patient care would
be characterized by the ability of the hospital system to respond promptly and accurately to changes in patient condition relative to any of these variables, while poor patient care would be the reverse. These concepts complement those of progressive patient care, which are based on the provision of a series of patient facilities dependent upon the state of the patient (17).

By formulating a patient care measure in terms of the ability of the hospital system to regulate his condition when he is unable to maintain survival states unaided, we can specify the kind of information needed by hospital system designers and managers to provide the requisite facilities. This provides a patient-centered basis for introducing into the hospital many of the recent developments in electronic equipment such as computers and measuring devices.

These devices may prove to be an invaluable aid to those responsible for patient care. Much remains to be learned about their use, however, before reliance can be placed on them for maintaining patient states. The use of detection and display devices, computers, and data processing equipment should be looked upon as ways of extending the sensory motor system of physicians and nurses and not as an infringement on the personal nature of the nurse-patient-physician relationship. Such electronic devices may assist in detecting changes in patient condition so that prompt corrective action is possible. Equipment for detecting and telemetering some types of physiological data has been developed as part of the man-in-space program (18). Although this equipment and the ideas related to its use may prove useful in hospitals, research is needed to test its feasibility. We cannot assume, for example, that because a signal generated by a patient appears on a display, it will be observed by a human monitor and appropriate action taken. In fact, the experimental evidence points to contrary conclusions. Mackworth and other investigators have found a characteristic decrement in the ability of observers to detect changes in signals as a function of time (19). The relevance of these experimental findings must be determined in the hospital setting.

We must be clear, also, on the difference between instrumentation and automation—words that are being used with increasing frequency in describing hospital operations. Instrumentation, as we use the term in a patient care setting, refers to electronic equipment which detects and transmits patient information. Such equipment would include the electrocardiograph, the telethermometer, or the body function recorder. Such equipment provides information for the physician and nurse just as the mercury thermometer, the stethoscope, and the sphygmomanometer do. Automation, on the other hand, refers to the replacement of a man by a machine for the performance of a specific task. An example of the implementation of automation in the hospital may be found in the laboratory where the electronic cell counter has, in part, replaced the human in identifying and counting blood cells. While we envision patient instrumentation as a powerful aid to the regulation of patient condition by the rapid accurate processing of patient information, electronic devices cannot replace the nurse or physician in making the
DANIEL HOWLAND AND WANDA E. McDOWELL

declarations essential for patient care. It is assumed that the information which can be acquired, processed, and displayed by these devices will assist the nurse, patient, and physician in performing their critical decision-making tasks. Such information can only supplement, not replace, direct observation.

In conclusion, we have defined patients as members of our society who are unable to maintain some aspect of homeostatic balance unaided, and hospitals as institutions provided by society to perform this function for them. Patient care was defined in terms of the ability of the triad, supported by the hospital organization, to maintain patients in specified states. If the hospital system is able to provide a high degree of regulation, it is providing a high level of patient care.

These definitions and the resulting measure of patient care provide the basis for experimental research and the construction of formal hospital system models. Such models can be used by designers and managers of hospital systems to assist in the operation of existing facilities and planning for future system development. Although a basic conceptual framework for measuring patient care has been developed, much hard work lies ahead before the concepts can be explicitly stated in useful terms for hospital design and operation. As Sheps pointed out:

The development of practical and valid methods of measurement will involve the expenditure of considerable money and time. However, in view of efforts and money now being spent on programs to raise quality, it would seem essential to direct some of these sources toward the development of appropriate methods with which to judge their effects (20).

A procedure for relating resource allocation decisions to their consequences for patients is required and long overdue. The central position of a patient care measure in hospital design and operation should provide the impetus to develop the requisite procedures.

REFERENCES

2. ABDELLAH, FAYE G., and LEVINE, EUGENE. Effect of Nurse Staffing on Satisfaction with Nursing Care. (Hospital Monograph No. 4) Chicago, American Hospital Association, 1958.
13. HOWLAND, op. cit.
20. SHEPS, op. cit.
APPLICATION OF SYSTEMS THEORY TO REGIONAL CONCERNS
The management of the land resource base in the United States is in a state of need. Not enough attention has been paid to the fact that outdated modes for allocation of land uses are no longer appropriate. There seems to be a lack of concern for future demands for land, especially in the private economic sector. Prime land is being gobbled up indiscriminately, because it is the least expensive to develop. The consequences of urban sprawl are a gradual destruction of prime, biotically productive land and the elimination of choices for uses for that land in the future.

The goal or aim of the model to be described is to set up a framework for decision making which will minimize the turning of prime, biotically productive land into concrete, or into "irreversible uses." It would be difficult and expensive to turn houses and highways back into farm land or green belts if we should need them someday, so why not minimize irreversible uses of prime land in the first place by putting a priority on keeping land in a reversible state. More productive plant varieties and production techniques have bought us some time, but we cannot count on these methods to bail us out indefinitely.

Wasteful methods of allocating land are a result of the American heritage. For three centuries, settlers and land developers have had a great deal of freedom in modifying the land base for whatever use they desired. Individual land ownership has been fostered, and the right to use one's land in any way an owner wishes has been accepted as an important part of the American property system. Zoning has not been successful. Too many variances have eroded many planning attempts for cities, and suburban enclaves have been preserved for a few individuals at society's expense.

Serious land-use problems have arisen in those areas of the United States where the most highly desirable land is located. For it is in these areas where the greatest number of choices for land uses exist and where the competition for land is acute. My special concern is for a rational management of the kind of land found in the Willamette Valley in Oregon, Washington's valleys, Santa Clara Valley, and California's Central Valley. These are examples of areas which are highly desirable for agriculture but also for housing, highways, and industry, because they are accessible, level, often well-drained, and near large, growing, urban areas.

They are our "prime" lands, and they are indeed being gobbled up fast.

When a resource becomes scarce, it is no longer possible to treat it in the same way as when it is ubiquitous or plentiful. It becomes a scarce, thus precious, factor. The resource can be used for some things only, not all things; therefore, priorities for use need to be established. These will usually be based upon expected monetary returns to private enterprise or on future benefits desired for society as a whole.

Since the resource of prime, biotically productive land is becoming scarce in many urbanized regions, it is now desirable to establish priorities for land use. Long-range plans which will permit maximum flexibility for land uses in the future would favor allocating land to those uses which are the most reversible, i.e., agriculture, timber production, fiber production, and recreation; they should receive priority over irreversible uses, such as transportation systems, housing, and industry.

The task of providing a new framework for the allocation of land is so large and complicated that no one discipline provides an adequate approach for its being accomplished. Moreover, the relationships of man to the land resource base are examples of phenomena that can best be understood when they are considered holistically rather than in fragments. In order to use the talents of many specialists and to arrive at some consensus among them as to the best ways to allocate land, general systems analysis offers an appropriate scientific method to use when trying to solve this problem.

The Man/Land system can be analyzed through a careful examination of its subsystems, feedback, flows, and inter-relationships among all the variables which can be identified. Because the following systems model has been constructed as an aid in decision making in regard to the allocation of the land resource base, the concept of primary variables is important; these are the ones that have the greatest influence upon the system.

There are five subsystems or components in the present study: the economic, the political, the geometric, the land-capability (sometimes called the ecological), and the behavioral. The urban sprawl problem can be approached from the viewpoint of any one of these subsystems, but that would be only a small part of the total picture. In

---

*This paper was presented at the Institute on "Application of Systems Analysis to Viable Solutions for Land Use Problems," supported by the National Science Foundation and held at Oregon State University, Corvallis, Oregon, June 26-July 22, 1972. Dr. Hollingshead was co-Director of the Institute.
this study, each variable and each subsystem were analyzed individually before the entire model was put together.

As we discuss the components of the model, imagine that you are looking at a region on the earth; for instance, the Santa Clara or Willamette Valley, with its imaginary boundaries, and envision all at one time the physical environment, spatial limitations, the economic allocation of resources, the political institutions which regulate the allocation of power, and the people and their behavior. All are inter-related and together comprise one large system with a capacity for viability or stagnation. In Appendix I, a list of subsystem variables begins on page 92, and a plan for locating the subsystems on the model appears on page 93.

ECONOMIC SUBSYSTEM

The economic subsystem is a two-dimensional system of supply and demand, a vast network of buying and selling land. Land in the United States is regarded mainly as a factor of production. Because prime land is now scarce, its allocation is necessary. Emphasis on regional growth and population increases contribute to increasing demands for prime lands, mainly for housing, highways, and industries, but also for airports, recreation, gravel pits, waste disposal, reservoirs, etc. In addition to supply and demand, other economic variables are short-run policies, the growth syndrome, externalities, and the "highest and best use" concept.

POLITICAL SUBSYSTEM

The political system is a set of social interactions between individuals or groups. These interactions are pointed toward an authoritative allocation of values, or, one might say, the allocation of power to regulate matters in society. Again, a scarcity of valued things prevails. Some freedoms must be curtailed in order to gain others.

The political process is a vast conversion of demands into outputs. Once the outputs—appearing as statutes, directives, or opinions—become apparent, then citizens either make new demands, while supporting existing outputs for the time being, or they give their unqualified support. This is a continuous conversion process and is carried out through hearings, legislation, lobbies, public opinion, law suits, and mass media. There is constant feedback between the producers of both inputs and outputs. Citizens of a society must consider themselves bound to the outputs, that is, obey the laws and rules, or the system breaks down.

GEOMETRIC SUBSYSTEM

The geometric subsystem deals with the land space on the earth's surface, in this case, the prime lands. In urban areas, this space is three-dimensional because of high-rise buildings; in rural areas, it is two-dimensional. Heretofore, only urban space has been considered "allocational" space, because there is so much competition for it. Rural space has been considered "locational"; that is, various enterprises have located there because of suitability or convenience without having to compete for space in a highly competitive market system. One of the theses of this paper is that prime land should all be allocated, both rural and urban, but not through the sole means of the market system. It is becoming too scarce.

Geometric space is the framework within which we find other kinds of space: psychological space, health and sanitation space, empty urban space, and multiple-use space. All of these kinds of space are perceived differently by diverse groups, and this adds to the complexity of the spatial subsystem.

LAND-CAPABILITY SUBSYSTEM

The land-capability system is the ecosystem, or the physical environment, for the purposes of this study. It actually represents the boundaries of the Man/Land system. It houses all of man's activities, along with natural phenomena. Because population is increasing while the earth's land area is finite, it is becoming imperative that we learn how to use space more wisely than in the past. It is my belief that, at our present state of knowledge, the best approach to land-capability is that of "intrinsic suitability," as espoused by Ian McHarg and his associates. It involves planning man's activities for entire river basins or regions and is based upon locating the natural processes of those regions, such as aquifers, marshes, forests, or flat land. These processes are derived from an area's natural features. We are learning that flood plains, steep slopes, and aquifer recharge areas should be left alone—that man's activities are least destructive on high, flat areas. Alluvial valleys are saved for biotic production, for they cover a very small portion of the earth's surface and are among its most fertile areas.

"Carrying capacity" is another important concept of the land-capability subsystem and is concerned with numbers of persons, consumption patterns, and human values.

BEHAVIORAL SUBSYSTEM

The behavioral subsystem deals with attitudes, customs, and aesthetics. Peoples' attitudes are the "shapers" or the basic building blocks of the subsystems. Man can be a conserver or an exploiter. His goals are reflected in his attitudes.

THE TOTAL SYSTEM OR MAN/LAND SYSTEM

The total system or Man/Land system is more than a grouping of the subsystems; it is a
unique system in itself and differs from each of its components. The Man/Land system is a combination of many types of variables, not all of which are quantifiable. Quantification of those variables and flows which can be measured is desirable, especially the spatial, economic, and ecological. The other flows are measured in a qualitative manner—intuition and logic are necessary. One could say that the whole system is linked through information—the idea, perception, knowledge of information. Things influence each other. This is a form of energy, most of which is dissipated at the urban sprawl box on the Man/Land diagram. This box signifies an end result in many ways. But, of course, urban sprawl causes other things, so some flows go back into the system in the opposite direction.

The man/land study deals mainly with a system based upon institutionalism rather than on an organism. Institutionalism, which occurs in systems dealing with human beings, is an ongoing, circular, systemic process, not an open-ended chain of events with clear-cut antecedents and consequences. There is much feedback involved because of human perception; both positive feedback, which is disorganizing or structure-elaborating, and negative feedback, which is stabilizing.

The model in the large diagram here is not to be confused with the structure of its components at a given point in time, for the purpose of studying a system is not to "describe" it but to show how it works. The political and land-capability subsystems are shown as they should be in order that land waste be minimized; the behavioral and economic systems are pictured as they exist right now. The flows between them are the important thing. Thus, a change in any variable will cause a series of changes in the system, for the energy flows will be affected.

In the list of variables (Appendix I) and on the model, the primary variables responsible for land waste appear in the economic and behavioral subsystems: affluence and wasteful use, a growth syndrome, a highest-and-best-use approach, short-run policies, increasing population (natural and immigration), demands for housing and highways, automobile takeover, detachment, old customs, a dislike for cities, and apathy. They are so dominant that the system cannot be changed until some of them change.

The political system has the potential power to regulate the framework within which the economic system may operate so that it will no longer be exploitative. For example, it can set up regional criteria in governing and planning in order to implement the goal to reduce urban sprawl and thus require business to operate within a spatial framework which controls land use. At the present time, however, the political system often favors local interests and special-interest groups.

The political system cannot change by itself or regulate the behavioral system, for it is itself controlled by the behavioral system. Since both the political and economic systems are mainly controlled by the behavioral system, this latter system is the one wherein changes must occur before the conservation of land can replace land exploitation.

Changes in attitudes concerning land can be brought about through knowledge of the entire man/land system, but especially of the land-capability and spatial subsystems, which have been largely ignored in America until right now. Quantitative and qualitative space are the environment of man and his institutions. Man's attitudes must be relevant to the physical and spatial constraints which they impose. Attitudes and customs in keeping with the view that man is part of nature are necessary.

It can be intuited that there exists a hierarchy of subsystems within the Man/Land system—a priority listing—if the goal of man to conserve his biotically productive land is to be achieved. Finite space and land-capability impose a set of natural boundaries upon man and occupy first place in the hierarchy. Knowledge of man's physical and spatial world affects the behavioral subsystem, which occupies the second place in the hierarchy. Attitudes and customs which reflect that man is a part of nature rather than dominant over nature influence the political subsystem, third in the hierarchy. It, in turn, sets the limits within which the economic system can operate, a structure based upon conservation rather than exploitation. The entire system is an ecosystem, for although there is a hierarchy of priorities for emphasis in decision-making, all of the variables are interrelated. The new hierarchy is a radical departure from American priorities at the present time.

RECOMMENDATIONS

The major control for organizing the Man/Land system is the decision-making structure composed of those persons with the authority for deciding upon land uses. Decisions are based upon the goals desired and upon the information available to the decision makers. Innovation is being limited by the inertia and momentum of past allocative judgements, for the current distribution of resources is the result of an historical process which is difficult to change.

A decision-making hierarchy should be established in which the interests of higher levels of governments have priority over those of the local levels. Nations and states traditionally favor policies that will bring benefits to the greatest number of their citizens; these levels of government are usually more democratic than lower levels, wherein vested-interest groups and private citizens sometimes exploit resources for their own gain at the expense of society. Knowledge concerning the
hierarchy of subsystems involved with land is of great importance at all levels, especially the concepts of "intrinsic suitability" and "carrying capacity," in order that decision makers can act wisely.

Since human behavior is unpredictable, it would be impossible to plan a perfect system for land conservation which is based upon logic alone. Innovations should be encouraged which are based both upon logic and intuition. Experiments in new kinds of cities, transportation, recreation sites, and industrial communities are necessary in the effort to conserve land. Diversity is desirable. Some things will work; some will not work. A great deal of thought must be given, however, to implementing some policies which we know will preserve land, for our system is now so negative that it is not only stagnant, it is deteriorating rapidly. Many specific suggestions for conserving land will be found in the Appendix, both for the long run and for the short run. Special emphasis is needed to change those variables which are the primary ones responsible for urban sprawl.

The underlying intentions of the model and the recommendations can be summarized as follows:
1. To provide a framework for keeping options open for as many choices as possible;
2. to provide for incremental decision-making and innovation within a large and long-term framework;
3. to provide for sound and rational land uses which are ecologically, economically, and culturally sound and manageable, and
4. to encourage both the individual and the citizenry to innovate and achieve.

All of the above discussion and the model are based upon the premise that it is wise to conserve our prime, biotically productive land. I sincerely believe that it is a desirable goal and that this model provides a framework for its achievement.

APPENDIX I

MAN/LAND SYSTEM (Model Variables)

+ Primary tools for minimizing urban sprawl
- Primary variables leading to land waste (urban sprawl)

A. Economic Variables (economic subsystem)
1. Demands for land
   a. Housing
   b. Industrial space
   c. Highways
   d. Recreation
   e. Biotic production
   f. Additional demands
2. "Highest and best use" approach
3. The growth syndrome
4. Short-run policies
5. Externalities

B. Political Variables (political subsystem)
1. Demands and support of the land-use power structure
   a. Citizen groups
   b. Jealous bureaucracy
   c. Large landowners
   d. Overlapping governments
      1. Municipal
      2. Metropolitan or regional
      3. State
      4. Federal
   +2. Political tools for guiding land development
      a. Zoning
      b. Tax policies
      c. Use rights
      d. Land purchases
      e. Eminent domain

C. Quantitative Spatial Variables (geometric subsystem)
1. Finite space—Euclidian
2. Psychological space
3. Health and sanitation space
4. Empty urban space
5. Multiple-use space

D. Qualitative Spatial Variables (land-capability system)
1. The ecosystem
   a. Functions
   b. Damaging factors
2. Ecological Planning
   +a. The ecological approach
   +b. An ecological value system
3. Carrying capacity
   a. Numbers of persons
   b. Consumption patterns
   +c. Human values
4. Problems of non-compatible land uses
   a. Agriculture
   b. Air corridors
   c. Litigation over land uses

E. Behavioral Variables
1. Attitudes (overt)
   -a. A wasteful use syndrome
   -b. The concept of freedom—individualism versus the group
   -c. Detachment—"It is fine for others, but not for me."
   -d. Reservation—clinging to space by a select group
   -e. Apathy
2. Customs
   -a. Customary space, i.e., backyards in suburbs or front yards in cities
A SYSTEMS ANALYSIS MODEL FOR MINIMIZING URBAN SPRAWL

- b. Customary economic, political, and personal decision-making
- c. Customary educational systems

3. Aesthetics (sensuous perception)
   a. Beauty
   b. Tranquility
   c. Solitude
   d. Stimulation

Table 1.
PRIORITIES OF SUBSYSTEMS WITHIN THE MAN/LAND SYSTEM

1. Finite space and land-capability subsystems.
   Impose boundaries for man's activities. Intrinsic suitability and carrying capacity important aspects. Based upon data of the natural features and their natural processes, along with spatial measurements.

2. Behavioral subsystem.
   A land ethic based upon knowledge of the ecosystem. A realization that man is part of nature. Attitudes and customs leading toward the protection of and improvement of land rather than exploitation.

3. Political subsystem.
   The tools for allocating the power to decide upon land uses. Such tools as statutes and zoning on a regional basis are made possible by the attitudes of persons responsible for inputs and outputs of political systems.

4. Economic subsystem.
   A market allocation system that fits into the framework prescribed by above systems.

APPENDIX II

SOME SPECIFIC RECOMMENDATIONS TO BE CONSIDERED FOR POSSIBLE CHANGES IN THE MAN/LAND SYSTEM—LISTED BY SUBSYSTEMS

Economic Subsystem

1. Land prices could be controlled like those of utilities. Reasonable prices for land to be set. Perhaps land is too much in the public interest to be allowed to remain in private ownership with uncontrolled prices.

2. Air and water pollution costs should be incorporated into the price system (external diseconomies internalized) so as to improve the quality of land.

3. Economic incentives to develop land for flexible uses. Make it costly to use land for irreversible uses.

4. Economic theories must not be so concerned with the idea of "scarcity." Quality is also important.

5. The United States Government could help to disperse people through economic enticements to locate in new urban areas, or government contracts could be awarded to those existing urban areas which can best tolerate an increasing population.

6. Apartment living should be made more attractive. There should be many kinds of apartment shapes and sizes and also a variety of recreational activities. There could be rewards for people who live in apartments, i.e., a lower cost of rent or purchase per room than in houses. Thus single-family dwellings would be considered a luxury and people would have to pay more for the privilege of living in them because they occupy more land.

7. "Automobile takeover" must be examined by the American people before we become completely dominated by this machine. Other forms of transportation should be encouraged to substitute for the automobile in congested areas, i.e., subways, monorails, jitneys, etc.

8. "Not-for-profit" community development corporations, such as state land development corporations, that can contract with profit-making businesses to build new communities that meet certain performance standards.
New construction methods, i.e., "modular units," may make possible construction of high-rise buildings at costs more nearly in line with present methods used for low buildings.

**Political Subsystem**

1. Federal policies on land conservation must precede state and local policies.

2. Hierarchical planning and decision making—the higher levels having priority over the lower ones.

3. Long-term planning and constant reappraisal of all the variables involved. Information from data banks will be important.

4. Recognition that suburbs can be "reservationist" rather than conservationist and a bottleneck to orderly growth.

5. State Agricultural Land Commissions should work with all other land development agencies (highways, dams, etc.) to make a case for preserving the best farm lands.

6. Principles of law, such as equal rights and equal responsibilities, applied to regional matters in preference to local ones.

7. Representatives to watch out for regional interests in lawsuits involving only local people.

8. State Land Boards to buy land on the urban fringe and hold it for future development. Land not to be released until the urban lands are fully developed.

9. Government purchases of land use rights for the general welfare, i.e., for recreation, scenery, historical sites.

10. Buy development rights from farmers for urban use. Where it is possible for governmental units at various levels to purchase the rights to convert farms to urban uses, this can be done to control the rate and manner in which land passes from rural to urban uses. Such a scheme for controlling land use has been seldom tried.

11. Legislation giving agriculture priority over other uses in certain areas. Animal densities not to be limited; large-scale spraying and fertilizing, noisy machinery, etc., allowed. This would protect the most suitable farm lands for modern mechanical agriculture from other non-compatible activities. Long-range planning required at the state level.

12. Rigid government control over freeway building and other large public works projects in order to conserve valuable lands.

13. A reappraisal of the federal right of eminent domain.

14. Statutes and policies to emphasize incentives for conservation rather than more police action.

15. School children should be taught about the practical political matters of hearings, petitions, letters of protest or support to legislators, the roles of committee chairmen and lobbyists, and how bills are brought to the floor for a vote.


**Geometric Subsystem**

1. Minimum amounts of land to be used for public needs—power lines, highways, parking, schools, municipal buildings.

2. Non-compatible uses separated. Priority-use zoning. Or require industry to be so clean that it does not need to be separated from other activities. Mechanized farm lands separate because of sprays, odors, noise, etc.

3. Multiple-uses for land wherever ecologically and economically feasible.

4. Minimize freeways. They tend to modify the areas through which they pass and require a great deal of space for themselves and for the activities which they generate.

5. New areas to be planned for maximum carrying capacity but with lots of open spaces between developed clusters and within them.

6. Open spaces for variety. Man needs them in urban areas for his psychological well-being.

7. More attractive and stimulating urban areas in order to limit urban sprawl (the flight to the suburbs).

   a. Small parks and open spaces greatly increased.
   b. Bicycle and walking paths.
   c. Balconies on apartment houses.
   d. Sidewalk cafes.
   e. Few or no automobiles in cities ( jitneys, rapid transit, moving sidewalks, etc.).
   f. Fountains, trees, and flowers in profusion.
   g. Underground wires.
   h. Small signs for shops—no billboards.

8. Long-range planning and education are necessary for the promotion of other types of dwellings than the traditional single-family home on a small plot of land, which is already so prevalent.

9. A change in suburban zoning laws that discriminate against multiple dwellings. If the empty spaces in many suburbs could be filled with apartments or other types of multiple-dwellings, the developer would not have to range so widely over the rural landscape.

10. There should be more experimentation in city living in an effort to make it more desirable. Perhaps some houses should be built in rows right next to a sidewalk and contain inside courts which would be quiet, because they are away from traffic noise, such as is done in many Latin countries. Apartments could provide sports facilities (pools, saunas, indoor tennis, etc.) on a much larger scale.
to give their residents a feeling of the outdoors and freedom even though everything is inside.

11. Public buildings, such as high schools, colleges, athletic stadiums, concert halls, etc., should all be located wherever possible near good public transportation to prevent highway crowding.

12. An alternative to providing miles and miles of urban freeways is to do nothing at all. Either the economic activity or the people would shift positions when the peak loads became unbearable, i.e., the spatial distribution of activities would change out of necessity. As it is, we are building freeways fast enough to handle urban traffic, but in the process we are building great concrete structures that are completely taking over the landscape in our urban areas.

13. Rapid transit planning that provides for a grid system for moving people and goods in an area rather than radially from a city center to a few places on the outskirts. It would encourage development everywhere in the city, not just on main arteries, and would help to prevent urban sprawl.

Land-capability Subsystem

1. Laws which limit the use of septic tanks except for farmers, and very low density housing, such as in forests. Developers would not be permitted to go out very far from the edge of a city because they would have to be hooked up to an existing sewer line.

2. Control of water and air pollution through the means of placing the responsibility for the costs of control upon the polluter and getting these costs into the price system.

3. The recycling of solid wastes. The recycling costs (mainly transportation) would be part of the original price of a good. Land would not be necessary for waste disposal to such a degree as it is now, and it would not have ugliness imposed upon it by junk automobiles, cans, non-returnable bottles, etc.

4. Require services before development, i.e., electricity, streets, water, and sewer. Approval of state funds for sewer or water districts, local road construction, etc. should hinge on how well local proposals compare with over-all planning objectives for the state. Pre-servicing could also be encouraged through making loans available for these purposes. The important point is to arrange the services ahead of demand in order to guide urban development.

5. Provide more guidance to utility districts.

Replacing unguided utility districts which are created after-the-fact with better regulated and financed ones would apply to the present-day semi-planned or unplanned areas which are growing in a haphazard way. A better solution is the first-mentioned one—pre-servicing, but it involves state-wide, long-range planning which takes time to implement.

6. Data banks with many kinds of information on the physical environment—continually updated.

7. The study of air corridors in regions to be developed so that externalities from air pollution will be minimized.

8. "Carrying capacities" to be studied in regard to human communities. Physical and psychological factors to be considered.

9. Forests surrounding cities wherever possible (or farmlands) to temporize summer climates, counteract polluted city air, and filter noises and smells of civilization.

Behavioral Subsystem

In order that attitudes, customs, and aesthetic values can be changed, all types of educational institutions should emphasize the following ideas:

1. Logic rather than empiricism. There is not always enough time any more for trial and error based upon feedback.

2. Entire processes and relationships.

3. Man's place in nature and as a link in the ecosystem.

4. Long-term planning and means to achieve goals.

5. Creative individualism—a freedom of the spirit.

6. An emphasis on society's welfare. The individual will reap the benefits of a more stable society.

7. Attitudes that will promote values that promote a healthy, more equitable economic system rather than those which promote economic gain for privileged interest groups.

8. Teaching youngsters to desire a more livable environment and to appreciate beauty.

9. That a planner must be an educator as well as a practitioner. Ideas, thoughts, and concepts not heretofore part of the public consciousness must be brought to light.

10. That diversity and experimentation are desirable.
Fig. 3. A Dynamic Response Model of the Land-use Political System

Fig. 4. The Political System (Detail).
Desire for Privacy
Desire for Own Home
Demand for Industrial Space
Demand for Recreation Land

Fig. 5. Economic Subsystem.

A SYSTEMS ANALYSIS MODEL FOR MINIMIZING URBAN SPRAWL
Any metropolis can be thought of as a huge engine of communication, a device to enlarge the range and reduce the cost of individual and social choices. In the familiar telephone switchboard, the choices consist of many different lines. Plugging in the wires to connect any two lines is an act of commitment, since it implies foregoing the making of other connections. The concentration of available outlets on the switchboard permits a wider range of alternative choices than would prevail under any more dispersed arrangement. It also imposes less stringent conditions of compatibility. The limits of the potentially useful size of a switchboard are fixed by the capacity of the type of switching and control equipment available.

The facilities of the metropolis for transport and communication are the equivalent of the switchboard. The units of commitment are not necessarily telephone calls but more often face-to-face meetings and transactions. For any participant to enter into any one transaction usually will exclude other transactions. Every transaction thus implies a commitment. The facilities available for making choices and commitments will then limit the useful size of a metropolis.

Contact Choices: The Product of Cities

From this perspective, the performance of a metropolis could be measured in terms of the average number of contact choices which it offers to its inhabitants within, say, one hour of round-trip commuting time, at the prevailing levels of effort and equipment. Efficiency in cities, as in other organizations, differs from effectiveness. Effectiveness is the probability of carrying out a given type of performance, regardless of cost, while efficiency consists in low cost for a given performance. The more persons or services available to a city dweller within a round trip of one hour, the more effective would be his city or metropolitan area, and the cheaper the cost of maintaining a metropolis that places, say 1,000,000 people and 50,000 public and private institutions, firms or service points within a given commuting radius, the more efficient the metropolis could be said to be. The effectiveness of a metropolis could be measured in contact choices within one hour of travel time, while the efficiency of the same city would be measured by the ratio of such choices to some unit of cost. How many choices with $100 per capita buy for the residents of city X? As in many problems of design, one criterion cannot entirely override another. Some increases in effectiveness may have to be sought even at the price of rising costs, and some gains in efficiency may be worth some concessions in performance.

According to this view, the essential performance of the metropolis is in the enhancement of the range and number of such choices, and the basic cost is the maintenance of a system of facilities that makes a wide range of choices possible. One might ask: how many choices can an individual buy at a cost he can afford—and how many such choices on the average can the community buy for different groups of people, at prices it can afford? For each type of city and for each type of communication and transport system, it might then be possible to sketch demand and cost curves based either on the best available knowledge, or on prevailing practice.

Large cities, of course, serve many other functions. They offer playgrounds for children, lanes for lovers, shelter for residents and transients. But houses, playgrounds, and lovers’ lanes are found in villages as well, and so sometimes are factories, power stations, mills, inns, manor houses, and castles. Almost any one kind of installation found in a metropolis can also be found in the countryside. It is the multiplicity of different facilities and of persons, and the wide choice of potential quick contacts among them, that makes the metropolis what it is. And this essential character applies to large cities in underdeveloped as well as in advanced countries.

This general function of the metropolis is facilitated by its geographic location at some nodal point in a larger transportation network. The more the arteries that intersect at the site of the city, the greater the opportunities the city has to facilitate a wide range of choice. Again, the larger the city, the more diversified its industries, repair shops, and service installations—hospitals, research institutes, libraries, and labor exchanges—the wider the range of possible choices among

*Reprinted by permission from Daedalus (The Journal of the American Academy of Arts and Sciences), Winter 1961 and from THE FUTURE METROPOLIS, edited by Lloyd Rodwin, published by George Braziller, Inc.

1 This would be analogous to measuring the performance of a switchboard or of a central telephone exchange in terms of the number of potential calls among which an average subscriber might be able to choose, say, for ten cents, or within thirty seconds time for dialing, automatic switching, signaling, and the first response of the called party.
The larger, the more diversified, the more highly skilled and educated the population, the greater the range of available personal choice either with respect to organizations or to opportunities in the world of culture, recreation, and the arts.

In terms of economics, particularly in regard to the location of industrial enterprises, many of these considerations appear as external economies, actual or expected. Roads, port and rail connections, municipal services, the supply of skilled and unskilled labor, and the availability of high-level professional and scientific talent—all appear as so many potential factors of production, and some of them may even appear as free goods, against which no additional items of cost need to be budgeted. As will be evident later, the expectation may not be an altogether realistic one: the effective attractions of the area for new industries may lead after a time lag of some years or decades to substantial problems of congestion and overload. Yet locating in or near a great city is not only an exercise in economic rationality. Often the decision is made in intuitive and human terms; and most often perhaps the economic reasoning and the human preferences for location may seem to reinforce one another. Both tend to seek a widening range of choice at low, or at least tolerable, costs of choosing; and just this is the special advantage of the city. The rising proportions of industrial staff whose jobs are oriented to communication, service or professional functions may reinforce this attraction.

The power of the metropolis as an engine of communication is thus attested indirectly by its power of attraction over people. Though this power has an economic component, in the aggregate it is far more than economic. "How ya gonna keep 'em down on the farm, after they've seen Paree?" asked an old song; and the sociologists and anthropologists of the 1940's and 1950's have been reporting the vast attraction of urban areas in Asia and Africa to former villagers, far beyond any immediate economic or social push. They are held even in the squalor of the shanty towns and bidonvilles. If freedom is the opportunity to choose, then the metropolis, in so far as it is an engine for facilitating choice, is also one of liberation. This liberation may be physical, in terms of the visits, the meetings, the sights now possible, or psychological and vicarious, in terms of the choices and experiences which can now be made in the imagination. In either case, it is a liberation whose reality and whose social, political, and psychological relevance cannot be doubted.

Communication Overload: The Disease of Cities

People come to large cities because there, among other reasons, they find a wider range of choice within their individual limitations than they are likely to find anywhere else. Inevitably this means that every metropolis must offer each of its residents enough freedom for a wide range of choices to be significant to them; and this also means enough freedom so that serious problems of peak loads and of recurrent, possibly growing, overloads are imposed on the city's many but limited facilities. Recurrent overloads are thus not an alien disturbance intruding into the even functioning of the metropolis. They are, on the contrary, an ever possible result of the essential nature of the metropolis as a device for facilitating a wider range of free choices.

To put it differently, the likelihood of such overloads is a result of the probability of coincidences in human choices and behavior under conditions of freedom. These overloads are not only the occasional loads, for which reserve capacities must be provided, but also the regular rush-hour loads, the result of relatively synchronized hours for work and recreation which in turn permit a larger range of choices than staggered hours would. Despite their origin, however, recurrent overloads will tend to paralyze many functions, and eventually to blight the very structure of a metropolis. It is for good reason that waiting-line theory has become a fast growing field in operations research and social science. Taken together, increasing overloads of this kind reduce or destroy many attractions of the metropolis as well as the economic value of many of the capital investments in it.

Even in the absence of such overloads, the very effectiveness of a metropolis may produce subtle changes in its culture and in the cast of mind of its residents. A wider range of relevant choices implies ordinarily an additional burden on those who are choosing. Some years ago, Clifton Fadiman wrote a thoughtful article "The Decline of Attention" in modern, and especially American culture.2 Since then Richard Meier has written of "attention overload" and of the "communication-saturated" society as characteristic problems of modern—and thus particularly of urban and metropolitan culture.3 These, too, are overloads in communication, but they occur not in streets and telephone lines but within the minds and nervous systems of people.

To increase the range of visible and relevant choices that confront a person usually means to increase the opportunity cost of whatever course he may eventually choose. Whatever he does will necessarily imply foregoing something else that also has appeared relevant and in a sense attractive. The wider the range of relevant choices we put before a person of limited physical and psychic resources and capabilities, the more acute and

3. I am indebted for the term to Richard L. Meier; see his "Characteristics of the New Urbanization" (multigraphed, University of Chicago, 1953), especially pp. 3-5.
pressing we make his problem of economy in allocating his own time, attention and resources; and if he has been raised in a "conscientious" culture, such as the American or Northwest European, we are quite likely also to have increased his vague but nagging sense of self-doubt and misgiving as to whether he has made the best choice, and thus the best use, of his opportunities.

Cities therefore may produce a pervasive condition of communications overload. Whereas villagers thirst for gossip, city dwellers with more ample choices may crave privacy. But the internal communications overload of other people makes them less receptive to our needs. Their limited attention or their real need for privacy may tend to exclude us, and in the midst of crowds of neighbors we may experience persistent loneliness. Such loneliness, inflicted on us by others, is the obverse of our own need for privacy; and our own limited capabilities for concentration, attention, and responsiveness will make both their and our loneliness less likely to be overcome.

What people cannot overcome, they may try to gloss over. The poets and the social scientists—both critics of our culture—have catalogued the many rituals of self-deception that men practice: the reading of mass media that purvey illusions of "inside information" to the millions commonly excluded from it; the fancy dress of conformity which they don, from ivy league dress to the black leather jackets of youth gangs, or beatnik beards and sandals. Even these foibles that convey a sense of belonging, of identity, should be seen in perspective. People indulge in them, not necessarily because they are more shallow or stupid than their forebears in a village or small town, but because the commitments the metropolis imposes—of greater freedom, wider choices, greater burdens on their attentions and their powers of response—have temporarily become too much for them.

This temporary overburdening may be particularly acute for newcomers from some radically different cultural background. Then the effects of psychological uprooting through contact with the wider opportunities of metropolitan life are superimposed on the effects of the shock of a new culture and the weakening of the traditional bonds of family and familiar authority.

Communication overloads may be reduced through effective cues for orientation. Consider, for example, the practice of the old city builders, who placed the most important structures of visual attraction, such as cathedrals, palaces, or monuments, at the nodal points in the street network of the city. The nodal points, as the term is used here, were those located at the main intersections of the city's traffic flow, and hence most often observed as the city's landmarks, and they were also those points most useful for orientation. The experience of visual beauty in a place of visual usefulness was thus often an inevitable part of a city dweller's daily coming and going. It is perhaps not too fanciful to surmise that this combined experience of perception and clarity of orientation in such cities as London, Paris, Bern, Cologne or Prague contributed, and still contributes to the charm of those cities and to the feeling of their inhabitants that they were members of a deep and rich culture. Bridges can fulfill a similar orienting function: the San Francisco Bay Bridge and the Golden Gate Bridge come readily to mind, together with the Embarcadero Tower of the old ferry building, visible for a long distance along the major artery of Market Street.

In many modern metropolitan areas, however, these conditions are no longer fulfilled. Major intersections in many American cities are often adorned with gasoline stations or car lots, with flimsy, low, shop buildings with large neon signs. At the same time, many of the largest, most expensive, and sometimes most impressive constructions are put on side lots, well away from the main intersections, as for example Rockefeller Center, the Lever and Seagram buildings, the United Nations building or the Museum of Modern Art in New York City. In Boston, the John Hancock building, tallest and most monumental in the city, is tucked away on a side street. Many of our visible landmarks are only of very limited help in orientation, and are best seen from afar or by special visit. At the same time, many of the major intersections passed daily by most of us are either nondescript or appallingly ugly and give subtle but depressing impressions of disorientation tiredness, or tension.

Such crucial traffic points cannot be easily abandoned. When elegant entertainment and shopping shifted from the central intersection of Times Square to the area of Rockefeller Center, the old subway system became less convenient to users of the Center, who have to make their way there and back by foot, bus, or taxi and who thus have increased congestion. This contrast between the changing fashions in regard to neighborhoods and the unchanging nature of fixed intersections in a major traffic network helps to make the market mechanism such an unsatisfactory instrument for the development of these crucial sites.

Overloads on some of the public and private
services available in an urban area may sometimes be reinforced in their effects by a shrinking, or even an atrophy, of these services. Services vulnerable to this kind of process include service and repair shops, stockrooms, and parts depots, hospitals and clinics, libraries, and museums. Many of the services of these institutions might be needed on Saturdays, in some cases, as is true of the cultural institutions, or on Sundays, or often for many hours on each work day. Institutions such as supermarkets and suburban shopping centers provide such longer hours of service, but many others do not, and some now curtail the amount of service previously offered. Much of this situation seems caused by rising labor costs, by fixed budgets, by the rising cost of able managers for small or middle-sized undertakings, or by the difficulty of dividing units of managerial effort so as to obtain management for some extra hours daily or weekly, and perhaps by some subtle development in American metropolitan areas that makes the personnel in service industries prefer shorter hours to more pay. This may be a rational choice, but it may become less so if too much of the new leisure is frittered away in waiting for delayed services. An increase in staff, with additional compensation for staggered hours (already practiced in suburban supermarkets), might be one approach an affluent society could well explore. In any case, free-market forces alone seem unlikely to overcome the persistant gap between the rising need for services in metropolitan areas and the actual volume of services rendered.

Suburbs: An Escape from Overloads

In the congested metropolis, a major effect of the cumulative overloads on communications, transport, and other urban amenities is frustration. Withdrawal to a suburb offers partial surcease. Taxes play a role in these frustrations. The late Justice Holmes once said he did not mind paying taxes, for this was the way he bought civilization; but exasperated city dwellers may flee to the suburbs from a metropolis where so much tax money buys so little in civilized living. The remedy is not to lower the urban tax cost as such but to improve the quality of metropolitan government and metropolitan living by attacking the whole range of overloads. Several lines for such an attack have been proposed, but most are proposals for escape. When put into practice they have not been markedly successful. For example, the shift in population to dormitory suburbs around the old cities has produced mounting burdens on commuting. A farther shift to some twenty-five or fifty miles from the city would make commuting prohibitive for many; whereas some men have been able to afford the financial and physical costs, their wives have found themselves marooned in a more or less rural environment, deprived of most of the choices and opportunities that make city life attractive.

The schemes for satellite towns are more far-reaching: each would be near the city, with separate though limited facilities for employment, shopping, services, and entertainment. Some towns of this kind have been built, but in Britain, at least, they have proved less popular than expected. Still more far-reaching schemes for decentralization would break up the large cities altogether in favor of a wide scattering of major factories and administrative offices over much greater regions. Such a proposal would require a heavy reliance on medium- and long-distance transport and on telecommunications, as well as the acceptance of rural (or nearly rural) isolation.

All these schemes are unsatisfactory in the same fundamental respect. For escape from the frustrations of the metropolis, they would sacrifice the primary purpose of the large city—a wider range of choice with a low cost. The search for more effective ways of dealing with urban problems cannot ignore this basic function of a metropolis; it must rather be the starting point.

A Strategy of Search for Solutions

The concept of a metropolis as a device for facilitating choice in communications can contribute first of all to answering some general questions, from which one may proceed to more specific surmises and to ways in which both the tentative general answers and the specific surmises can be tested. The first questions might be these: what is the usual ratio of the cost of transport and communication facilities to the cost of shelter? how does this ratio change for different types of cities? how is it influenced by an increase in the scale of a city, as measured by its total population?

There are several ways of exploring this inquiry; they should give us interchangeable indices of the same underlying fact. The proportion of communication costs to shelter costs could be measured in terms of the ratio of total capital investment in communication and transport facilities to the total capital investment in shelter. Or it could be measured in terms of the ratio of current expenditure of communication and transport to current expenditure on shelter; or in terms of the ratio of total manpower employed in communication and transport to the manpower employed in the construction and maintenance of shelter. Doubtless, a range of further indicators of this kind for

7. See Paul F. Lazarsfeld's discussion on the interchangeability of indices in his article "Evidence and Inference in Social Research." Daedalus, 1958, no. 4, pp. 99-130 (On Evidence and Inference).
related ratios might be developed, but those already given should serve amply to illustrate the point.

One could also study the ratios of some appropriate nonmonetary indicators, such as the physical proportions of certain relevant facilities. The known ratio of the area of land that is devoted to streets in a city to the land area devoted to dwellings and gardens could perhaps be used more effectively within the context of the other ratios noted above.

Other types of large-scale organization could be studied. As taller skyscrapers are built, what is the change in the ratio of space devoted to elevator shafts to the total volume of the building? As corporations grow bigger, what is the ratio of telephone calls and written messages to some measure of the total volume of company activities? Such questions are aimed not merely at promoting speculation but also at suggesting a principle that may be tested: as the size and functions of a city grow, the proportion of resources devoted to transport and communications may have to grow faster, or at least as fast, if increasing overloads are to be avoided. It may be that some lag may produce no ill effects. Then the crucial question would be: what lag in the growth of such facilities is acceptable? Research may disclose a range of acceptable or desirable proportions for investment in such facilities and thus offer a potentially useful tool to planners.

What would life be like in an otherwise normal metropolis if its transport and communication facilities had been deliberately somewhat overdeveloped by present-day standards? Suppose its streets and intersections were hardly ever jammed, its parking spaces rarely unavailable, its public transport frequent, rapid, clean, and uncrowded, its telephone lines usually free, with quickly available connections? If this sounds too much like utopia, it might still be asked: how much improvement in well-being in a city could be purchased by how large an investment in drastically improved transport and communication?

Some years ago, Sigfried Giedion drew attention to the late nineteenth-century shift to the pavilion system in large exhibitions, away from the earlier practice of centralizing all major exhibits in one giant building of the Crystal Palace type. Giedion suggested that, as the exhibitions and the crowds of visitors grew larger, they gave rise to intolerable demands for more corridors to keep the crowds moving. The solution devised was to break up the exhibition into scattered pavilions, and to let the visitors make their way from one to another across the network of footpaths or across the open ground. People preferred to walk hundreds of yards along some crowded corridor or hall. The principle may be relevant perhaps to the metropolis and the problems of urban decentralization.

Again, the question of cost arises. The shift to the pavilion system made the visitors themselves responsible for keeping dry and warm, a cost previously borne by the management of the single central hall. When a shift occurs from a compact city to a spreading network of suburbs, costs are also shifted from the city government to suburban families, who must now maintain one or two cars and pay toll rates for most of their telephone calls. In addition, there is now the financial, physical, and nervous cost of commuting. The decisive factor is the increase in delay in arriving and the danger and tension. The ten or twelve miles between Wellesley and Boston may require twenty-five to thirty minutes with light traffic and good weather, in bad weather or dense traffic, forty-five minutes or more; and there are perhaps a hundred intersections. Over an adequate expressway the same trip might take fifteen to twenty minutes, with less tension and fatigue. A radial and peripheral system of improved expressways, permitting safe traveling speeds of seventy miles an hour—assuming corresponding improvements in the safety features of cars—would permit a city to double its effective radius and quadruple its potential area of integration. Our road experts have told us that "speed kills" if resorted to at the wrong time and place. But our city planners might well remind us that delays, too, can kill when their cumulative burden is added over a long period to an intensive working day.

Safe speed is not cheap. It cannot be achieved except by planned investment under public guidance. But it could do much to humanize life in our cities. The day may come when a profession of specialized expediters may watch over the smooth and quick flow of traffic and communication in our metropolitan areas, to identify and remove bottlenecks and overloads before their effects become cumulative and choking.

The same considerations apply even more strongly to public transportation. Improved and publicly subsidized rail transport—on the ground and underground—offers perhaps some of the most promising opportunities to combine high speed in mass transportation with safety at tolerable cost. The old-style commuting trains that take forty minutes for twenty miles not only exhaust their passengers but also drive more and more people to the somewhat faster highways. A drastic improvement in the speed and caliber of public transportation might relieve the pressure on the road system. Similarly, an extension of local telephone call rates to the entire suburban area—on the analogy of the successful principle of uniform postal

rates—might reduce some of the need to travel back and forth and thus further reduce the pressure on the transport system. Still another step might be the partial staggering of service hours, so that more stores and service facilities would be available for more hours daily, thus reducing the peak loads when all stores open or close. Rotating assignments and staggered hours might require more employees, but it might pay off in higher profits for the stores and in greater freedom for the community.9

None of these improvements would be cheap, and none easy to achieve. Such improvements, however, might be a key factor in rehabilitating our metropolitan areas. What is needed is a realistic analysis of the problem of peak loads and of the rising capital requirements for transport and communication. Only a substantial investment in transportation and communications can make metropolitan decentralization practicable, and only a substantial strengthening of public control over strategic land sites can restore beauty to our cities. Ways will have to be found to let planners use the powers of the community to guide urban growth toward a clear and pleasing pattern of new and old landmarks where people can once again feel well-oriented, exhilarated, and at home.

The various lines of research suggested in these pages have a common origin and a common goal. Our inquiry has centered on the function of a metropolis in aiding its residents in their choices and in their search for responses. The ranges and costs of such choices and responses are basic to our analysis. Proportionately accelerated investment in communications, together with an improved knowledge of the general order of magnitude of these proportions, suggests a possible approach to urban decentralization. It also points up the need for greater clarity and beauty in our cities, and perhaps also for more responsive government, capable of integrating a wider range of metropolitan and suburban services, if the expanded metropolis is to become a genuine home for its people.

9. For a discussion of some limiting factors and of the forces tending to pull the working hours of the whole community into a single rhythm, see Vilhelm Aubert and Harrison White, "Sleep: A Sociological Interpretation," Acta Sociologica, 1960, 4: no 3, 1-16.
TOWARDS A SCIENCE OF REGIONAL SYSTEMS*

Amit K. Maitra

We are concerned in this paper with decisions relating to long and medium term regional development. These decisions are made by the various levels of administrative structure within an institutional framework. The administrative agencies at the local, regional (state), and national levels are entrusted with different degrees of authority to devise and implement proper schemes for bringing about change in the geographical sub-areas within a country.

Political leaders and other experts at the policy making level most often have preconceived ideas about the developmental process. The decision makers attempt to find rules which when applied in a given socio-economic situation would foster greater economic development. Decision makers normally act as subordinates to elected leaders and in this capacity have certain discretionary authority to devise schemes that would eventually serve the needs of the general population. Elected leaders most often do not have the competence to evaluate the outcome of any particular developmental policy in great details. Furthermore, because of the relatively short span of time the leaders have at their disposal to devise and implement any plan, they are mainly interested in changes that would be immediate as well as impressive enough to gain public attention. If, however, the policy maker evaluates the impact of any decision rule which has technology as a variant, desired outcomes could indeed be achieved. Unfortunately, the actual case is often somewhat different. Politicians desire changes that are big, impressive, and need fantastic sums of money. The short-run improvements are more visible and more compelling. Thus a policy which produces improvement in about five to ten years is usually one which may often degrade the situation in the long run, beyond ten years. Again, programs designed to produce long-term improvements may initially cause depression in the desired behavior of the socio-economic system. This is particularly risky politically. But, in every case, at the outset very little attention is paid to examine how new technological change in a given socio-economic situation would interact with the social, political, and cultural factors. The reason for this neglect is that politicians and other decision makers have the general notion that man or society would change and adjust as technological changes are brought in. Moreover, in order to guard against any possible technical failures, the executives direct the engineers and other technicians involved in the project to work out every single engineering detail very precisely. Yet recent studies by different scholars indicate that the outcomes of these decisions are far less than satisfactory and may appear to be counter-intuitive as well as counter-productive in terms of the ultimate social gains. This has been noticed in most situations where new technology interacted with a wide range of social factors.

We are, therefore, confronted with the necessity of an extended knowledge base for policy making with respect to regional developmental efforts. This need motivates the current study that has been undertaken by a group of scholars of the University of Minnesota and the University of Rajasthan, comprising a case study of a developing region in the State of Rajasthan, in northwest India. The study has been designed to reveal how an irrigation system, which is thought of as the technological input in the area, interacts with a variety of other factors and thus changes the course of entire developmental activities in rather unexpected manners. The interest has mainly been focused to analyze the following questions:

- how to ensure better understanding of the process of decision making by which societies would respond to planned technological changes;
- how to design an analytic tool that could be used to predict the consequences of specific policies on overall development;
- how to use the analytic tool for determining the future policies necessary for maximum feasible rate of growth of a region.

In the following section a theoretical construct is developed relating processes of decision making to behavior evaluation techniques. The study introduces a methodological solution, based on ideas borrowed from cybernetics and engineering control theory, for understanding regional dynamics.

*Paper delivered at the American Association for the Advancement of Science annual meeting, Section on History and Philosophy of Science ("The World System"), December 28, 1972, Washington, D.C.

The work of this paper was in part supported by the Ford Foundation, and the Center for Comparative Studies in Technological Development and Social Change, University of Minnesota. I would like to thank John Brandl, James Jernberg, Darrel Frohrib, Vernon Ruttan (all from the University of Minnesota) for their critical comments and intellectual contribution. In particular, I am indebted to Richard N. Blue, Assistant Professor of Political Science, and Robert T. Holt, Professor of Political Science and Director, Center for Comparative Studies, without whose support and stimulating ideas this paper would not have been possible.
A region is an economic and socio-economic structure reflecting the patterns of human interactions. The need of a region is heterogeneous in character and the various categories of need will have to be satisfied in the future. The possibility of satisfying the particular categories of future need in given proportion will depend on the decisions taken in the present. The question, therefore, arises as to who should perform this task of decision making and in what way.

The overall regional dynamics model as has been established with the help of the control theoretic framework contains a directing and controlling sector (i.e., the government or a set of decision-making agencies that collectively and individually face the decision problem). In technical terms this has been regarded by Simon as the Inner Environment). Consideration has to be given to the control process in its entirety. Control process involves preparation of regional development plans and programs; their implementation through executive actions and review; the feedback mechanism necessary for adoption and adjustment.

Before elaborating further on this approach, I think it is desirable to clarify the meaning attached to the term regional development. For the purpose of this study the term would be used to mean the increase in welfare of the people of a region, expressed by such indicators as income per capita, its distribution among the population, the availability of social services and institution, the rate of unemployment and economic growth. It will be assumed implicitly that these above mentioned changes will occur as a result of technological change that might take place because of decisions made by the "Controlling Sector." Development is, therefore, a subset of processes taking place in a partially controlled environment which is different from the above mentioned controlling environment. This "Controlled Environment" represents the outer environment or the society in general. Thus far we have been able to trace main component parts of a regional dynamics model. The component parts are

(a) a directing or controlling sector (Inner Environment);
(b) a controlled environment (Outer Environment).

In order to understand the behavior of the decision-making body of a total system, attention will have to be drawn on the inter-relationship of the whole system and its dynamic nature. The dynamic nature of the system cannot be evaluated without taking note of the information feedback mechanism that operates between the Inner Environment and the Outer Environment. This feedback mechanism is actually the vital linkage provided by the political process, as indicated in Figure 1. The Government's plans are directed toward the development process taking place in the outer environment where two types of information are continuously generating—one is purely statistical in nature, the other is political in nature. In order to carry out the control processes, the I.E. (controller) must make use of both these types of information for transmitting proper decision (control) inputs to the outer environment in the next phase of action programs. Analysis of this feedback information system is vitally important for the I.E. (controller). This analysis will help the I.E. to perceive and assess possible future improvements that could be made to some medium and long-term regional development plan made by the controller. However, due to the complex and long-term character of the regional development schemes as well as comprehensiveness of the developmental objectives, information generated by the social and political processes is always very complex and involves time delays.

From what has been said so far, it should be clear that regional dynamics has been conceived as a set of activities that are connected both in time and space by a set of decision-making and behavior evaluation criteria. Overall system control and the availability of information on the environmental conditions restrict the decisions and decision rules. Decisions made by the I.E. affect the state of the system as a whole because of the functional inter-dependencies mentioned above. The dynamics of the states, \( x_i, i = 1,2,3,\ldots,n \) can be represented by a system of differential equations which specify the deviations of the actual from the desired states. These equations can be represented in the general form

\[
\frac{dx_i}{dt} = f_i(x_1, \ldots, x_n, u_1, \ldots, u_r, t) \quad i = 1, \ldots, n
\]

\( u_k(t), k = 1,2,\ldots, r \leq n \), are the time paths of the decision variables.
Due to the dynamics of the regional system, there would be a large set of outcomes and feedbacks in a continuous setting. This is $y_j(t), j = 1, ..., m \leq n$. Typically, the outcomes would contribute to insight into the process of regional development and to a systematic organization of the empirical facts referring to previous development and the actual state of the region concerned. The basic point of view referred to here could be expressed by splitting up the concept of outcome into two chains,

$$y_j(t) = \int_{t_0}^{t} g_j(x_1(\theta), ..., x_n(\theta))d\theta$$

The second form deals with the outcome that is based on a long-term developmental strategy and is shown by the entire course of evaluation of the regional system from time $t_0$ to time $t$ of interest; the first form, on the other hand, deals with the operational short-term or year to year level of outcomes. The fundamental assumption is that any long-term strategy for regional development can only be implemented through the short term operational control and implementation procedures. The two chains showing different outcomes are inseparably linked together at every stage. Furthermore, the instantaneous outcomes always impose constraints on the joint and simultaneous variations of the decision variables, $u_k(t)$. Some of these constraints may assume the form of an explicit limitation, such as

$$-a_k \leq u_k(\theta) \leq b_k; \quad t_0 \leq \theta \leq T$$

or they may be implicit in the manner in which control is organized. For instance, long-term strategy provides guidelines for short-term outcomes desired and ensures the efficiency of the planning objectives. The relevant theories here are theories for regional development control. These serve as a basis for formulating the operational ideas which must be quantified on the basis of empirical facts referring to actual states of the economic, social, and political situations of the region. The effect of the decision inputs should be examined against decision criteria established by the long-term project goals. Thus we may say that a subset of $s(s < r)$ of the decision variable $u_k$'s may be dependent on the others of the other $u_k$'s. The factual process of regional development shown by these decision variants gives rise to experience about the quality of plan and the realism of the long-term strategy, the adequacy of the institutional machinery represented by the I.E., wisdom of objectives and goals, as well as the efficiency of the model upon which the decision strategy was based. This feedback mechanism may give rise to adjustment of the I.E. and to improvements in the means. Also the implementation procedure might be changed. The confrontation of the factual developments with the plans should normally induce a new search resulting in a continuous variation of the various elements in the control framework. Thus we can come to the hypothesis that the regional development control is a dynamic process where feedbacks built into the control framework make for a sophisticated model.

A FRAMEWORK FOR FORMULATING REGIONAL DEVELOPMENT STRATEGIES AND OPERATIONAL PLANS

The characteristics of the regional system described above have been explicitly considered in a simulation model developed by J.W. Forrester. H.A. Simon has also suggested an appropriate decision making apparatus that can conveniently be designed to show interaction of the I.E. and the O.E. At this point, I intend to characterize the regional development strategies by carefully integrating some of the ideas of Forrester-type social system model and the more traditional economic analytic tools. It is my hope that this characterization of the model would give a detailed description of the simple idea of regional interactions. Figure 2 shows the structure that has been assumed. It interconnects the effects of population, capital investment, natural resources, and the fractions of capital devoted to agriculture and industry.

It is true that the system as suggested in Figure 2 is highly simplified. It draws our attention to a few major factors and omits most of the substructure of the regional social and economic activity. Nevertheless, the system representation of Figure 2 is highly complex and comprehensive in nature. The diagram in fact summarizes numerous time varying non-linear relationships in a way that shows the future possibilities. The powerful dynamic force suggested here, explicitly or implicitly (whichever way we choose to emphasize), indicates the eventual shifts that will occur in the components of the model.

Rectangles in the diagram refer to system levels which are variables. There is yet another kind of variable—the rate variable. The dotted lines in the diagram show the feedback loop structures which control decisions or actions. An intermediate concept is given by the circles which summarize many influences and affect the rate variable. These are called the multipliers. The action is accumulated to generate level. Information about the levels would be partially the basis on which further controlling rates would be manipulated. The diagram shows how each rate of flow is assumed to depend on the levels of natural resources, population, and capital investments.

The government or the decision-making body is implicit everywhere in the structure of the model, but no attempt has been made to identify the decision making-body by any kind of explicit notational device. In fact, the idea is to make proper representation of the decision-making body at a
Fig. 628 to NATURAL RESOURCES H. E. Power Generation

- Natural Water Runoff Riv. bed
- Water For Gang & Channel
- Water For Developing Region (Perceived)
- M. E. Power Generation

- Increased Agricultural Prod. Mult. due to greater water availability
- Table Multiplier to increase Agricultural Production Potential
- Table Multiplier to generate more power for new industries

- Quality of Life due to Production
- Production Potential due to Capital Investment
- Capital in Agriculture
- Capital in Industry

- Natural Death
- Normal Birth
- Migration

- Death Due to Natural Calamity Drought, Famine

- Growth Rate

- Capital Investment

- Total Capital Investment/Population

- Yearly Water Availability

- Rate of Usage

- Natural Resources
later stage when we integrate our findings. The identification requires an additional step in our analysis. The controlling agency should learn accurately the nature of the dynamics involved in a regional system. So, for the time being we will turn our attention to a description of the surface behavior of the component parts of the regional dynamics model suggested in the figure. The purpose is to explore how the real world operates. Forrester claims that a model of this type is far more explicit than the normally used intuitive models or mental models.

We can begin our investigation from the right hand corner of the diagram where the circle reads 'Quality of life due to production.' The multi-directional information loops determine the dynamics involved in raising the average quality of life for the region as a whole. The QLPM (quality of life due to production multiplier) serves the purpose of an index at any particular time. The index determines the rate of optimal utilization of resources for some specific time horizon. Better quality of life in a particular region attracts more people into the region. New migrants influence economic activities of the region. The regional programs, therefore, introduce new elements which set the criteria in sectoral programming, as suggested by the two multipliers, capital investment in agriculture and capital investment in industry. To meet the regional political, social, and economic demands, the sector programs are guided by regional distribution of demand, and regional population density as indicated by the loop structures. All these influences are exogenous from the point of view of a given sector and show evidence of the past trends, which, according to Forrester terminology, may mean time delays. The above criteria point out how the different sectors can choose efficient solutions which result in the desired regional pattern of social targets. It is quite natural that conflicts would develop between interest relating to the pattern of social targets and the criterion of efficiency in terms of the optimal utilization of resources. Social targets are determined by competitive political pressures from various sections of the population. With the aid of this revealed social preference, different projects to be undertaken by the government (state) can be suitably ranked in order of importance attached to the different projects. However, there is no guarantee that even with the best intentions the decision makers in the government will be able to specify and also maximize the political objective functions.

In recent times, project ranking formulas are widely used in government. The objective is always to maximize the sum of the scores of all the projects in order to derive an index of the QLPM. The complexity that arises out of the pursuit of the multiple objectives is due to the fact that there are interactions among the goals that are defined for the sub-function of the system. We can write the outcomes as

$$y_i = Y_i(u_1, ..., u_m) \quad i = 1, ..., r.$$  

We have already seen that the variations of some of the u-variables are constrained by the levels of the remainder. This suggests that the interactions among objectives may be represented by relationships between the outcome increments, $dy_1, ..., dy_r$, such as

$$\Phi_i(y)dy + \Phi_2(y)dy_2 + ... + \Phi_r(y)dy_r = 0$$

In this pattern of dependence, the set of weights $\Phi(.)$ are variables. The values of these variables depend on the levels of outcomes achieved. Our task is to identify the weighting functions $\Phi_1(.) ... \Phi_r(.)$ as the partial derivatives of some function, so that the construction of a measure of tradeoffs of objectives can be achieved. The top decision-making body of the system is constantly confronted with these tradeoffs in order to find a basis for an expression of preference. A note of caution must be inserted here. There are certain circumstances where this process of building up an index from a set of individual tradeoffs might be impossible even in principle; furthermore, in other situations practical computational difficulties might arise too.

The problem of finding an index arises because of two factors:

1. the comprehensiveness implied by the maximization criterion relating to social targets for the economy as a whole;

2. the multitude of individual decisions which must be taken in order to reach these targets.

It should be evident from Figure 2 and the previous discussion that decisions relating to micro-activities cannot be made simply on the basis of overall criteria of social rationality. The dynamic process outlines how the level of resources, capital, etc. constantly play the important roles in guiding the future decision strategies. The sub-units of the government's administrative structure are responsible for making micro-decisions. This shows the involvement of multi-level system of decision making. So, a distinction could be made among the following levels of decision making:

- macro-economic decisions related to the basic economic structure of the region, based on QLPM (Figure 2);
- intermediate level decisions related to the intra-sectoral development;
- operational decisions related to specific sets of activities such as location of investment project, etc.

In order to derive marginal substitution ratios or proper tradeoffs between different projects being planned and implemented by the different levels of government body, the levels must be adequately informed of social targets. The levels
should also be properly motivated to make decisions directed toward the implementation of such targets.

Looking at the regional decision problem against the background outlined above, it is now possible to go to our next stage of constructing an extensive model that would identify all the dynamic properties as they have been related to regional developmental planning. In the process of presenting the integrated form of the model, taking note of all the ideas discussed so far, I think we will be able to reveal in what way the answers to numerous questions should be expressed in the process of decision making at the various levels of government structure. The particular construction of the overall dynamic model in the following section should clarify how in practice the decision-making structure could include all relevant factors and forces which influence the solution of a set of decision problems. Although this form of representation may be applied for computational purposes too, we will not discuss any such specific experience but simply present the issues.

SPECIFIC FEATURES OF THE REGIONAL DEVELOPMENT MODEL

We now proceed to initiate a mathematical model of multi-level systems incorporating the ideas presented in the preceding sections. A glance through the previous sections suggests that our analysis of the regional dynamics model has thus far been based on a fundamental, yet subjective distinction. In the preceding discussion, the two concepts, controlling sector (I.E.) and controlled sector (O.E.), have been illustrated by their usage in regional dynamics. The distinction is drawn on the basis of what we consider as something we can play with (that is, what we define as the controlling mechanism I.E.) and what we conceive of as the representation of the society (that is, our environment). Some essential characteristics which most I.E. should have in common, like the dependence of the I.E. upon actual performance of the O.E., etc., have also been pointed out.

In this section, I intend to develop specifically
(a) some concepts for classification and studies of the controlling mechanism (I.E.) in greater details;
(b) some ideas of hierarchical systems which might provide conceptual foundations for handling the problems of interactions of the system variables in the O.E.

The controlling mechanism can be conveniently sub-divided into a set of sub-units linked together by flows of information. In Figure 3 an attempt has been made to give a picture of the decision-making process and the environment. The element of the controlling system that performs the planning process is assumed to be located in the top of the control unit, representing the National Planning Commission. We term this the Supremal.

The planning process is supplied with technical information relating to
—trends and state of the development process from local information generating and collecting units, representing the government agencies at the state level; we call these agencies the Infimals;
—development objectives and goals which are formulated by processes of political and economic adjustments between the supremal and the infimals.

We shall learn more about the adjustment process that actually aids the decision rule at a later stage. Figure 3 has been drawn to indicate how the factual developments in the outer environment are being influenced by the actions of the infimals and the supremal. In fact, there is not much difference in Figures 2 and 3, except that Figure 3 attempts to characterize every aspect of the complex process involved in a regional dynamics. With the help of the multi-level hierarchical structure, it should be possible to examine the relationship that exists between different levels of the I.E. and the different sectors of the O.E.

The decision problems for the supremal are more complex than infimals. The latter are concerned with particular local changes which directly and explicitly affect the environment. The important task of the supremal is to assess and evaluate changes that are caused by actions of the infimals. Supremal's main function is, therefore, to coordinate the actions of all the infimals. Supremal is concerned with broader and slower aspects of the overall systems behavior. It is actually not possible for the supremal to respond to variations in either the environment or the process itself. Supremal cannot act more often than the infimals.

There are two information signals connecting the supremal with the infimals. The downward signal specifies the decision problems for the infimals, while the upward signal furnishes technical information about the lower levels to the supremal. Figure 3 identifies all major types of information flows and information systems involved in the preparation and implementation of operational plans. We now plan to give detail attention to the identification of sources and the determination of mathematical notation of information systems for usage in planning.

For the purpose of simplicity and brevity, we would picture the planning process as constituting two phases. In the first phase, the supremal unit exchanges information or message with the infimals. In the second phase, these messages are taken into consideration for making plans of action (decision inputs) for the various infimals. Ultimately, these decision inputs are transformed into outcomes by actual implementation in the proper environmental components. The transformation of decision inputs into outcomes is a dynamic process evolving in actual time. Figure 2 has already shown how the transformation takes place in the real situation. Figure 3 points out that the operation of one
TOWARDS A SCIENCE OF REGIONAL SYSTEMS

HIGH LEVEL
POLICY-MAKING &
COORDINATING
AGENCIES

STATE
ELECTRICITY
BOARD

PUBLIC
WORKS
DEPARTMENT

PLANNING
DEPARTMENT

AGRICULTURE
DEPARTMENT

CIVIL
BOARD

CONTROL
INTERFACE

PHYSICAL SYSTEM :: TRANSPORTATION

PHYSICAL SYSTEM :: WATER DISTRIBUTION

PHYSICAL SYSTEM :: AGRICULTURE

PHYSICAL SYSTEM :: POWER REQUIREMENTS

ECONOMIC SYSTEM

SOCIAL ORGANIZATION

LOCAL POLITICAL SYSTEM

INNER ENVIRONMENT

OUTER ENVIRONMENT

:: Authority Channel

:: Information Channel
infimal is likely to be influenced in the time sequence by the operations of other infimals. After a certain stage in the planning process, communication is automatically established between the supremal and the infimals as well as between the infimals. It is, therefore, important to know the language in which communication takes place. In decision-making systems where numerous infimal units communicate with the central decision-making agency (supremal), the messages that flow are very complex in nature, and so the specification of the language is also a very cumbersome procedure. But one can hardly ignore this step, because it constitutes a very important feature of model construction for the decision hierarchy. Here we will adapt some mathematical notion developed by L. Hurwicz.

The process of information flow that takes place between infimal units and the supremal is thought of as iterative in nature, i.e., it consists of a sequence of stages. In Figure 3, the infimal units have been identified by the actual agencies serving under the direction of the State of Rajasthan in India; but, for the purpose of building a mathematical expression, let the infimals be labeled as $1, 2, 3, \ldots , n$. We may denote by $m_t^i$ the complex of messages sent out or received by the $i$-th infimal unit at time $t$. The iteration begins with an initial message $n$-tuple $m_0$ to be followed by $m_1, m_2, \ldots , m_{T-1}$ and the terminal message $n$-tuple $m_T$. The adjustment process is merely the time sequence of its stages, that is,

\[
\begin{pmatrix}
  m_0 \\
  m_1 \\
  \vdots \\
  m_t \\
  m_T
\end{pmatrix}
\]

where $m_0$ denotes the initial stage, and $m_T$ the final stage. Written out more fully, an adjustment process can be represented by a matrix whose rows represent time points and columns represent the infimal units:

\[
\begin{pmatrix}
  m_0^1 & m_0^2 & \cdots & m_0^n \\
  m_1^1 & m_1^2 & \cdots & m_1^n \\
  \vdots & \vdots & \ddots & \vdots \\
  m_T^1 & m_T^2 & \cdots & m_T^n
\end{pmatrix}
\]

In order to understand the process fully, we must specify how the message $n$-tuple of a given stage is formulated in relation to the messages of the preceding phases. According to Hurwicz, the rule which relates the message of a given stage to its predecessor is called response function. The response function of the $i$-th unit at stage $t$ would be denoted by $f_t^i$. The response depends on two factors—the earlier messages, and the unit's information concerning the environment.

The symbol for a complete specification of the environment is $e$. The description of the environment can be split into separate descriptions $e_i$, so that $e = (e_1, \ldots , e_n)$. In Figure 3, the model of complex environment has been shown on a stratified basis in reference to physical subsystems, social and economic subsystems. The figure is justified if the coordination problem can be sufficiently simplified with respect to the overall systems behavior. In a model representation of this form, changes in decision procedure necessitated by changes in the operation of a sub-process can be localized and accounted for. The potential for increased flexibility is there.

Now the dependence of response on the preceding messages and the knowledge of the environment may be expressed by the following equation system:

\[
m_t^i = f_t^i(m_{t-1}^i, m_{t-2}^i, \ldots , m_0^i ; e) \quad i = 1, 2, \ldots , n \\
t = 1, 2, \ldots , T
\]

The determination of the initial message can similarly be written as

\[
m_0^i = f_0^i(e) \quad i = 1, 2, \ldots , n.
\]

In iterative dialogue approach involving individual infimal units, there is an underlying assumption. In idealized form this assumption states that at the initial stage of the adjustment process each infimal unit knows only of its own environmental component and knows very little of other units' environmental conditions. We can then say that, for a given set of incoming messages, a given infimal's response function would depend only on its own environmental component and not on those of the other units. This is true only at the beginning; as time matures, the response functions assume more complex forms. For the initial condition we thus have

\[
m_t^i = f_t^i(m_{t-1}^i ; e^i) \quad i = 1, 2, \ldots , n \\
m_0^i = f_0^i(e^i) \quad t = 1, 2, \ldots , T .
\]

From the above arguments, we can easily conceive of the complex nature of information transfers that take place between the environmental components and the infimal units as well as the supremal unit. The computational and other difficulties associated with this phenomenon of adjustment process hardly need to be pointed out. In Figure 3, the matrix, shown as a rectangle on the right hand side of the diagram, represents the adjustment process that takes place as a result of interactions with the different environmental components. The other matrix, situated in between the environment and the decision hierarchy, indicates how infimal decision strategies interact with each other to produce a set of action proposals which will actually be implemented to bring about changes in the environment. The message $n$-tuple $m_T$ produced in the final stage would serve as the basis for decisions as to actions to be taken. Initially there might evolve a set of
inconsistent action proposals, which might be termed as draft plans. The objective is then to transform this draft plan into a real plan. This is done by scaling down or up all demands until they come reasonably close to resource supply at a particular situation. The transition from draft plan to real plan is iterative in nature, as we have seen earlier in the case of project ranking formula.

The operation that transforms the terminal message into a draft plan is given by the relation
\[ b = d(m_T), \]
where \( b \) denotes draft plan. In time, transition is made from draft plan \( b \) to the actual or real plan \( a \),
\[ a = r(b), \]
where the functional symbol \( r \) represents the realization function. Since \( m_T \) indirectly determines \( a \), we may write
\[ a = \phi(m_T) \]
with \( \phi \) called outcome function.

A plan should specify the present or future actions of all the units, and we may indicate this by writing
\[ b = (b_1, \ldots, b^n) \]
\[ a = (a_1, \ldots, a^n) \]
just as
\[ m_T = (m_T^1, \ldots, m_T^n) \]

CONCLUSION

A plan, either draft or real, may be thought of as a decision information flow matrix. In the terminology I have adapted, the choice of the response functions \( f \), of the outcome functions \( \phi \) specifies an adjustment process which in Figure 3 is denoted by the decision information flow matrix. To sum up, therefore, we can say that the concepts developed here provide a guideline for the formalization of the multilevel institutional framework. The purpose of this paper has been to point out the basic features of a regional development model and how the model should be established, operated, and expanded. To a large extent, the regional development will be conditioned by the particular institutional system. This descriptive phase or the model building phase of the study is aimed at describing the tasks and the institutional structure of the regional development organization within the complete administrative system of India. Particular attention has been given to the relationships between long term strategic planning, short term operational planning, and the organizations for the execution and review of actions and projects we found in Indian case. It must be mentioned here that it was not possible to include description of all the various units of the regional development administration, decision making, execution, and review processes. The aim has been projected to describing the basic structure and inter-relations existing between the more significant levels of national and state administrative structures which are supposedly more involved in the control and direction of regional development in the Rajasthan Canal and Chambal Basin areas. As stated at the beginning of the paper, we borrowed ideas from engineering control theory in order to establish our framework of the model. We could do that by postulating regional development as a controlled process. The positive aim of the empirical investigation will be at explaining
- the overall system of economic planning and management;
- the methods and techniques used in planning and decision making;
- the problems of regional development in a particular situation;
- the structure of the administrative system, etc.

This investigation should also reveal the actual inter-relations of plans and decisions, the functioning of information systems discussed here, and the constraints to be taken into consideration in the policy making, analysis, and administration. The project is divided into few phases. Hopefully, in the long run, the project might serve a practical purpose for the Government of Rajasthan in India and the other governments of similar regions in the other corners of the world.

REFERENCES


E.

APPLICATION OF SYSTEMS THEORY TO NATIONAL PROBLEMS
I

I am not a political scientist; only a student of communication, and, in particular, of the part which human communication plays in the regulation of human societies. This study leads back—or on—along two closely related paths. One is the study of systems generally, in the search for principles of regulation common to them all. The other is the study of communication generally, in the search for better ways of understanding those levels of communication which distinguish human societies from other types of system. These studies are new, many sided, and rapidly growing; I claim no expertise in either. But they seem to be the fields in which I can most usefully think aloud in the presence of political scientists.

They have also a topical relevance, for they seem to me to provide apt language for describing simply and sharply the principal threat which shadows the world's political perspectives and the principal dimensions along which escape will have to be sought. I will first describe this threat as a breakdown in the conditions which make possible the regulation of political systems such as support us now. Then I will analyse this breakdown a little further, first as an ecological trap and then as a failure of communication.

II

Let me begin with a rather arid summary of what I understand by the regulation of a political system.

By a social system I understand a set of ongoing relations between persons and organisations, governed by mutual expectations which are usually embodied in roles. It is, of course, a very complex pattern. Each of us forms part of several sub-systems and each of these is incorporated in varying degrees in others. Whether we focus our attention on the family, the neighbourhood, the city or on the factory, the university, the trade union, we distinguish something which we regard as a continuing entity but only to the extent and in the field in which it maintains, through time, two sets of relationships which are themselves intimately linked—the internal ones which relate its members to each other and the external ones which link it, as a whole, to its surround. The entity is in fact a pattern of relationships, subject to change but recognisably extended in time. This way of regarding the objects of our attention helps to resolve the ancient dichotomy between the individual and society and many other pseudo-problems resulting from the tendency, built in to our language, to regard the objects of our attention as "things," rather than systematically related sequences of event.

Within this comprehensive picture I will distinguish a political system as constituted by those relations which a society seeks to regulate by the exercise of public power. This definition would be too narrow for some purposes but it distinguishes one group of relations which deserves a name. The departmental organisation of central and local government distinguishes a host of relations which it is the function of these departments to regulate—the relation of roads and road users, houses and home seekers, schools and school children, sickness and hospitals; the level of employment, the balance of trade, the balance of payments, the balance of international power, and so on. Every political activity is directed to the regulation of some set of ongoing relations, whether internal to the system controlled by the regulator or external, between that system and other systems.

Regulation operates by manipulating one or other term of the relationship or both. We may build roads or restrict traffic, build schools or abstain from raising the school leaving age, increase the armed forces or cut our international commitments. Equally, of course, we may fail, partly or even wholly, in our regulative efforts. But even where we fail, I regard the relations in question as having been brought within the political system by the decision to treat them as regulable by acts of public power and thus to separate them from the host of other relations which are left to the regulation of the market or of the family or of other determinants.

Even my casual list of examples shows how changeable is the content of a political system thus defined; for most of the relations it mentions were not regarded as necessary or even proper subjects of regulation a few decades ago. If we tried to distinguish the changes which forced these new regulative tasks on to the public power, we should have, I think, to distinguish at least three kinds:

GEOFFREY VICKERS

the physical, the institutional, and what I will call the appreciative. In the first I include all the physical changes of an island increasingly urbanised, mechanised, and populated; in the second, all the changes in the institutions by which we carry on our collective living. In the third I include all the changes in our ways of appreciating our situation; what we notice and what we ignore; what we regard as acceptable or unacceptable, important or unimportant, demanding or not demanding action by us. I regard this appreciative system as no mere derivative of the other two. They interact mutually and determine each other.

Consider one example. For many millennia the river Thames has earned its name as a continuing entity. It is in fact the way in which water from a stable catchment area finds its way to the sea. It expresses the relationships, changing but continuous, between rainfall, contours and porosity of the area, vegetation, and a host of other physical variables.

Throughout this time until very recently its valley provided a habitat for many species, including men, who long ago learned to live above its floodmarks and to cultivate its alluvial soil. Then we began to incorporate this river, once an independent variable, into our own man-made socio-technical system. We controlled its floods with barrages and dykes. We adapted it for transportation. We distributed its water. We used it as a sewer. Our demands rose and began to conflict with each other, making necessary, for example, the control of pollution. Now these demands have begun to conflict in total with the volume of the river. We plan to supplement it by pumping out the deep reservoirs. Soon, unless some other solution appears, we shall be supplementing its flow by pumping desalted water from the sea. By then the Thames as an independent physical system, part of the given environment, will have virtually disappeared within a human socio-technical system. We controlled its floods with barrages and dykes. We adapted it for transportation. We distributed its water. We used it as a sewer. Our demands rose and began to conflict with each other, making necessary, for example, the control of pollution. Now these demands have begun to conflict in total with the volume of the river. We plan to supplement it by pumping out the deep reservoirs. Soon, unless some other solution appears, we shall be supplementing its flow by pumping desalted water from the sea. By then the Thames as an independent physical system, part of the given environment, will have virtually disappeared within a human socio-technical system. We controlled its floods with barrages and dykes. We adapted it for transportation. We distributed its water. We used it as a sewer. Our demands rose and began to conflict with each other, making necessary, for example, the control of pollution. Now these demands have begun to conflict in total with the volume of the river. We plan to supplement it by pumping out the deep reservoirs. Soon, unless some other solution appears, we shall be supplementing its flow by pumping desalted water from the sea.

Regarding the content of a political system as the relations which it aspires to regulate, I will describe as its setting the standards by which these relations are deemed acceptable or unacceptable. Such standards are essential to regulation. The problems of the traffic regulator are set by the standard of congestion which is regarded as unacceptable. Without such a standard there would be no problem and nothing to regulate. All regulation depends on setting standards by a process of human valuation.

Many people dislike applying mechanical analogies to human affairs but I find it useful, for contrast as well as for similarity, to compare political governors with engineers before the instrument panel of some mechanical assembly. The engineer watches dials, each of which displays the course of some important variable, showing how closely it approximates to some desired standard or how dangerously it strays towards some critical threshold. These standards and thresholds are the settings of his system; and these signals of match and mis-match alert him to the need for regulative action. The picture serves equally for the political governor. He too watches the course of a limited number of variables—limited by his own interests in them and further limited by the number which he can usefully attempt to watch and regulate; and he too depends on signals of match and mis-match for his guidance.

There are differences also. The indices which the political governor watches are for the most part not mere observations of the present state of critical variables but estimates of their future course, based on his latest knowledge of them (which is usually imperfect) and worked up by a process of mental simulation. A more important difference is that half his skill consists in setting the standards which he shall try to attain. For unlike the engineer, who controls a system designed to be controllable, the politician intervenes in a system not designed by him, with the limited object of making its course even slightly more acceptable or less repugnant to his human values than it would otherwise be.

In our society, as in many others, this setting has changed startlingly in recent years. The content of our political system—the sum of relations which we aspire to regulate—has grown and is growing in volume; and the standards to be attained have risen and are rising. The action needed to attain and hold these standards requires more massive operations, supported by greater consensus over far longer periods of time than in the past. On the other hand, the situations which demand regulation arise and change with ever shorter warning and become ever less predictable, as the rate of change accelerates and the interacting variables multiply. Clearly the task of the political regulator becomes ever more exacting.

By contrast, the capacity of political societies for accepting regulation is being eroded by several factors. The capacity for collective response is dulled when the situation which should evoke it is not present to experience but is a mental construct, based on uncertain predictions. It is further dulled by those policies of collective security which cushion the individual against even such present experience as he might otherwise have. It is further limited by the need for greater consensus and by the increasing vulnerability of that consensus to the resistance of protesting or predatory minorities. Above all, it is limited by the emergence of time thresholds, which deny the opportunity needed for the gestation of innovation. These factors, some of which I will explore in greater length in a moment, create, as it seems to me, a wild and growing disparity between the
least regulation that the situation demands and the most that it permits. This is the dilemma which preoccupies me and which I want now to examine more fully from the two angles which I described earlier—first, as an ecological trap; and secondly, as a failure of communication.

III

The first contribution we can draw from the study of other ecological systems is that these gloomy anticipations, even if they were fully borne out, would be in no way surprising. We have no reason to assume that political societies will prove to be regulable at any level which we would regard as acceptable. Many species have perished in ecological traps of their own devising. We may already have passed the point of no return on the road to some such abyss.

A population in a favourable but unfilled habitat normally multiplies at a constant rate until it meets or breeds limitations which slow and in time arrest its further growth. It may then stabilise, at or below its maximum, in the same or an altered form, with oscillations of less or greater amplitude; or it may even disappear, because in its period of expansion it has either unfitted itself for life in a limited environment or unfitted its environment to support even a limited population. These are the ecological traps I mentioned, in their most acute form. We call them traps only because our interest is engaged by the species they ensnare. From other viewpoints, such as an interest in the continuance of organic life, the replacement of one species by another is of no importance or appears as a salutary bit of regulation. If our kind exterminates itself and leaves an earth habitable only by creatures tolerant of a high degree of radiation (cockroaches are, I believe, favoured for the succession) only a judgment which values man and his works will notice any serious discontinuity. But we are human and we must take our human value judgments seriously. Here again we are reminded that all our thoughts about political regulation assume and depend absolutely on our human value judgments.

Not all species have worked themselves into ecological traps; many—the trilobite and the porcupine are stock examples—have established a stable relation with the milieu which has supported them for millions of years. In the past many human populations have similarly attained and held, for periods respectable by our human time-scale, a state of dynamic balance with their milieu, including the other species with which they shared it. Men have shared the Amazonian jungle with its other fauna for several millennia, without substantially changing the jungle or their neighbours or themselves.

The stability of these societies results from the fact that their way of life does not of itself disturb either the milieu or the society itself in its physical, its institutional, or its appreciative aspect. Each generation, taking over the skills, the institutions and the ideas of the one before, finds them as apt as ever to the milieu in which they have been developed and the purposes which they have been designed to serve. They may in time become inept through some independent change in the environment, slow or sudden—an eruption or a change of climate, an incursion of predators, or the impact of a foreign culture. But no significant change is generated by the activities of the system itself.

In the history of the human race, stability in this sense has diminished towards a vanishing point which, I think, is now in sight. The skills, the institutions, and the ideas of hunting tribes served their needs far longer than did those of agricultural peoples, because they did not generate the changes which would have made them obsolete. Even the agricultural epoch, those last three hundred generations which span the whole era of recorded time, gave birth to civilisation which deserve to be called stable over several centuries, in that their skills, their institutions, and their ideas did not change rapidly from one generation to another. We now seem to be approaching a point at which the changes generated within a single generation may render inept for the future the skills, the institutions, and the ideas which form that generation's main legacy to posterity, and the next generation's principal heritage. If this is true, it looks to me like an ecological trap identical in character with those which I described earlier, though the determinants of the trap are social and cultural, rather than biological.

IV

But is it true? The analogy is obviously far from exact. Ecological traps arise because biological evolution works too slowly to adapt some species or population to some environmental change or rate of change. Need we assume any significant limits to the far more rapid processes of cultural and political development?

I think we must. The reasons appear when we define the conditions that make regulation possible. They are, I suggest, four.

First, the regulator must be able to discriminate those variables that are involved in the relations it seeks to regulate and to predict or control their future course over a period at least as long as the time needed to make an effective response.

Secondly, it must be able to preserve sufficient constancy among its standards and priorities to make a coherent response possible.

Thirdly, it must have in its repertory, or be able to discover, some response which has a better than random chance of being successful.
Fourthly, it must be able to give effect to this response within the time which the first and second conditions allow.

Some would add a fifth condition—that the results of the response must be sufficiently distinguishable in the future course of affairs to prove or disprove its aptness and thus give the opportunity to learn by trial and error. I do not include this, because it seems to me that in important political decisions we often must and do get on without it. The results of most of our important decisions return to us long after the event, in manifold and often recognisable and indistinguishably mixed with other changes; and even when we think we can trace them, we cannot compare them confidently with what would have been the results of any alternative decision we might have taken; for at these we can only guess. I suppose we do learn something from experience, even in politics; but the process is so obscure and so seemingly remote from the ways in which we learn from simpler and more repeatable experiences that I will omit from consideration the limitations which are involved in this condition, except in so far as they are also involved in the other four.

The other four conditions are limiting enough. The most obvious disparity is between the "lead times" needed to mount any regulative action and the future span over which any reliable prediction can be made. The first grows ever longer; in important fields, such as changing land use and changing educational needs, it is reckoned in decades. The second grows ever shorter; changes unforeseen a year before may make nonsense of well matured plans which already commit huge resources. Plans for the reorganisation of ports have recently been made obsolete by developments in "containerised" transport. Examples could be multiplied. It is quite possible for the world as we know it now to become unregulable in important fields, in that it might pass the point beyond which any considered action might have a statistical probability of being worse than random. There are many situations in which to be systematically late is to be systematically wrong.

But the condition which produces this unhappy result is not primarily due to exploding technology but to the limitations of human communication. The long lead times which intervene between the emergent need for action and its achievement are partly due to the delays inherent in the processes of generating a sufficiently agreed view of the situation, a sufficient consensus on the course to pursue, and sufficient common action to achieve it; and all these are collective processes, mediated by communication. Even the confusion and loss which rapid technological innovation produces by its unplanned impact on other parts of the process only express a human failure to achieve, at that level, that phasing of complex activities which, at simpler levels, is recognized as a proper technological necessity and a proper technological skill. The far more difficult conflicts between nations, classes, and cultures—between rich and poor, communist and anti-communist, negro and whitey, Arab and Jew—remain insoluble so long as they reflect the lack of any common basis for communication. Technology has made this lack into a threat, by reducing the distance between cultures and increasing the distance between generations; but it did not cause the lack and it cannot abate the threat.

Since until recently science has tended to blind us to the extent to which all our activities, including science, depend on communication, it is worth recalling that human societies differ from others primarily because humans talk—and even listen. Nearly all we know comes from communication, rather than from observation or direct experience. Even the ways in which we classify experience were taught us through communication; most of them are inherent in whatever language we use. Our interests and our standards were formed largely through communication; the rebel, the deviant and the prophet, no less than the most conforming member of the Establishment, define themselves, by protest if not by compliance, by reference to the specific culture which made them human by claiming them from infancy as members of a specific communicative network. No one can make another do anything at all, except by the obscure process of communication. Even an atom bomb, we are told, is valued for the sake of the message it sends out while it is intact, rather than for the energy it would release if it went off. Change and stability in human societies are mediated hardly at all by those transfers and transformations of energy with which we model the physical world, but almost entirely by those transfers and transformations of information which physicists have at last learned to distinguish and have thus made respectable subjects for scientific speculation.

V

Since we depend absolutely on communication, within societies, and between the generations, developments which threaten these communications with failure are a lethal form of trap. By failure of communication I do not mean failure in the means to transmit, store, and process information. Of that we have already more than we can use. I mean failure to maintain, within and between political societies, appropriate shared ways of distinguishing the situations in which we act, the relations we want to regulate, the standards we need to apply, and the repertory of actions which are available to us. This fabric, on which communication depends, is itself largely the product of communication. Demands on it are rising.
We need to consider what chance there is of meeting them and at what cost; especially at a time when new techniques for handling information are finding their way into the regulative process at all levels, based on assumptions about how that process works which are not, I think, well validated and at the same time changing the process more deeply than we realise.

So let me turn to the threatened failure of communication in the three roles I have mentioned: its role in defining problems, in evaluating programs, and in securing cooperation and concurrence.

First, about the way in which problems are defined. In more static societies the relations to be regulated and the situations which involve them are usually familiar and are often equally visible to all. In our society, the more numerous relations to be regulated combine in subtle and often novel ways and the situations which involve them have to be anticipated by techniques of simulation, often involving the combination of large volumes of information. So the task of defining the problem is by no means simple. Ways of seeing the problem become obsolete, no less than ways of solving it.

To revert to an earlier example, it was possible until recently to think about the distribution of water in southern England, whilst taking for granted both its supply and its disposal. Today, distribution can be usefully considered only as part of a much wider problem; hence, amongst other things, the merging of river and water authorities and the setting up of a water resources board.

It was possible until recently to think of traffic regulation simply in terms of better roads. Today, as Professor Buchanan1 has so lucidly shown, it makes no sense except as part of the problem of providing buildings with accessibility and thus as part of the still wider problem of three-dimensional town planning; for towns contain and relate the buildings where all journeys begin and end and of which accessibility is one among many necessary attributes, none of which can be fully enjoyed without denying others. Town planning, in turn, cannot be usefully approached until we have rid our minds of the concept of a street as a multi-purpose space, a concept as inexact today as the multi-purpose great hall of a medieval house would be to modern domestic living.

This need constantly to restructure problems makes novel demands on communication. For policy making is a collective activity and the first condition of the communication which makes it possible is that the participants should be talking about the same thing, or at least should know when this is not so. Most of the discussion which goes into policy making is directed to reaching agreement on how the situation can most usefully be regarded; in other words, what is the complex of relationships most significantly involved. Policy making is vastly complicated when this cannot be taken for granted but must constantly be reviewed.

A very simple example is provided by the efforts made here and in U.S.A. to coordinate the policies of the three fighting services into a single defence policy. Fifty years have passed since the war which added a third fighting service and made it apparent that all important future activity in that field would be combined operations. Yet neither country has yet succeeded in fully establishing what I will call a common ideology for the three services. Where this is lacking, the participants in the potential dialogue lack in effect a common language in which they can fruitfully disagree—a condition which is by no means confined to the fighting services.

The situation to which the policy maker attends is not a datum but a construct, a mental artifact, a collective work of art. It has to be simplified, or it becomes unmanageable; yet if it is over-simplified, it will be no guide to action. It has to reflect present and future reality; yet if it departs too sharply from the familiar thinking of the past, it will not be sufficiently shared by those for whom it has to provide a common basis for discussion. It has to be not merely discovered but invented, not merely invented but chosen from among several alternative inventions, each a valid but differently selected view. Most difficult of all, it must not obscure the views which it supersedes.

Crime, for example, has long been regarded as a violation of the social order, demanding correction, and a violation of the moral order, demanding expiation. We have learned to regard it also as being, sometimes, a cry for help, demanding response and a protest against the social order, demanding attention. Each of these views invites action partly inconsistent with the others; yet we cannot afford to suppress those on which we cannot act.

Thus the definition of political problems becomes more difficult, as the relations to be regulated become more numerous and involve more diverse conventional views. The more thorough the analysis of the situation the more complex it is found to be; and this complexity consists in the variety of inconsistent values which call for optimizing or at least "satisficing." The problem of "traffic in towns" as it was handed to Professor Buchanan looked far simpler than the problem which he handed back to the policy makers. His analysis disclosed the multi-valued choice latent in any adequate analysis of "the situation."

1. Colin Buchanan, professor of Transport at Imperial College, London, and chairman of a committee which produced for the Minister of Transport a report on urban traffic ("Traffic in Town," H.M.S. Office, 1963), commonly referred to in Britain as the Buchanan report.
VI

Now for the evaluation of programs. Here again it is convenient to look first across the Atlantic. For as you know, the defence department of the U.S. government has pioneered the application to government procedures of methods first developed in industry during and after the last war, to increase both "efficiency" and "effectiveness." Added interest attaches to these methods because of the President's direction that other departments should adopt them. They are clearly and modestly described in a little book recently produced by the Rand Corporation.2

The object of these procedures is, first, to ensure that action serves policy; secondly, to improve the information on which to choose between one program and another; and thirdly, though much more modestly, to improve the information which may help to guide the distribution of resources between one field of policy and another. These objects are to be attained by budgeting activities according to the policies they are supposed to serve, rather than according to the departments which carry them out; to budget them over periods long enough to disclose their real costs and benefits; and to interpret costs and benefits according to the meaning which policy gives them. Benefit means success in implementing the policy concerned; cost means the loss of whatever else the resources so used might have achieved.

This may not sound revolutionary; but it is. Its first finding is to dispel the assumption that action now in train serves any policy at all. This is not to be assumed. It may have become self-perpetuating. It may serve some policy long abandoned or some purpose which some other policy is designed to frustrate. It may be an expression of empire building or Parkinson's law, of pure competitiveness or the tyranny of technological or administrative fashion. These risks are not excluded by the present system of annual budgets for action departments. These may be needed for other purposes but they are inept for keeping action in the service of policy. They should be subordinated to a separate system of long-term program budgeting.

The procedure depends on and assumes the clear, prior definition of policy in terms of objectives. This, the memorandum insists, is a primary, creative act of choice, with which the program budgeter cannot help. For choice involves valuing, and valuing is by implication excluded from the rational process. No calculus can compare the relative values of atom bombs and medicine; or even those less dramatic disparities which trouble the policy maker at every level, such as the relative importance of improving primary or secondary education. The program budgeter can answer no question of value. All he can do is to "sharpen his (the policy maker's) intuition" by telling him what he may expect from what he is doing or contemplating now, and what he might hope to achieve by applying the same or different amounts of resources in some other way.

So to benefit from this procedure, the policy maker must set his own house in order. He must define his objectives and group them in a small number of fields, which should be so far as possible distinct. The memorandum recognises that these are searching requirements, for it treats as open the question how far this procedure will prove applicable to civil policy and civil departments.

For first, the grouping of objectives is itself a creative act, affecting both how they are seen and how they are valued. It determines which shall most actively compete and which of the policy implications of any proposed action shall be most effectively hidden. Within the defence department, it was relatively easy and obviously useful to group together, for example, all weapon systems which contribute to a policy of strategic retaliation, irrespective of the service which controlled them. Similar groupings in the civil field might prove more arbitrary.

Further, however objectives are grouped, they will overlap. Water conservation is an aspect of agriculture and of health, even of agricultural price policy, as well as the conservation of a natural resource. Yet the development and conservation of natural resources has a strong claim to be considered a major policy field in its own right.

Thirdly, however policy fields are grouped, they will not coincide with the boundaries of departments organised for action. There is at present in the United States no Secretary of State responsible for the conservation and development of natural resources; and if there were, he could not gather into one executive department activities which today are divided between eleven.

You will notice that this procedure bears the marks of its origin. It does not distinguish between objectives to be attained once for all and standards to be maintained through time, which I regard as a more adequate description of the setting of a regulator. It assumes that the policy maker's only problem is the apportionment of scarce resources between conflicting claims. It does not take account of possible conflicts between the objectives themselves—for example, between defence policy and other aspects of foreign policy.

It therefore assumes that costs and benefits can be usefully, if not completely calculated in terms of the policy which the action is designed to further.

and need not follow its endless repercussions in other fields—except, of course, the repercussions of its claim to resources. (This is reflected in its more commonly used name of cost-effectiveness.) It assumes, in brief, that planning and programming involve the comparison of different means to attain an agreed end, the means being neutral.

Now though this is not true even in industry, still less in defence, it is, I think, a more useful assumption in those fields than in the fields of civil policy; and it may well be that methods devised to evaluate weapon systems will prove inept to evaluate poverty programs. But these simplified methods should not on that account be dismissed. Some simplification there must be, and therewith much exclusion and distortion. It seems to me nonetheless useful to experiment with them, both to explore the limits of their usefulness and to discover what passes for planning in their absence.

In the meantime we should not assume that program budgeting models the process by which all plans and programs are compared. Even a chess-playing computer, I understand, when involved in the complexities of the middle game, cannot guide its play by a strict analysis of costs and benefits but has to rely on general principles. And unlike chess, politics, except perhaps in the defence department, is not a "zero-sum" game: indeed, is not a game at all, in that it has no built-in measure of success.

Thus program-budgeting, as I have so far described it, is a procedure designed not for regulators but for operators, seeking defined and single-valued objectives. This, however, is not all: for as the Rand memorandum candidly puts it, "program budgeting begins with structuring the problem." The program-budgeter, in other words, can help the policy maker to define what the problem shall be deemed to be. If he wishes to define the problem in a way with which his budgeting techniques can deal, he will do his best to have it simplified to the pursuit of one defined, single-valued objective. If, on the other hand, he uses his analytic skill as, for example, the Buchanan committee used theirs, he is likely to define a problem of multi-valued choice which will defeat his budgetary techniques.

Thus the multiple analytic skills of the program budgeter reveal a dilemma inherent in the policy making process. The more crudely simplified the objective the more efficiently it is likely to be pursued. Given a single-valued objective and a repertory of "means" assumed to be comparable simply by their cost in resources, it may be possible to demonstrate objectively which means is the "best." But no political problem can or should be stated in these terms; the more truly we present to ourselves its multi-valued nature and the multi-valued effects of all the means by which we might pursue it, the more impossible it becomes to compare either the costs or the benefits of alternative solutions.

The solution to any multi-valued choice is a work of art combining in an unique way the regulation of the various relations involved. The problem of the policy maker is to choose between such solutions (if he is happy enough to have more than one to choose from) or boldly to tell the planners to think again. But he cannot tell in advance what combination will prove attainable or even preferable. The Buchanan report, for example, discloses to him the values which he needs to combine, the time scale on which he should think, and the limitations which he should not forget; but only a definite plan for a defined area can show him one of the unnumbered combinations between which, if he knew them, he might choose.

A foreign policy, an educational policy, no less than a development plan, if they are to be more than a reaction to the most obvious pressing danger, need to display the characteristic qualities of a work of art; and call, in consequence, for common attitudes, as well as skills in those who assess and support them, no less than in those who design them.

VII

I come to my third query: the role of communication in securing the concurrence needed to carry a policy into effect. In concurrence I include everything from understanding commitment to grudging acquiescence, so long as it secures what is needed from the individual concerned. We do not know much about the many and various ways in which communication secures agreement or exacts obedience but we can recognise situations in which its task becomes harder, and the illustration I have just used supplies one example.

Since program budgeting requires the policy maker to group his objectives in fields which cannot correspond with the boundaries of departments organised for action, it follows that each executive agency must expect to implement more policies and to frame none; and that each policy making authority must rely for the implementation of its policies on more executive agencies and must expect to control none. This is bound to make more acute, within the regulator itself, the ancient problem of controlling executive power.

We already know this problem. Policy makers insist, if they can, on controlling the departments which implement their plans, because, as mere coordinators, their communications are usually too weak to support their responsibilities. Even the "overlord" with overriding executive authority over several departments, often has a hard struggle to assert it. We may well conclude, then, that if action is to be subordinated to policy, the power of communication to secure agreement, or
This growth of power alarms many, as well it may; but we can evaluate it only in the context of the dilemma, which I have described, within the machinery of government and which is writ even larger in the relations of government and governed. The same nexus of communication must mediate both the dialogue which keeps policy under review and the cooperation which allows it to be executed. Within the government machine, oriented as it must be to action, the chief danger may be that the momentum of action will defeat policy. Between government and the governed, the more threatening danger may be the opposite; that the debate on policy will frustrate any sustained action. Either would be lethal. Both can be excluded, I suggest, only by the kind of consensus I have described, sustained at national level. This in these days can be sorely tried. The agonies of American admirals and generals, which the procedure of congressional committees allows us to share, are similar in kind but, I think, far less radical in their scope than those which beset many responsible trade unionists in Britain today, invited to regard the preservation of collective bargaining as only one of several competing "values."

The dialogue between government and governed has become an immensely difficult exercise in communication. First, it involves much more searching demands on the governed. I said earlier that in principle we can regulate our relations with the milieu by manipulating either the milieu or ourselves. The first is more comfortable and more congenial to a technological age but the second is increasingly the pattern of the future. To control the flow of the Thames we had only to discipline the river but to control its pollution we have to discipline ourselves and this transition is typical.

Further, this dialogue becomes more comprehensive; for all a government's actions are themselves communications, demanding explanation. When unemployment was not deemed to be regulable, people responded to losing their jobs by finding others, if possible in other and perhaps more stable occupations. When they lose their jobs through what they regard not as an act of God but as an act of state, they may respond differently. A doing is not the same as a happening. It is a message, as well as an event.

Thus every government becomes involved in an endless exercise of explanation, the difficulties of which are enhanced by the three factors I have examined. The situation is often a novel view of

the relations to be regulated, not yet shared by all parties to the dialogue. The solution, if commended by an adequate "cost-benefit" analysis, will almost certainly not be the one which those most concerned would favour—still less, if it relies on the more aesthetic criteria on which we fall back when cost-benefit analysis fails. And its implementation will sometimes require wider concurrence than the traditional powers of government can achieve, even extended as they now are.

So the crucial purpose of the debate is to generate trust or distrust. And this, perhaps the most important political function of communication, depends not only on the constraints which are placed upon criticism but also on the incentives which are offered to it by the shape of political institutions. The fact that in this country [England] we have at all times an alternative government in being and more or less identified mobilises the attack on the government's credibility in a way which is I think peculiar to us. Consider the relatively sheltered position of an American president compared with a British Prime Minister.

VIII

Clearly the demands on communication are rising in the regulation of political societies, both within the governmental regulator and between it and the governed. Can we hope that increased understanding of this essentially human art and increased skill in its use will match the rising need?

The traditional means of making people do as they are told are weakened in practice and discredited in principle. On the other hand, more potent and subtle means of moulding opinion and attitudes are available. Whether these formidable engines will be either properly controlled or properly used remains at present, I think, an open question.

Our former means of estimating the results of actual or hypothetical courses of action have been enormously extended but at present they serve single-valued rather than multi-valued choices and are likely, I think, to reinforce the tendency to oversimplify which always besets harassed regulators when faced by multi-valued choices. They are designed for operators, rather than regulators. It is no accident that they entered the machinery of government through the defence department.

On the other hand, they have vastly fortified the process by which situations are defined, and the readiness to redefine them. And this is of the greatest value at a time when traditional definitions of situations and traditional attitudes towards them are becoming increasingly inept.

This in turn affects the structure of the appreciative system on which all communication depends. In the political context it is a solvent of ideologies. But communication is also a builder of ideologies. It is, I think, in its impact on political ideology that our increasing understanding of communication is most important.

We have come a long way since Condorcet, waiting for the guillotine, could look forward to a day when the sun would shine on a world of free men with no master but reason. He did not doubt that free men would find in their reason a sufficient guide to what T. H. Huxley was later to call "the State of Art of an organised polity." How much more modest is the role allowed to reason in the Rand memorandum, blandly relegating all questions of value to human intuition!

If indeed we have reached the end of ideology (in Daniel Bell's phrase) it is not because we can do without ideologies but because we should now know enough about them to show a proper respect for our neighbours' and a proper sense of responsibility for our own. The critique of ideology is, I believe, the most important political function of communication; a critique which needs to be creative and conservative no less than destructive. These self-spun webs alone support us in the abyss of non-humanity. Major operations on them are fearful enterprises. We shall be better equipped for such enterprises if we share a common understanding of the strange fabric we are working on and the still stranger instrument we are working with.
An integrative model that contains a high degree of generalizability and appears to be congenic to general systems theory is currently being used widely as a model for planning and executing public policy. Although the planning-programming-budgeting-system (PPBS) has operational utility as a management device, it provides a conceptual framework for delineating the complexities of purposeful, goal-seeking public policy. The fundamental problems shared by men of reflection (scientists) and men of action (policy makers) are the relationship of parts to wholes and the levels of integration of parts into wholes. PPBS, which was largely developed by interdisciplinary researchers of Rand Corporation and popularized by the federal government, is more a mode of thought than a mechanical tool. It transforms the conceptual clarity of general systems theory into a methodological, policy-making strategy for use in numerous and diverse contexts ranging from defense planning to the management of elementary schools.

One of the original aims of the Society for General Systems Research being to investigate like forms of systems in various fields, the paragraphs that follow are presented with that aim in mind. The objectives of this essay are to (1) describe why the PPB System may be conceived as a derivative of general systems theory; (2) portray some of the structural-analytical properties of PPBS; (3) review some of the governmental units that have attempted to install this system, and (4) consider some of the limitations of systems analysis in the conduct of human affairs.

GENERAL SYSTEMS THEORY AS THE BASIS FOR PPBS

The major disciplinary trends in the theory of general systems have been convergent rather than divergent, and they have served to enrich a common conceptual scheme. If it can be conceded that the present status of general systems theory is one of a family of theories awaiting integration and formalization, but being mutually very compatible, then the proposed basis for planning theory (PPBS) can be regarded as a member of that family. The general theory of systems implies methodology in its strict, philosophical sense, and the methodology of systems is represented symbolically by the design of models. The isomorphies that were found to exist among models developed to portray biological, physical, and social systems are among the most exciting and fruitful intellectual discoveries of the past two decades. Models of organizations have been developed and administrative theories formulated in the area of management science. Concepts such as wholeness, hierarchical differentiation, orderliness, interdependence, open systems, and steady-state behavior seem especially applicable to those formal, social organizations that are governed by public policy.

Among the most prominent of the family of emergent economic planning models is the PPB System. Although PPBS is a particular type, or subclass, it embodies universal principles that apply to systems in general. It can be described best as a general theory of planology because it is appropriate for planning in nearly any type of organization.

In the context of management science, systems analysis is defined as a way of examining the differentiated components and identifying the mutual interactions of some integrative unit. A system is a set of two or more interdependent elements and their relations, while analysis provides glimpses into the unit's parts and operations. The essence of the systems approach in organizations is to construct a model of the means-ends configurations by comparing optional courses of action systematically (quantitatively when possible), using a generalized sequence that can be retraced and verified by others. Contrasts to this approach are represented by leaders who create public policy by means of charismatic paternalism, sophistry, illogical judgment, unaided intuition, and action based upon irrational premises. Systems procedures are being used widely as research techniques to provide supportive data for public policy. Scholars who contribute interdisciplinary data to policymakers are increasingly choosing to regard intricate topics of human inquiry as systems, and they are reformulating issues of conflict in systemic terminology. Similarly, PPBS is an endeavor to frame a coherent, logical, generalizable scheme in

terms of which the discrete elements of a given system can be interpreted. Without belaboring the issue of possible homologies involving an organismic system and a PPB System, let us consider some of the properties of the latter.

STRUCTURAL-ANALYTICAL PROPERTIES

OF PPBS

PPBS is neither a revolution in government management nor is it merely old-style approaches couched in new terms. What is new is the way in which separate concepts that evolved in the past have been brought together in a comprehensive system. The integrated PPBS actually provides relatively few innovations for the individual elements of planning, programming, and budgeting; its value lies in its systematic coordination of all of these. A conceptual basis for the total allocation process is provided. Planning, programming, and budgeting constitute the process by which objectives and resources, and the interrelations among them, are combined to achieve a coherent and comprehensive program of action for any organization conceived as a whole. The need for a holistic approach arises from the indissoluble connection between the allocation of human and material resources, or budgeting, and the formulation and conduct of policy. Assumptions related to the philosophy of science are implicit in this attempt to model the process of policy-making. For example, one assumes that a unit of government can determine its policies most effectively if it chooses rationally among alternative courses of action that are placed in order of priority on the basis of anticipated benefits and costs for each alternative. As an operational tool, PPBS is being used to manage human conduct in such diverse contexts as corporations, universities, the Department of Defense, public schools, municipal and state governments, research and development centers, the YMCA, and others. Many public organizations and public officials exhibit a kind of cultural tendency by claiming that all values (or programs for the public) are important, and further, that all are equally important. If all values are of equal importance, then no value can serve as a guide to conduct. In this respect, PPBS provides a priority scheme for assigning relative valuations to the alternative courses of action.

Presented below are twenty major characteristics of PPBS:

1. Strategic planning. The proposed planning orientation includes a time horizon of at least five years, as opposed to the present usage of tactical planning (one year time spans). The extension of the time horizon is a way of attempting to exert control over the future instead of merely reacting to it and being controlled by it.

2. Programming. Activities of an agency are grouped according to common objectives in order to determine the manpower, materials, facilities, and time required to support a program.

3. Budgeting. The nucleus of PPBS is the program budget. This differs from conventional object-of-expenditure budgets in that it emphasizes programs to be accomplished rather than objects to be purchased. The program relates programs to resources that are transformed into budget dollars over at least a five-year period.

4. Output emphasis. The budget is structured on the basis of outputs, missions, functions, activities, services, or programs, rather than on conventional input items. A program is related to operational objectives, and consists of activities to be performed, subprograms, and program elements such as human resources, materials, space, and facilities.

5. Input-output coordination. Program budgeting seeks to relate inputs to the programs of an organization in a way that enhances a rational mean-ends calculus.

6. Evaluation. Comparison of desired outcomes, as expressed in programmatic objectives, may be made with actual accomplishments. Performance indicators are utilized, such as indices that measure cost-effectiveness.

7. Quantitative analysis. In order to analyze comparative benefits, quantitative measures are applied if they are available. These include such techniques as input-output analysis, benefit-cost analysis, operations analysis, linear programming, simulation, queuing, gaming, and others. Qualitative measures are also included and they should play a major role in public planning.

8. Multiple options. Program budgeting provides a framework for the consideration of all relevant alternatives for a particular course of action. These are then placed in order on the basis of desirability, feasibility, least cost, available resources, and other criteria.

9. Program review and revision. In addition to program formulation and analysis, procedures for periodic review and modifications of programs are specified. This process is dynamic, viable, and promotes the adoption of innovations. Intraprogram studies are encouraged.

10. Future needs identification. More explicit assumptions can be made about future demand and production functions of the organization. Risk is reduced and assumptions are specified. Total cost


646
implications of long-term undertakings are expressed. Scenarios are designed to portray future environments.

11. Economic rationality. There are at least two schools of budgetary theory, one subscribing to economic rationality and the other to political rationality. Program budgeting is essentially an economic concept designed to serve in the political arena. It represents an encroachment of economics upon politics, and is an embodiment of classical political economy.

12. Flexibility. Program budgeting does not impose arbitrary constraints. Programs can be defined in any way suitable to a particular organization. This approach is appropriate both for internal programming and control and for development of future policies and programs.

13. Openendedness. Expenditure items are not treated as givens. The amount available for the total system is determined and the highest utility alternatives are selected for the specified budget programs. The major objective is not expenditure control in the fiduciary sense.

14. Policy determination. Budgeting becomes an integral part of the administrative process with PPBS. The financial administrator shares in policy formulation, and provision should be made for a continuing dialogue among the policy maker, systems analyst, budget officer, and the organization members affected by such policy.

15. Decision centers. Within organizations using PPBS, decision centers and cost centers are developed so that administrators can have at their fingertips historical and projected information from all phases of activity. These are retrievable from a computer data bank. Simulators of the budget can be developed.

16. Cost neutrality. Program budgeting is neutral on the issue of cost reduction. A "cult of efficiency," measured by less spending per se, is not the criterion of success for PPBS.

17. Structural variability. Organizational structural variability is maintained, because the operations may be either centralized or decentralized. If the latter, participatory planning is encouraged so that policy-making is shared by all members of the organization. If the former, lowest departmental level budget estimates form the building blocks for the next level where they are aggregated, reviewed, and transmitted upward to the highest level.

18. Accountability and performance measurement. Program budgeting can be used for control and internal management purposes to review personnel data, output data, and resource data. It seeks to measure performance and affix accountability. Cost accounting procedures are generally a part of the program budget format, although it may be necessary to redesign an existing accounting system so that it will provide information to meet line-item, legal accounting requirements, and also provide data on program costs.

19. Concise budget document. The actual program budget document should be concise, but complete, and should be understandable even to a lay reader. Local school budgets frequently use terminology and nonprogrammatic categories that are too technical and unclear for their citizenry. With PPBS, the programs are likely to be stated in such a way that the public can grasp their contents easily.

20. Preservation of the past. The program budget encompasses the best features of previous budgeting formats: the executive budget; line-item, object budgets; and performance budgets. It need not replace conventional budgeting procedures. For the time being, line-item and program budgets probably should be maintained concurrently as a means of describing an organization's expenditures on both an input and output basis.

In summary, the structural aspects of program budgeting involve formulating a set of categories oriented toward "end-objective" activities that are meaningful from a long-range planning point of view. These characteristics are in marked contrast to conventional governmental budgeting, which stresses functional and/or object-of-expense categories and a very short time horizon. Analytical procedures are an integral part of program budgeting, and their primary objective is to examine systematically the alternative courses of action in terms of utility, effectiveness, and cost, with a view to clarifying the relevant choices open to policy makers. An information system is designed to support PPBS by means of (a) progress reporting and control, and (b) providing data necessary to carry out the analytical process.

IMPLEMENTATION OF PPBS IN GOVERNMENT

Thus far the discussion has focused upon the conceptual-descriptive aspects of the PPB System. In what follows, the operational-technical aspects are considered in an examination of a number of government units that are phasing-in PPBS. As of mid-1968 it is difficult to identify organizations in which this system is completely operational, and it seems fair to assert that PPBS is in a formative rather than final stage of development.

Federal Level

PPBS was formally introduced into the Department of Defense in 1961, and, after enjoying some degree of success, spread throughout the federal government. A Presidential directive called for the adoption of this system for the entire fiscal 1968 budget. In presenting a rationale
for PPBS, the President criticized conventional, non-systematic planning and budgeting practices. The shortcomings may be summarized as follows: (1) under the existing format, objectives of agency functions and activities have too often not been explicated with clarity and concreteness; (2) alternatives have not been sufficiently provided for consideration by decision makers; (3) program review for decision-making has been frequently concentrated within too brief a time period; (4) accomplishments have not always been specified adequately; (5) future-year costs of present decisions have not always been considered; and perhaps most important, (6) formalized planning techniques and systemic thinking have had an insufficient effect on decisions pertaining to public governance.6

To what extent was PPBS fulfilling its expectations at the federal level? In his budget messages for both fiscal 1968 and 1969, the President contended that the concept had helped to provide better methods of accomplishing program objectives and had contributed to greater overall efficiency. He cited eight particular programs to which systems analysis and programming had been of value: disease control, child health, urban planning, agricultural research, tax administration, vocational rehabilitation, veterans' pensions, and employment programs in depressed areas. An Office of Executive Management (OEM) was created in order to integrate separate agencies. OEM was formed to analyze government structure, provide assistance to agencies in developing information systems, coordinate operations with state and local programs, and to assist in fiscal affairs related to PPBS.

One way of illustrating the systematic sequence of a federal program structure is to start with a general objective, for example, "elimination of poverty." This can be broken down into a number of more specific objectives, such as "income maintenance" and "community improvements." These lead to narrower subobjectives, for example, "educational programs for disadvantaged youths" and "adult manpower development." From this highly differentiated level of subobjectives, the rationale is provided for such programs as Head Start, Elementary and Secondary Education Act—Title I, and Student Scholarships. Ultimately, the specific elements and means for accomplishing the end objective of elimination of poverty are found at the lowest level, or program element level, of a program structure. Current expenditures as well as the future projected costs can be determined for the general objective and for each program element.

Another illustration is contained in the analysis of the benefits and costs of Job Corps training portrayed in Table 1. Does the investment in Job Corps training meet the economic criterion of efficiency? Do benefits exceed costs? A case study based upon data provided by the U.S. Office of Economic Opportunity compares the improvement expected in lifetime earnings of Job Corpsmen with the costs of training. Two measures of earnings improvement are used: one based on the additional earnings associated with educational gains, the second based on a comparison of wages earned by former Corpsmen and a control group without Job Corps training. The future stream of earnings is discounted to arrive at present value. The example shows the importance of the discount rate in determining whether a program passes the test of efficiency.7

### Table 1

**Benefit-Cost Estimates for the Job Corps**

<table>
<thead>
<tr>
<th>Benefits: Present Value of Future Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thousands of Dollars</strong></td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

**Costs:**

- 3% Discount Rate: $5,664
- 5% Discount Rate: $5,664 (5-Months Training)

![Benefits and Costs Illustration](image-url)

PPBS has served to exert influence against the inertia of entrenched federal programs, and it seems to be moving the nation closer to the day of a single federal budget document that is prepared almost entirely on a systematic program basis.

**State Level**

Among the more notable attempts to develop program budget procedures on a demonstration basis is the State and Local Finances Project, conducted at George Washington University. Supported by the Ford Foundation, the project is part


of a program of research and education on advanced fiscal planning. Participating in this "5-5-5 Project" are five states: California, Michigan, New York, Vermont, and Wisconsin; five counties: Dade (Florida); Davidson (Tenn.); Los Angeles (Calif.); Nassau (New York), and Wayne (Mich.); five cities: Dayton, Denver, Detroit, New Haven, and San Diego. The objective is to demonstrate the conceptual and operational feasibility of PPBS procedures for state and municipal governments.8 However, the stage of planning development varies greatly among these fifteen governmental units. Descriptive information, guidelines, and technical manuals are generally available upon request from each unit.

California and New York have made substantial progress in developing program budgeting procedures. In relating the approach (or Planning and Budgeting System [PABS]) as California has labeled it) to its many departments and agencies, the State admitted it had difficulties responding to the basic question "What is a program?" To answer this question, and to train department personnel in the intricacies of PABS, the State stressed ten principles:

1. Define the system and specific problems (needs);
2. Meaningfully describe objectives;
3. Quantify outputs (benefits);
4. Quantify inputs (costs);
5. Provide multi-year considerations;
6. Conduct analysis of alternatives;
7. Crossover of input data to program format;
8. Specify assumptions and uncertainties,
9. Record relationships; and
10. Conduct evaluation.

Since 1964, the State of New York has been developing a PPBS format that systematically relates the expenditure of State funds to the accomplishment of carefully conceived goals: In order to establish program planning on a recurring basis, an annual cycle was devised to provide for the evaluation and coordination of program plans in line with statewide goals and to help coordinate annual budget requests with a detailed projection of long-range needs. There are six major phases of the annual planning cycle: (1) organization phase; (2) agency report preparation; (3) Executive Office conferences with each agency; (4) program coordination and summarization; (5) integration of program plans and budget requests; and (6) budget development. The State was conceived as a system and it was differentiated according to nine central government level programs: business and industry, education, government affairs, health, housing and community renewal, personal safety, recreation and cultural enrichment, social development, and transportation. Each of these major parts of the total system is subdivided into a number of highly differentiated parts (or functions). Within each State agency, program activities are grouped into a set of program categories (major program areas), each of which relates to one of the minor functions. These program categories are not necessarily consistent with appropriation categories or with existing organizational structures. Most program categories contain one or more subordinate levels, called subcategories and program elements (components).

Local Level
In addition to the five cities listed above as participants in the "5-5-5 Project," a number of municipal governments are installing PPBS. To illustrate, New York City is utilizing PPBS and it has contracted with Rand Corporation to help analyze city problems in a number of key areas (police, fire, health, and housing). As the nation directs greater attention to urban affairs, it is probable that systems procedures will be employed to an even greater extent in city government. More than 5% of Rand's 1968 budget involved its New York City research project, and it is likely that this program will be expanded. A number of first steps were taken in 1967-68 toward the PPBS format by the City:

1. recruitment efforts to employ program analysts;
2. establishment of a Policy Planning Council;
3. analysis of the capital and operating budgets by objective;
4. review of projects by community area within the City;
5. establishment of a Division of Program Planning;
6. PPBS probes in selected key areas;
7. adoption of cost-effectiveness procedures;
8. installation of PPBS by the Board of Education;
9. development of information and scheduling systems; and
10. proposals for a think tank to analyze city problems.

An illustrative program structure for a city government that might wish to install PPBS is shown in Table 2. It encompasses eight program areas. This design was presented to a U.S. Senate subcommittee that was studying intergovernmental relations between and among the federal, state, and local levels.9

HARRY J. HARTLEY

Table 2
ILLUSTRATIVE PPBS PROGRAM STRUCTURE FOR THE GENERAL PROGRAMS OF A MUNICIPAL OR STATE GOVERNMENT

I. Personal Safety
   (Protection from personal harm and property loss)
II. Health
   (Physical and mental well-being)
III. Intellectual Development and Personal Enrichment
    (Education, libraries, museums, etc.)
IV. Satisfactory Home and Community Environment
    (Creation of a livable and pleasant environment)
V. Economic Satisfaction
    (Satisfactory individual work opportunities; welfare assistance)
VI. Leisure-time Opportunities
    (Indoor-outdoor recreational and cultural opportunities)
VII. Transportation—Communication—Location
     (Highways, urban transit, parking, air transport, etc.)
VIII. General Administration and Support
      (Activities not assignable to the other program areas)

The eight program areas represent general categories that must be differentiated in greater detail. For example, within General Program Area I, Personal Safety, there are eight parts, or programs:

1. law enforcement;
2. traffic safety;
3. fire prevention and control;
4. safety from animals;
5. protection from and control of disasters;
6. prevention of food, drug, and occupational hazards;
7. unassignable research and planning, personal safety, and
8. unassignable support, personal safety.

Furthermore, using the general systems notion of hierarchical organization, there are subprograms within each of these eight programs. For example, within Program I, Law Enforcement, there are five subprograms:

1. crime prevention;
2. crime investigation;
3. judging and assigning of punishment;
4. punishment and safekeeping of criminals; and
5. rehabilitation of criminals.

The logic of this procedure is clear. The first step is to identify the essential properties and programs of a given system. Then, more extensive differentiation of parts of the system is made in order to study the relationship of parts to wholes. Finally, the parts and their relations are integrated into a whole for which public policy may be made at any level of government—federal, state, or local.

LIMITATIONS OF SYSTEMIC APPROACHES TO POLICY-MAKING

General systems theory and the planning-programming-budgeting system can provide better means of decision-making for public institutions, but it is essential to note that the systems approach is not a panacea. Brief mention is made below of ten limitations of the systems approach in the context of public policy. The intent is to call for refinements and improvements in systems techniques rather than to provide a self-defeating detraction from general systems theory.

(1) Political elements may act as a barrier to systems procedures. Public policy is formulated in a political environment where distrust of economic rationality is widespread. The criticism of "McNamaraism" by the Congressional-military alliance was very great. One critic, Adm. Hyman Rickover, referred to PPBS as a religion of efficiency and "fog bomb" in testimony before Congress. He scoffed that, on a cost-effectiveness basis, the colonists would not have revolted against King George III and John Paul Jones would not have engaged an inferior ship against "The Serapis."

(2) Some critics argue that systems analysis can provide very little assistance in solving problems of social change. The 1968 Washington, D.C. forum on "Systems Analysis and Social Change," sponsored by the Operations Research Society of America, had a number of experts who suggested that a modern social "system" actually may not be a system. A City has little of the arrangement and almost none of the harmony that are necessary if a system is to be defined as a "harmonious arrangement or pattern." PPBS was successful when attached to the inertness of complex military machines, but it may be less appropriate for hydra-headed "living problems" of riot control and slum removal.

(3) The systems approach is not a "mathematical messiah" that can quantify and analyze the entire output of a social organization. It is neither a substitute for good management nor a remedy for organizations that do not have sufficient resources to achieve their objectives. PPBS is not put forth as a method destined to drive all the conventional planning methods into obscurity. However, persons who place too strong confidence in mathematical general systems theory may misuse PPBS by trying to subject to mathematical analysis systems so complex that they cannot possibly yield to such analysis.
PPBS: THE EMERGENCE OF A SYSTEMIC CONCEPT FOR PUBLIC GOVERNANCE

(4) Generic models should be altered to fit specific situations. Models and procedures formulated in one discipline are not always transferable to another field. One should be aware of the possible loss inherent in the adaptation process. Subclasses of systems possess a relative scope of applicability; thus policy-makers should be aware of what borrowed models can do and cannot do. The pathetic fallacy of the expert is to believe that his technical concepts and language can be used to explain the universe.

(5) The quantum jump of technology and science far transcends any comparable advance in human wisdom. A wisdom lag exists. We can analyze intricate contemporary problems with computers, but oftentimes we do not know what to do with our analyses. The tragedy of our era is that human intellectual capacity, as addressed to problems of human relationships, seems, if anything, regressive.

(6) Elaborate analysis may be based upon poor data or questionable premises. Analysts may be sophisticated in certain techniques, but not in estimating the value and relevance of data.

(7) Systems procedures may foster a centralizing bias that reduces individual freedom and privacy. As government units have grown and increased their complexities with data banks, information centers, and other computerized devices, policy-making has become much more centralized within a tightly defined chain of command. The distance between the leader and the led has been increased, thus reducing the individual's democratic rights of decision, dissent, and deviation. How to balance the advantages of efficiency obtained from centralized decision-making against the human survival values of individual decision-making at the "point of stress" is a basic problem of public governance in our time.

(8) Innovative systems procedures such as PPBS are "doomed to success" in many organizations. The reason is that evaluation of newly installed techniques is often conducted by the same persons who introduced the innovations. It is unlikely that such officials will claim that systemic thinking is anything less than a smashing success. More definitive verificational research evidence is needed to determine whether PPBS has, in actual experience, fulfilled its promise. Most of the literature describing systems concepts deals with their potential value, but does not validate it.

(9) Most government units presently have inadequate staffs and insufficient resources for comprehensive systems planning. Deficiencies exist in the training programs of administrators, in the number of personnel, and in the usage of electronic data processing. The result has been that the approach of a number of units, such as New York City, has been "...deliberately opportunistic, rather than systematic and comprehensive, with a focus on sectors of high apparent yield."12

(10) In the application of systems procedures to formal organizations, there appears to be a strong tendency to measure only that which is most easily measured. One organizational theorist asserts that this can have very undesired effects from the standpoint of organizational goals. "Frequent measuring can distort the organizational efforts because as a rule, some aspects of its output are more measurable than others. Frequent measuring tends to encourage over-production of highly measurable items and neglect of less measurable ones."13 The result is "goal distortion."

PLANNING SOCIAL CHANGE*

D. W. Griesinger and C. G. McClintock

1. HUMAN POVERTY AND SOCIAL CHANGE PROGRAMS

Introduction

There is growing concern about the seemingly inextricable plight of the nation's poor. Major domestic programs are being considered and some already have been implemented to help the poor and to improve the overall quality of life in America. Therefore important questions abound with respect to planning for social change—What are the dimensions of poverty? What are the side effects of governmental intervention in the lives of the poor? What is the relationship between poverty and unrest? How can social and psychological outcomes of social change programs be assessed and compared?

The relevance of many of these questions has already been painfully demonstrated by inadequate consideration of the social and psychological impact of urban renewal programs in several of our major cities. However, recent attacks on problems of urban life have benefited from past experience. As a result, increasing emphasis is being placed on citizen participation throughout the planning process. This is particularly evident in the Model Cities Program, as indicated by H. Ralph Taylor** when he observed:

A relationship must be worked out in which City Hall and the neighborhood people recognize their respective needs, and agree to negotiate their differences. Lacking this, a plan to meet citizen participation standards will not be developed, the overall Model Cities plan will not be workable, and the city will, therefore, be ineligible for program execution funds.

It is not at all clear, however, that citizen participation alone is sufficient to assure acceptable social and psychological outcomes of a massive poverty relief program. Unless an adequate data base and means for assessing its availability, and unless some reliable notions of cause and effect are employed, endless debate or ill-founded choice can lead to further frustrations in our cities, at great expense to our society.

Many of these issues, already raised by the Model Cities Program, will continue to be basic to anti-poverty programs for years to come. The thrust of this report is to establish a conceptual orientation and methodological procedure for assessing the many dimensions of poverty, for establishing objectives for social change, and for relating potential social and psychological outcomes to program alternatives.

In the remainder of this section, we will discuss economic and cultural dimensions of poverty and identify broad objectives for social change programs. In Section 2 we attempt to characterize the informational and measurement procedures required to assess poverty, and to set forth some of the requirements for a decision model for social change programs. In Section 3 a theoretical orientation is established and a social dynamics model is presented, together with examples of its applicability to planning problems encountered in an anti-poverty program. A more formal characterization of the social dynamics model can be found in the Appendix.

Economic Objectives

Poverty, both historically and currently, has been most frequently defined in economic terms as a discrepancy between one's current economic resources, regardless of their source, and those resources required to procure a minimal standard of living. A minimal standard of living, in turn, is generally defined in terms of survival, in terms of achieving some minimal level of nutritional support and limited access to other commodities and services requisite to maintain life. Through recent history, what has changed is not so much this definition of poverty, but what is considered to be a minimal economic standard for sustaining life.

The pervasive utilization of an economic definition of poverty, and the establishment of programs to reduce it, derive primarily from two considerations. First, the overt effects of economic poverty, particularly on children, are easily discernible—the debilitating effects of malnutrition and its correlated physical diseases and high death rates. Hence this form of poverty cannot be easily hidden or rationalized. Second, one can obtain a relatively simple quantitative index of economic poverty that enables one not only to assess the current status of an individual, a family, a group, or a society, but also to measure improvements that attend various welfare, economic, educational,
or other remedial programs. In effect, one can establish more or less rigorous standards of what is and what is not poverty by utilizing quantifiable economic indices, and then justify programs of change on the basis of anticipated and measurable changes on these indices. For example, the President in his Economic Report of 1964 emphasized the magnitude and severity of poverty in this country in such economic terms:

Americans today enjoy the highest standard of living in the history of mankind. But for nearly a fifth of our fellow citizens, this is a hollow achievement. The forgotten live without hope, below minimum standards of decency. The per capita money income of these 35 million men, women and children was only $590 in 1962—against $1,900 per capita for the Nation as a whole. We cannot and need not wait for the gradual growth of the economy to lift this forgotten fifth of our Nation above the poverty line. We know what must be done, and this Nation of abundance can surely afford to do it (pp. 14-15).

Given the general social objective of reducing economic poverty, a great variety of programs with more or less limited objectives have been initiated within our society. To merely list such programs would be a major task, much less to describe their specific objectives. One can, however, characterize the objectives of most of these programs in terms of three fundamental strategies summarized in Table 1.

Social and Psychological Objectives

A more recent view is that poverty also includes other major dimensions of living, and that a minimal standard of living implies more than mere physical and economic survival. This change in definition is in part dictated by three factors: the growing recognition that 1) the quality of life is not solely determined by economic factors, 2) people cannot be motivated to change their behavior solely by economic incentives, and 3) that economic solutions to poverty do not automatically reduce many of the social ills of the society.

Kenneth Galbraith, the economist, in defining insular poverty, cites the importance of several non-economic factors in determining the behavior of those who reside in economically impoverished areas (1):

Insular poverty has something to do with the desire of a comparatively large number of people to spend their lives at or near the place of birth. This homing instinct causes them to bar the solution, always open as an individual remedy in a country without barriers to emigration, to escape the island of poverty in which they were born. . . . In some circumstances escape may not be possible. Especially in the urban slum, race or poverty may confine individuals to an area of intrinsically limited opportunity. And once again the environment perpetuates its handicaps through poor schools, evil neighborhood influence, and bad preparation for life. The most certain thing about modern poverty is that it is not efficiently remedied by a general and tolerably well-distributed advance in income. Case poverty is not remedied because the specific individual inadequacy precludes employment and participation in the general advance. Insular poverty is not directly alleviated because the advance does not necessarily remove the specific frustrations of environment to which the people of these islands are subject.

It is only one step further to the observation that there exists to a greater or lesser degree a culture of poverty that socializes its own young in

<table>
<thead>
<tr>
<th>Table 1</th>
<th>STRATEGIES FOR ECONOMIC CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRATEGY</td>
<td>PROGRAMS</td>
</tr>
<tr>
<td>1. Support the individual economically</td>
<td>Welfare, Aid to Disabled, Community Health, etc.</td>
</tr>
<tr>
<td>2. Change the individual’s economic potential</td>
<td>Training and skill development, remedial medical, etc.</td>
</tr>
<tr>
<td>3. Change the individual’s economic opportunity structure</td>
<td>Federal incentives to private industry to utilize economically deprived, direct Federal employment of economically disadvantaged, etc.</td>
</tr>
<tr>
<td>a. Create economic opportunities</td>
<td>Fair employment legislation for minorities, women, etc.</td>
</tr>
</tbody>
</table>

654
PLANNING SOCIAL CHANGE

a manner that perpetuates the culture. The products of this socialization experience hold to values and patterns of behavior that are not necessarily consonant with those of the larger society. Children within this culture learn its norms and its values. Life becomes a matter of "marginal" survival through avenues that the ghetto social system defines as legitimate—which may indeed include activities inimical to the major society, such as crime, prostitution, welfare fraud, etc. It seems evident from those who have lived in and reported on this culture that satisfactions of the moment often take precedence over the uncertainties of delayed rewards; that consumption is emphasized, whereas production or work values are not; that schools are irregularly attended, and that many members are relatively unattracted by the incentives (often inadequate, to be sure) that society offers them for their own economic advantage.

In a very fundamental sense, the Model Neighborhood Program is revolutionary in its recognition of the complex determination of insular poverty and poverty cultures. In establishing the overall social objective of "improving the quality of urban life," it implicitly recognizes that the problems of the poverty stricken—the elderly, the poorly educated, the minorities, the handicapped—must be met initially in their indigenous social framework. Any program aimed at eradicating poverty rather than merely temporarily alleviating it must operate among the poor; it must, in effect, foster and support the emergence of a different culture whose goals are more consonant with those of the major society. Such changes must occur not only within the ghetto culture, but also within the major society, for the latter plays a substantial role in defining the opportunities and constraints operating upon the ghetto dweller.

Changing a culture implies more than sending in various teams of outsiders to guarantee survival or better housing or swings for the playground. It implies more than the traditional economic programs outlined previously that make individuals their principal target. It requires that the local community and its inhabitants become a part of the planning and decision process itself. Changes in human values and behavior follow from the recognition of problems, feelings of dissatisfaction with the status quo, opportunities to express this dissatisfaction in meaningful and constructive ways, and active learning and participation in change. Culture changes involve more than individual change; they imply changes in human institutions, norms, and complex patterns of human interaction. Hence, comprehensive poverty reduction programs must take as a major objective not only alleviating economic poverty, but also modifying those social, political, and economic processes—both within the ghetto and within the impinging major society—that give rise to and perpetuate a culture of poverty.

In terms of establishing social programs that successfully produce cultural change, we are confronted with difficult scientific, pragmatic, and political problems. There is no agreement on which variables or indices are useful to measure culture deprivation as distinct from economic deprivation, and it is unlikely that simple measurement techniques exist for assessing these indices once defined. Furthermore, the relationship is unclear between various proposed programs for culture change and their potential effects. Additionally, we are confronted with decisions whose implementation will undoubtedly result in a redistribution of political and social power within and between various segments of our society, and that consequently will generate resistance, if not overt conflict. The government, in effect, will be involved in the change of aspirations, norms, and beliefs of a significant sector of our society—and we lack both the conceptual orientation and the measurement procedures for defining, assessing, and controlling this process. We are faced with specifying and measuring social objectives such as community involvement, cultural norms, and political participation, as well as psychological states relating to aspirations, levels of dissatisfaction with one's physical and social environment, expectations concerning future outcomes, etc. It is only through understanding and measuring such processes that one can hope to design and implement programs to eliminate the poverty cultures currently operating in this country.

Understanding, defining, and measuring poverty in economic terms is essential to this process, to be sure, but it is not sufficient. For culture change emerges through modification of norms and values, which in turn depend upon the active, concerned, and goal-oriented behavior of the members of the target culture—behavior that is not totally determinable by economic incentives. Finally, changes in poverty cultures occur in juxtaposition to a majority culture that is not indifferent to or independent of the costs or the effects of such changes. In effect, the majority culture places restraints on change that are mediated by the various "gate-keepers" who have one foot in poverty culture and one in the majority, e.g., employers, social workers, politicians, teachers, landlords, real estate personnel, government employees, doctors, and union personnel. To understand the forces for and against change in poverty cultures, it will be necessary to assess the degree of influence that these representatives of the major society exercise over the change process. One major social objective will indeed be to find methods for assessing, modifying, and controlling the constraints that these groups exert for and against change within the poverty culture.
2. THE DECISION PROCESS

Information Requirements

It is not our intention in this paper to specify all of the types of data that are requisite for assessing the current status of a poverty culture and its inhabitants. However, such data would include an estimate of the current physical status of the neighborhood area, including the condition of housing, the schools and other buildings, the streets, and the transportation systems, if any. It would require information concerning the economic status of the community and its members, including conditions of employment, family income levels, welfare payments, rental costs, capital investment in the community, and tax structure. It would require information regarding the social organization and disorganization of the community as related to such areas as family structure, recreational patterns, crime and delinquency, ethnic and racial groupings, and political structure and power.

Furthermore, detailed data would be necessary to characterize 1) the types of services provided citizens under existing welfare, legal, medical, recreational, and educational programs; 2) the problems to which services are directed; 3) their effectiveness, and 4) the degree of current coordination involved in meeting these problems. In addition, information would be required concerning the economic, social, political, and psychological potentials of the neighborhood, its inhabitants, the various service agencies and other groups that currently affect the operation of the neighborhood. Some measure of the morale of the community would be requisite—feelings of satisfaction and dissatisfaction with specific issues and activities, feelings of hope and despair with one's physical and social milieu, the presence or likely emergence of activist groups attempting to foster change in behalf of various segments of the community, etc. Finally, one would have to assess the constraints imposed on the ghetto community by the "gatekeepers."

The collection and organization of such data is an immense task, and to accomplish it effectively would require a prior conceptual orientation concerning what data is relevant and why. This, in turn, requires some fundamental understanding of the types of data that are relevant, given some preconceived goals of a development program. Finally, the information requirements outlined above relate only to assessing the current status of the neighborhood. There are also a number of additional informational requirements that attend subsequent stages of the decision process.

Information Procedures

To obtain an adequate representation of the current status of a community, a number of techniques would necessarily be required. First, demographic data derived from the national census or specifically designed local censuses should be obtained. Furthermore, the records of the various actors in the community, e.g., welfare agencies, hospitals, and law enforcement and zoning agencies, would contain relevant data. Persons familiar with various aspects of the community and its current status could be sampled for their opinions when more objective data is unavailable. Also, sample surveys of neighborhood residents could be obtained to estimate perceived or experienced areas of strengths and weaknesses, neighborhood morale, etc.

Such an effort would require considerable sophistication regarding the availability of data, methods for assessing its reliability, and methods for obtaining new data. This would include not only determining what data are relevant, but also procedures relating to collecting new data, selecting appropriate samples, and storing and retrieving information. Furthermore, there are many problems to resolve and techniques to be understood if it is necessary to combine various data to form indices; for example, a general index of quality of housing.

In effect, even in the initial phases of data collection, one is confronted with the need for conceptual skills and research methodologies to which one may not have immediate access, including problems in survey methodology, the use of computers for data storage, retrieval, and index construction, the development of questionnaires, interviewing procedures, etc. It is perhaps even more critical that one may not be aware of what are the most relevant types of data necessary to obtain the most precise and valuable estimate of the current status of the neighborhood. Such deficiencies unquestionably exist in the various model neighborhood-planning grant applications that already have been submitted to the U.S. Department of Housing and Urban Development (HUD).

Goals and Priorities

Examples of the types of general goals that a program aimed at reducing human poverty should consider have been specified by HUD in defining its Model Cities Program. These are outlined elsewhere (2). We will not reiterate these here, but will attempt to characterize some of the problems that attend translating such general goals into more specific operational ones.

One of the major problems is to redefine a general goal such that one can specify in detail the end states desired. For example, if we acknowledge that a program should improve education, it is incumbent upon the decision maker to define in an operational way what he means by improved education: e.g., more technical skills, more abstract knowledge, higher scores on achievement tests, or greater social adjustment. If one is going to include a number of criteria in defining a better education, then it is necessary to weight the
importance of the criteria, since one may be faced with alternative programs that more or less adequately meet these various objectives. Finally, in a comprehensive program, relative importance between major objectives is also a consideration, since all goals cannot be equally met given limitations in financial and human resources.

There exists of course the difficult problem of deciding who defines the goals and the priorities—the national government, the political agency of the area, the various public and private agencies whose programs serve the community, or the community itself? Even if one could identify the goals and the priorities of the organizational units that impinge upon or form a neighborhood, how would they weigh these? It is evident that this is the major task of any planning group, subject to the political realities of the environment in which it operates. This indeed involves the processes of political bargaining, influence procedures, and other considerations in the negotiation of agreements between parties with varying goals and orientations.

Costs and Benefits

If indeed one has a set of objectives that can be operationalized, then one is confronted with the task of considering alternative procedures in meeting these goals, taking into consideration the costs and benefits (utility) and the likelihood of success (probability) of each alternative. In the conceptual jargon of decision theory, rational decisions require that one select that alternative which maximizes one's expected outcome, the latter being some function of one's utility or goals and probability estimates of success. The decision maker then selects that alternative or combination of alternatives which brings the highest expected rewards.

Thus any decision takes into consideration projected costs and benefits. The advantage of formalizing or systemizing this process lies in the identification of a wider range of potential alternatives, a closer scrutiny of the rewards and costs of particular alternatives, a more precise consideration of whose costs and benefits one is affecting, and a more comprehensive review of possible secondary effects upon neighboring systems.

It is not necessary to consider in detail here the many problems that would confront planners attempting to attach benefits and costs to alternative programs designed to meet operationally defined objectives. However, we will list a few of the more fundamental issues that confront a decision-maker charged with the task of coordinating, initiating, and implementing a variety of community service programs. Prior to this, we note that there is an additional payoff to increasing the cost-benefit effectiveness of decisions based upon a more or less systematic analysis; that is, such analyses help to assure that estimated expenditure of resources (dollars, time, services, etc.) remain within a specified budget.

An enumeration of some of the major problems involved in a cost-benefit analysis of alternative programs to reduce poverty would necessarily include the following:

How does one obtain estimates of costs and benefits of activities that in part can be judged in terms of market value, e.g., housing construction, disease prevention, and highway relocation?

How does one assign benefits and costs to fundamental social phenomena such as aesthetic appreciation, cultural education, political participation, psychological well-being and happiness, and other non-market values?

Who should be assigned responsibility for defining market and non-market values as regards what values are relevant, how they are to be scaled, and what utilities are to be attached to various alternative programs?

How does one deal with data that range from that which is easily quantifiable and more or less objectively defined (dollars per square foot of construction cost) to that which is difficult to quantify and involves estimates of subjective human status (happiness)?

It seems indeed unlikely that a strictly quantitative and objective approach is possible in assigning costs and benefits to alternative community services. However, the consideration of both market and non-market values is of great importance; an attempt to assign costs and benefits to alternatives in terms of these values forces the decision maker to confront fundamental issues, to recognize that, in community welfare and planning, there are few absolutes; to systematically consider and utilize the consumers of programs as sources of information regarding their utility structure, and hence to make decisions based upon better estimates of both the physical and social environment.

Decision Making—the Art of Compromise

Thus far, we have described the decision process as one of defining goals, gathering relevant information, assigning costs and benefits to alternative programs, and selecting that alternative with the highest expected payoff. This is an adequate descriptive model for a single decision maker, or of a group of decision makers with more or less common values, operating upon the environment. However, it is a gross misrepresentation of the task confronting those involved in culture change. The latter are confronted with a situation in which there are multiple actors with multiple values who attach differing weights to various goals and subgoals, who vary in their estimates of
the costs and benefits associated with various alternative programs, and who possess more or less political, economic, and social power relative to facilitating or impeding the functions of a planning group. It must be recognized that at the local level the list of relevant actors in the negotiating game could include the many subagencies of City Hall, and the many private and public groups functioning in the neighborhood, whose participation would be essential to the implementation of various programs.

Thus, any planning agency is faced with evaluating goals, subgoals, costs, and benefits of the various groups who are directly or indirectly involved, and selecting those alternatives for proposed implementation that provide the most satisfaction to the most actors. This is not a simple optimizing problem faced by an individual decision maker operating upon the physical environment. Rather, it is a complex process in which a great number of social-political variables must be considered.

In effect, we are confronted with a multi-actor game in which various actors assign varying costs and benefits to various alternatives. The process of obtaining estimates of the "utilities" for the relevant actors and program alternatives, and of providing mechanisms for negotiation and compromise between actors, is undoubtedly the most critical problem to be confronted. To complicate the matter even further, one must recognize that the information base, the identification of goals, the number and kinds of actors, and the decision rules may change throughout the planning process, and throughout the various stages of implementation.

### Decision Making—A Preliminary Schematic

For any social change program initiated by government, some individual, agency, or organization will be charged with the execution of the planning function. This agency will be responsible for integrating the values and resources of the government and community organizations with those of neighborhood groups in such a way that goals and aspirations, needs, and values can be included in renewal plans for all participants.

Although each alternative for renewal will likely bring hardship to some and benefit to others, we emphasize that a primary purpose of the planning function is to seek out the costs and benefits accruing to all affected parties, not only as they perceive them but as the facts portray them over a spectrum of cost dimensions, including economic, political, and social factors. Without an adequate assessment of social costs and benefits, a decision model is not functional, and even informal decision processes will tend to over-emphasize economic and physical factors, or bog down in time-consuming delays and endless arguments.

As more and more facts are brought to bear on any one alternative, perceptions of the costs and rewards will also change. If objective measures are available, there may result a more or less unanimous conception of the consequences of a proposed action. Such is the case in the economic dimension when a contractor submits a bid to build so many housing units for so many dollars. In the social dimension, measures may be far more subjective; unanimity, even among those affected, may be difficult to achieve. Part of the planning function, then, is to achieve some consensus on the subjective consequences of a particular alternative.

Alternatives are of two types: alternative objectives and alternative means to attain a given objective. For any particular renewal objective, a principal goal of the planning agency can be represented schematically in tabular form as shown in Table 2. Opposite each alternative means to the renewal goal are listed the associated costs and benefits, in each dimension of experience, for the local government and for each affected group. Conceptually, the planning goal is agreement on the alternatives and on the costs and benefits that should be entered in such a table. Ideally, the finally approved renewal plan would be drawn from among the alternatives a, b, ..., ab, ..., abc, ..., etc. The ultimate choice would rest upon the finally agreed upon costs, c, and benefits, b, for each

<table>
<thead>
<tr>
<th>Renewal Alternative</th>
<th>City Costs</th>
<th>City Benefits</th>
<th>Each Relevant Group Costs</th>
<th>Each Relevant Group Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>a, b, c, ab, ac, abc</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 2

**TABLE OF ALTERNATIVES**
participating group and value dimension; i.e., $c_{ai}$, $b_{ai}$; $c_{ai}$, $b_{ai}$; ..., etc., where the subscripts $a$, $b$, ... refer to the renewal alternative, the subscript $i$ refers to the value dimension (social, economic, etc.), and the primes distinguish among relevant groups. A fully successful planning stage should result in value entries in the matrix of alternatives that are mutually agreed upon by all major participants, at least for all competitive alternatives. This agreement may result from the invocation of objective criteria, or evolve through persuasion, negotiation, compromise, or mediation. We emphasize that objective criteria are currently available only in a few value dimensions; in the other dimensions, these presently can be resolved only by interaction among the principal participants, perhaps in a bargaining situation. If the parties refuse to interact, the conflict may not be resolved.

Theoretical and Methodological Requirements

If one has a number of broad social objectives, and a set of decision rules concerning how to select programs to achieve these objectives, then several additional theoretical and methodological issues follow. For example, the mere activity of selecting problem areas to enter into a cost-benefit matrix implies that one has a theoretical rationale for identifying and measuring their relevance and hence inclusion. The attachment of rewards and costs to various alternative means of attacking a problem area implies that one possesses some set of more or less formalized hypotheses concerning the relationship between a given alternative and the changes that will follow. The comparison of rewards and costs accruing from some limited set of alternatives implies that one has the capability of consistently measuring and scaling both the forces one is proposing to manipulate and the outcomes of the manipulation.

Unfortunately, within the social sciences there do not exist logically and empirically validated theories of social organization and human behavior sufficient to provide criteria for selecting the appropriate forces to change, for measuring and projecting the costs and benefits that would accrue from such change, for assessing the political feasibility of programs, or for measuring the obtained effects upon the quality of life of urban residents. This does not mean that we are without information, that there do not exist sets of hypotheses or educated hunches concerning possible sets of relations between forces operating within an urban context and their outcomes. But, without a well developed theory, it behooves us to use the more promising of these hypotheses or hunches, always working toward a more comprehensive and valid theory and continuously sharpening our measurement procedures. For we are confronted with the political and social reality that, regardless of our state of knowledge, we must initiate massive programs of change as urban conditions become increasingly less tolerable for the residents of these areas, and for the society as a whole.

A Theoretical Orientation

The remainder of this paper is concerned with the study of techniques for defining, evaluating, and comparing the potential social and psychological outcomes of renewal alternatives, in order to assure that social factors are adequately accounted for in the decision process. For example, important cost measures in the social dimension are, on the one hand, individual and group dissatisfaction, anxiety, and alienation, and, on the other hand, initiative, effort and other human resources. Social and psychological benefits are measured in terms of satisfaction, fulfillment, joy, ease, and contentment.

The model to be presented in the next section relates social and ecological forces to human behavior, effort, and discontent. Because each renewal alternative alters the social and ecological forces on each group of participants in different ways, the effort required, the discontent experienced, or the satisfactions achieved by each group of affected individuals will differ. These social costs and benefits, we believe, can be estimated from a combination of social field data and the model of social dynamics to be outlined in the next section.

3. A MODEL OF SOCIAL DYNAMICS

Introduction

The variables and principal constructs of the social dynamics model fall into three classes: behavioral, emotional, and environmental. The model formally relates these factors, providing a means of assessing the implications of environmental change—social or physical—on changes in the behavior and sentiment of the affected population.

As a conceptual introduction to some of the important features of the model, consider the following example. Imagine a man in a mountain valley as shown in Figure 1. Suppose we measure the man's position by the coordinate $\sigma$. When he is at the valley floor, his position is $\sigma_0$; when he visits the lake over the next ridge, he will be in the vicinity of $\sigma_0$. If the effort he must expend to reach a position $\sigma$ is measured by the height of the mountain at that point above the valley floor, then it is clear that the position of least effort for the man is at $\sigma_0$. In the absence of other considerations, it seems likely that he will remain near $\sigma_0$ because the forces on him there are minimal. If, however, other circumstances motivate him, such as thirst and the knowledge of the lake at $\sigma_l$, he may exert the effort necessary to climb over the mountain to obtain a drink.

Over a period of time many men may visit the valley. Suppose that an observer counts the density of footprints on the valley floor and along
Fig. 1. Position as a function of effort in a mountain valley.

Fig. 2. Probability for being at a given point along the trail.

the trail leading from the valley up the mountains on either side and over to the lake. These footprints, made by many persons and representing a variety of states of motivation, will constitute a frequency distribution for persons reaching various positions \( \sigma \). From this data can be constructed a probability density function \( P(\sigma) \), such as that shown in Figure 2, giving the probability that a person on the trail will be found in the vicinity of point \( \sigma \). This function has the general feature that the probability is maximum in the regions requiring least effort; nevertheless there is some probability for climbing up the steep trail and over to the lake. This probability density is derived from the behavior of men in various motivational states. However, one can imagine subgroups of men with approximately the same motivational energy. For each subgroup of \( n_i \) individuals in the same motivational state, there exists a corresponding probability density function \( P_i(\sigma) \). The resultant \( P(\sigma) \), which is the observed distribution for the entire group, is made up of the sum of the \( P_i(\sigma) \) weighted by the relative population of each subgroup; thus,

\[
P(\sigma) = \frac{n_0}{n} P_0(\sigma) + \frac{n_1}{n} P_1(\sigma) + \frac{n_2}{n} P_2(\sigma) + \ldots + \frac{n_k}{n} P_k(\sigma)
\]

where \( n_i \) is the population of motivational state \( i \), for \( i = 0, 1, 2, \ldots, k \), and \( n = \sum_i n_i \) is the total population.

The three classes of constructs of the model are now evident. Behavioral variables, \( \sigma_j \), measure performance in the \( j \)-th activity or behavioral dimension. In the present example, we have considered only one dimension in which \( \sigma \) measures the location of a man on the mountain path. However, the model is not primarily concerned with individual behavior, but rather with the probability density function describing the distribution of behavior for a group of people.

The probability distribution, \( P(\sigma_i) \), in the \( j \)-th activity depends on the environmental force variables—the height of the mountain or the steepness of the grade. In general, the environment is represented in terms of the effort or energy expenditure required by an individual to achieve a behavioral position \( \sigma \). The probability distribution also depends on motivational or emotional variables, in particular, on the number of individuals in various states of discontent within the group population (various states of thirst in the present example).

In this example, the barrier to movement is physical (the mountain), but the idea of constraint is extendible to various social dimensions. Indeed, social forces can seem like mountains, and personal goals analogous to lakes on the other side. Without sufficient discontent with the present state, the energy that must be expended by an individual to reach a goal is more than he is likely to have available. Unless the barriers to his progress are reduced, he will not change his position.

A primary objective for poverty relief programs is the deliberate removal of barriers to social progress for people who have found themselves inextricably trapped in poverty states by a variety of social forces. The alternative to planned change is to rely on the discontent of the poor to force the necessary gains.

Social Forces

Social forces are of many kinds and operate in a variety of behavioral dimensions. They consist of the opportunity structure for receiving a given level or quality of education, including factors such as patterns of educational discrimination, adequacy of physical facilities, and quality of instruction; they reflect the physical availability of adequate low-cost housing; they reflect the match or mis-match between available jobs and the skills of the unemployed; or they involve patterns of social disapproval directed toward persons who deviate from ghetto norms. But, whatever the source of the force, it is represented in the model by a function \( S(\sigma) \) that, in analogy with the height of the mountain above the valley floor, represents a cost in terms of the energy (or effort) that an individual must expend to achieve the behavioral level \( \sigma \).

*This simple example neglects the interaction of persons with each other, e.g. imitative behavior, etc. The model, however, can treat the more general case.
PLANNING SOCIAL CHANGE

We have chosen a formal analogue from quantum physics to define social forces in terms of observable behavior distributions, and to prescribe procedures for predicting the resultant changes in behavior when these forces are manipulated. The rationale for making this choice is contained in (3) and (4) and will not be pursued here. For the convenience of the reader, formal details of the model are outlined in the Appendix so that we may proceed here unencumbered by mathematical detail.

Dimensions of particular relevance to the poverty relief programs are those in which social and psychological poverty are measured, e.g., quality of housing, range or quality of health services, socioeconomic status, annual income, and educational attainment. Across the range of values of the variables assumed on each of these dimensions, the status of a target population and the goals of re-form can be represented by probability distributions. The measurable $P(o)$, then, is the distribution of the target population on the variable $o$. For instance, $P(o)$ might be the annual income distribution per family in an area of interest.

From $P(o)$, the model describes the procedure for obtaining the force function $S(o)$. We will refer to $S(o)$ as the "force potential" with respect to the behavior or performance variable $o$. A simple example of the results of this procedure is illustrated in Figures 3 and 4. Figure 3 is a probability density function derived from 1967 annual income data for negro males, ages 25 to 34. Figure 4 is the potential $S(o)$ (to within an additive and multiplicative constant) derived from the income data according to the prescription contained in the Appendix. (The curve shown in Figure 4 is actually $\frac{1}{k} (S(o) - \lambda_0)$ where $k$ and $\lambda_0$ are constants.)

---

Fig. 3. 1967 income distribution for U.S. Negro males, ages 25 to 34.

Fig. 4. 1967 income opportunity potential for U.S. Negro males, ages 25 to 34.

Fig. 5. 1967 income distribution for U.S. White males, ages 25 to 34.

Fig. 6. 1967 income opportunity potential for U.S. White males, ages 25 to 34.
In contrast, it is interesting to compare the force curve in Figure 4 with 1967 census data for white males as shown in Figures 5 and 6. Two prominent features are evident in comparing Figures 4 and 6, namely the displacement of the minimum in $S(\sigma)$ to higher $\sigma$ for whites (i.e., the most probable income is higher for whites) and the relatively steeper slopes on the $S$-curve for the black population (i.e., the force is more constraining). The significance of these features with respect to discontent is appreciable, as we will now discuss.

Discontent

Individuals subject to the same social forces can experience various amounts of discontent. At least conceptually, a population can be subdivided into relatively homogeneous subgroups with respect to the amount of discontent experienced. To each subgroup there exists a probability distribution, $P_i(\sigma)$, where $i = 0, 1, 2, \ldots$ identifies the subgroup. The model predicts the existence of a set of specific levels of discontent $\lambda_i$ that result from the imposition of a force potential $S(\sigma)$. In addition, the model predicts the corresponding probability distributions $P_i(\sigma)$ (see Appendix). The distribution of the population over the various levels of discontent is not currently available from the model but may be sought empirically.

If we arrange the $\lambda_i$ for a given potential in ascending order, we can draw what will be called an "energy level diagram," as shown in Figure 7.

![Energy Level Diagram](image)

**Fig. 7.** An energy level diagram.

By superimposing the potential on this diagram in such a way that $S = 0$ at $\sigma_0$, one can illustrate that the greater the state of discontent, the greater is the likelihood that a person will deviate from $\sigma_0$; i.e., the greater the magnitude of $\lambda_i$, the higher up the mountain one might climb. If the potential $S$ were to represent a pattern of sanctions with respect to deviation from a community norm, the model would say that the greater the unrest or discontent, the greater would be the probability that behavior is deviant from the norm.

Residual Discontent and Quality of Life. A person subject to social constraints is rarely completely satisfied or content. The state of lowest discontent in a given social context will be called the "residual discontent," the magnitude of which is assumed to be a function of the social constraints. In the model, the residual discontent is measured by the magnitude of $\lambda_0$ relative to the minimum value of $S(\sigma)$, as shown in Figure 7. In general, the more constraining are the social forces, the greater is the residual discontent. We propose to employ the magnitude of the residual discontent as a measure of the "quality of life" for social change alternatives.

For example, if we follow the procedure outlined in the Appendix, we can compare the residual discontent for the white and negro populations shown in Figures 3 and 5. The result indicates that there is almost three times more residual discontent deriving from economic factors among the negro males than for the corresponding group of white males.

Riots and the Venting of Discontent. Consider the income data for the 1965 Detroit riot area population (5). The income distribution for non-rioters living in the riot area is shown in Figure 8. This distribution can be fit reasonably well for incomes greater than about $4,000 by a normal distribution curve with a mean of about $6,500. Because it represents non-rioters, it is assumed that this distribution represents the performance of individuals who are predominantly in a state of residual discontent.

The model predicts that forces leading to normally distributed (Gaussian) behavior for the persons in the lowest state of discontent, $\lambda_0$, lead to bimodal distributions for individuals in the first excited state, $\lambda_1$ (see Figure 14 in the Appendix). It is therefore interesting to compare the income distribution for the Detroit rioters with the non-rioters in the same area. This comparison is shown in Figure 9. The striking bimodal feature of the distribution for rioters suggests the interpretation that rioters had, indeed, been in the first level of discontent.

The data does not support the view, however, that rioters and non-rioters perceived identically the same forces. While both distributions correspond to linear restoring forces centering on about $6,500 per year, calculations based on this data suggest that the rioters perceived the forces on them to be as much as four times greater than the forces perceived by non-rioters. This disparity is not particularly discomfiting in view of the marked differences in the composition of the two groups—the most probable age for a rioter was under 25, and for a non-rioter was over 35. These differences cannot be resolved without access to a wider range of data.
PLANNING SOCIAL CHANGE

The model, therefore, permits a description of riot behavior as the "venting of discontent," the amount of discontent available for release per individual rioter being given by the difference in energy between a rioter's pre-riot discontent and the next lower discontent level possible; in the Detroit case, $\Delta_l = \lambda_1 - \lambda_0$.

Persons in the lowest state of discontent (that is, in the residual state) are, in general, nonparticipants in the riot because the residual discontent is not ventable. There are no states of discontent lower than the residual state; thus it is not possible to vent the emotional energy of residual discontent except by actually reducing the environmental constraints. This is illustrated in Figure 10, which shows two potentials and their corresponding energy levels. Notice that, as the constraints are relaxed, both the residual discontent, $\lambda_0$, and the ventable discontent, $\Delta_l$, are reduced in this example. Thus, in addition to improving the perceived quality of life through the reduction of residual discontent, proper manipulation of the forces on a population can also reduce the amount of discontent per person available for destructive activity.

Consider, for example, the 1967 economic opportunity potentials for blacks and whites presented earlier in Figures 4 and 6. If potentials similar to these were to represent the conditions in a city that was engaged in anti-poverty planning, we would be free to ask what the residual discontent is for the city's blacks relative to the whites.

Using techniques discussed in the Appendix, we find that the residual discontent for the blacks is about 2.5 times that for the whites (that is, $\lambda_0$ black/$\lambda_0$ white $\approx 2.5$). The energy level separation in the two cases is such that the ratio of ventable discontent for blacks relative to whites is also of the order of 2.5.

The city can reduce both the ventable discontent (per person) and the residual discontent by adopting programs that would change the economic constraints on the blacks to conform more closely to those experienced by the whites. As the potential for the blacks approaches congruence with that for the whites, the ratios of discontent obviously approach one. It is also possible, however, to adopt a change program that will increase ventable discontent while reducing residual discontent. Such an undesirable option can be treated equally well by the model.

Formulating Goals and Evaluating Progress

One of the first jobs for an agency preparing a plan for change is to determine the relevant dimensions of social, economic, and political poverty in the target area. In effect, the planner is concerned with identifying and assessing those aspects or dimensions of human existence that are candidates for modification through alternative programs.
For one population the dimensions may be housing, education, and income; for another, crime and unemployment; for yet another, transportation, rats, and health services. The planner must also determine an order of priorities, establishing the urgency for remedies in each dimension for each target population, these priorities being closely related to the explosiveness of the situation in each case.

The problem dimensions frequently can be expressed by a variety of performance indicators, which are seldom independent. Interaction with members of the affected community often can uncover a variety of complaints from which root problem areas can be identified. Theoretical hypotheses, sophisticated survey procedures, and data analysis techniques, such as factor analysis, are available, if necessary, to assist in defining a basic set of dimensions \( \{d_i\} \) for the problem area.

For each relevant dimension, a measure of the current distribution of the population across that variable may be obtained by survey or other measurement technique. An estimate of the discontent with existing conditions can also be made by attitude sampling. If, for example, inadequate housing is believed to be a relevant dimension, a housing scale can be constructed including, say, a measure of rooms per person and the availability or quality of facilities. The housing status of the residents of the target area is then classified according to this measure in terms of a frequency distribution. Figure 11 is a hypothetical example of the status of a population with respect to housing on a 12-point scale.

For illustration [see also the Census Bureau housing classification scale (6)], suppose that this scale is built upon three indicators: 1) rooms per person, 2) electricity and plumbing adequacy, and 3) state of repair. Suppose a value is assigned to each indicator that reflects its relative distance from a given reference group, in this case the remainder of the community within which the target population is imbedded. An index of quality can then be obtained by averaging across the three indicators. The example in Figure 11 would thus represent a quality of housing distribution skewed to the low end of the scale; the distribution for the remainder of the community is shown by the dashed histogram in the figure.

Data of this form in each relevant problem dimension is the basic information requirement for the implementation of the model. The data in this form both establishes the present status of the target population and provides the framework for quantitatively stating program objectives. If an objective of a program is to improve the quality of housing, then it is necessary that the planner estimate the distribution that he would like to attain. This may be a distribution close to that for the community as a whole (shown in dashed lines in Figure 11) or a more conservative goal.

The quantitative statement of existing and proposed performance distributions is necessary to estimate the program implementation dollar costs, to assess potential changes in residual and ventable discontent, and to provide standards against which to assess program progress and effectiveness subsequent to implementation. Furthermore, by comparing performance distributions across subpopulations of the poor, those groups with the greatest constraints, and correspondingly highest residual discontent, can be identified. This information can be of use in assigning program priorities.

Clearly, all programs to improve the mean value of a given performance variable are not equivalent in terms of the social and psychological outcomes; the nature of the distribution about that mean is also extremely important, as the model emphasizes. Thus we hold that it is not sufficient for the planner to specify objectives in terms of average values; if social and psychological outcomes are to be assessed, then the relative distributions of performance that are expected from the proposed change must also be specified.

Having obtained the initial frequency distributions of target area populations over each of the relevant performance dimensions, and having selected distributions corresponding to program objectives, it is then necessary to determine those forces that have been responsible for producing the present condition. One must identify procedures for changing these forces so as to effect the desired objectives.

We have already given examples of the derivation of force potential from distribution functions. Let us now discuss the use of the model as a theoretical and methodological tool for uncovering the sources (causes) of these forces. The basic procedure begins by dividing the target populations into subgroups according to hypotheses or educated hunches (with the benefit of the best of the social science literature) as to the source of barriers that appear in the force potential; then the force potential for each subgroup is computed, using the model. From the results, it can be determined whether or not the hypothesis according to which the subgroup was selected was significant in affect-
PLANNING SOCIAL CHANGE

ing the barrier. Perhaps this will be clear if we consider the following example.

Figure 6 shows the 1967 economic opportunity potential for whites in the age group 25 to 34 years. Notice the barrier peaking at about $10,000. Notice also in Figure 4, which shows the corresponding potential for blacks, that the barrier is considerably higher and occurs at about $8,000. In view of the educational differential between whites and blacks in this age group, we might hypothesize that the barrier is to some degree the result of insufficient education to qualify for higher paying jobs.

To test this hypothesis, from the 1967 census data we have obtained distributions of incomes for U.S. males ages 25 to 34 who are high school graduates and who are college graduates. For these two income distributions, we computed the corresponding economic opportunity potential. These are compared in Figure 12. Notice the drastic reduction in the height of the barrier for college graduates compared to high school graduates. These results are consistent with the hypothesis that educational attainment is an important factor in changing the barrier height.

Pursuing the same example, one can compare the residual discontent for the two groups of people. We have performed this calculation and have obtained the estimate that the residual discontent with respect to economic opportunity is reduced by a factor of almost five for the college graduates compared to the high school graduates.

Summary

In this section we have tried to sketch the principal features of a social dynamics model and to present examples of its applicability to urban social planning problems. In summary, the model presents a conceptual framework for data collection and for the analysis of social planning alternatives. It provides means for exploring the relationship of social change options to the social and psychological outcomes, particularly in terms of residual and ventable discontent. Furthermore, it suggests a technique for testing the efficacy of social change alternatives with respect to cause and effect relations between social forces and behavioral performance. Finally, it provides a framework not only for quantitatively stating planning objectives, but also for subsequent appraisal of progress and program effectiveness.

REFERENCES


Fig. 12. Comparison of 1967 income opportunity potential functions for U.S. high school and college graduates, ages 25 to 34.
In References (3) and (4) we hypothesized that if a social force is defined by
\[ F = -\frac{\partial S}{\partial \sigma} \] (1)
where \( \sigma \) is the deviation from some reference behavior, then the sanction pattern, \( S \), determines the probability of behavior of an individual according to the equation
\[ S(\sigma, t)\psi = \frac{1}{2} k \frac{\partial^2 \psi}{\partial \sigma^2} + i \frac{\partial \psi}{\partial t} . \] (2)

The solution \( \psi(\sigma, t) \) is to be understood as related to the behavior expected of the individual at time \( t \). The product of \( \psi \) with its complex conjugate, \( \psi^* \), gives the probability density that at time \( t \) the behavior in a given activity is \( \sigma \). We call \( \psi \) the probability amplitude.

The mathematical expectation for the result of any observation of the behavior of an individual can be computed from the probability density according to the prescription
\[ \langle \sigma \rangle = \int_{-\infty}^{\infty} \psi^* \psi \sigma \, d\sigma. \] (3)

We have represented the potential, \( S \), explicitly as a function of \( \sigma \) and \( t \); however, let us consider the simple case in which it is constant in time and dependent only on the deviation of behavior from a given reference, so that
\[ S(\sigma, t) = S(\sigma). \] (4)

If we write
\[ \psi(\sigma, t) = \psi(\sigma)\varphi(t), \] (5)
then
\[ \frac{\partial^2 \psi}{\partial \sigma^2} = \frac{2}{k} [S(\sigma) - \lambda] \psi, \] (6)
and
\[ \varphi = \varphi_0 e^{-i\lambda t}, \] (7)
where the separation constant \( \lambda \) is called the eigenvalue of the equation.

Under these conditions, Eqs. 1 and 6 operationally define "force" to within a multiplicative constant. This can be seen explicitly as follows. From Eq. 6 we have
\[ [S(\sigma) - \lambda] = \frac{k}{2} \frac{\partial^2 \psi}{\partial \sigma^2} = \frac{k}{2} f(\sigma). \] (8)
Taking the derivative with respect to \( \sigma \) gives
\[ F = -\frac{\partial S(\sigma)}{\partial \sigma} = \frac{k}{2} f'(\sigma). \] (9)

For an observed probability density \( \psi^* \psi \), the right-hand side of Eq. 9 is determined to within the constant \( k \).

Consider, for example, a linear restoring force; that is, the force on an individual is proportional to the deviation from some desired behavior, \( \sigma_0 \), and always in the direction of greater conformity. Such a force may be written as
\[ F = -\gamma(\sigma - \sigma_0), \] (10)
where \( \gamma \) is a proportionality constant. The corresponding energy level diagram and force potential are shown in Figure 13.
that the higher the level of discontent, the higher the probability of deviant behavior. Qualitatively these properties are representative of results to be expected for a wide range of social forces.

The lowest level of discontent, the "residual discontent," for a linear restoring force leads to a normally distributed behavioral distribution; that is, \( \psi_0 \neq \psi_o \) corresponding to \( \lambda_0 \) is a Gaussian distribution as shown in Figure 14. A case in which the force constant [Eq. (10)], \( \gamma \), equals \( 1/2s_0^2 \) leads to behavior that is normally distributed about \( \sigma_0 \) with a standard deviation of \( s_0 \), with a residual discontent of \( 1/2s_0^2 \), and a ventable discontent (level-spacing of \( 1/s_0^2 \)). The first excited state results in a bimodal behavioral distribution with maxima at \( \sigma = \sigma_0 \pm \sqrt{2} s_0 \). This illustrates several important features of the model. The more confining the force (i.e., the greater \( \gamma \)), the greater is the conformity of that part of the population that remains in the state of lowest discontent; however, the greater also is the magnitude of the residual discontent. Thus, one cannot achieve conformity by impressing coercive forces upon a population without a resulting increase in residual discontent; in many cases an increase in ventable discontent also results. Furthermore, to assure conformity under any set of forces, means must be available for persons to vent (i.e., dissipate) their discontent.
PREDICTION AND MEASUREMENT IN TOTAL SOCIETY SYSTEMS

P. N. Rastogi

Prediction and measurement are two basic concerns of scientific enquiry. (Auguste Comte envisaged for sociology the goal of knowing in order to foresee and foreseeing in order to control.) This paper seeks to structure these two concerns in the context of total society systems. The theoretical approach developed here has been applied to five diverse societies—Brazil, France, India, Nigeria, and U.S.A. The model results are highly consistent and show a correspondence within \( \pm 2\% \) with the empirical indices like economic growth, price-rise, and unemployment. The structuring process permits the realization of predictive inference and retrodictive comparison in an unambiguous manner. The scope of the model encompasses exogenous pressure, internal system processes, and endogenous events. The latter are interpreted as processes extended in time.

Society as a Complex Adaptive System

Social existence of living beings including humans is based on the values of their survival. The concept of "basic human needs" was postulated by Malinowski in this context. Institutions of society are evolutionary structures. They emerged; exist, and change for the realization of basic values like sustenance, health, sex, security, and orientation. An institution is a social structure articulating human values. The structure consists of roles, statuses, norms, and rituals. Economy as a societal institution is related to sustenance, polity to regulation, marriage to sex, military to security, religion (or ideology) to orientation, and so on. The institutions are interrelated and together constitute a system. The interrelationship follows from the interrelated nature of human values. The structures for their fulfilment cannot function in isolation. Society may then be viewed as a complex system of interacting institutions. The outputs of these interactions issue forth as a number of social indicators in terms of which we may judge the performance, i.e., relative viability or weakness of the institution(s) concerned. The interaction of primary societal institutions and a resultant minimal set of performance variables or indicators may be outlined in the following a matrix form:

**INTERACTION BETWEEN INSTITUTIONS: THE INDICATOR VARIABLES**

<table>
<thead>
<tr>
<th>THROUGHPUTS</th>
<th>II</th>
<th>ECON</th>
<th>GOVT.</th>
<th>REL-IDEOL</th>
<th>ED</th>
<th>HL</th>
<th>MIL</th>
<th>FAML</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>ECON</td>
<td>EG</td>
<td>ET</td>
<td>ES</td>
<td>LP</td>
<td>LE</td>
<td>PMP</td>
<td>PR</td>
</tr>
<tr>
<td></td>
<td>EAL</td>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UE</td>
</tr>
<tr>
<td></td>
<td>PR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>GOVT</td>
<td>EG</td>
<td>AE</td>
<td>ES</td>
<td>LP</td>
<td>LE</td>
<td>PMP</td>
<td>PG</td>
</tr>
<tr>
<td></td>
<td>EAL</td>
<td>ET</td>
<td>ET</td>
<td></td>
<td>RD</td>
<td></td>
<td>PG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PR</td>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>N</td>
<td>UE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>REL-IDEOL</td>
<td>EG</td>
<td>AE</td>
<td>ES</td>
<td>LP</td>
<td>PMP</td>
<td>PG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>ET</td>
<td>ET</td>
<td></td>
<td>RD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>UT</td>
<td>UE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>S</td>
<td>EE</td>
<td>AE</td>
<td>ES</td>
<td>LP</td>
<td>LE</td>
<td>PMP</td>
<td>ES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ES</td>
<td></td>
<td></td>
<td>RD</td>
<td></td>
<td></td>
<td>PR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>MIL</td>
<td>EG</td>
<td>ES</td>
<td>ES</td>
<td>RD</td>
<td>PMP</td>
<td>ES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EAL</td>
<td>ET</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PR</td>
</tr>
<tr>
<td></td>
<td>PR</td>
<td>AE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>FAML</td>
<td>EG</td>
<td>AE</td>
<td>ES</td>
<td>LP</td>
<td>LE</td>
<td>PMP</td>
<td>PG</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Given obvious space limitations, the above variables are not properly elaborated here. PG, population growth rate, refers to increase in the numerical size of a society. EG, economic growth rate, refers to the change in the material resource base of a society. UE, unemployment, indicates the proportion of total working force that remains unabsorbed in the society's economic activities. PR, price rise, is a measure of relative scarcity and economic pressure on the society. ED, education, indicates the level of knowledge and communication in terms of literacy and stress on research and development. HL, Health, refers to the physical vitality level in terms of average expectancy of life in a society. EAL, economic activity level, indicates the relative volumes of productive investment in economy. ET, ethnic tension, refers to the disruptive aspect of internal divisiveness based on racial, cultural, and ideological factors. AE, administrative effectiveness, refers to the level of corruption and efficiency in the administrative machinery of a society. ES, expectations of socii, refers to their evaluation of ruling leadership of Government with respect to their ability to solve the society's major problems. PMP, politico-military pressure, signifies the impact of external and internal stresses and disturbances on the government of a society. It is held that this minimal set of performance variables encompasses the most salient aspects of the total society systems for the purposes of present study. They represent the state variables whose values define the state of a societal system.

Society is also an adaptive system. An adaptive system requires a control element that changes the course and responses of the system in the direction of increasing adaptation and/or protection from disturbances. Government or polity serves the function of such a regulator. It regulates the course of variables like unemployment, price rise, pressure from other societies, internal divisiveness, etc., in order to maintain system stability and realize system goals. If any of the regulated variables go "off" their courses, it would signify a state of system disturbance, i.e., a failure of the regulator. Under such conditions, policy and/or structure changes may ensue in the system.

The problem of measuring the "regulatedness" of variables in a common frame now arises.

The Measurement Schema

The rationale here is provided by the dialectical concepts of polar antinomies. According to this approach, a performance variable may range along a performance continuum whose two poles are given by the minimum and maximum performance limits. These limits are obtained by the observation of societal data in space and time. Thus, for example, economic growth per annum may range from a minimum of 0% (corresponding to periods of economic recession) to a maximum of 15% (corresponding to postwar performance of Japanese economy). Population growth may similarly range from a minimum of 0% (i.e., constant population) to a maximum of 4% in terms of Latin American data. Other performance variables may similarly be provided with performance limits.

This aspect brings us to the related issue of assessing the performance on a higher level of abstraction in terms of which the different performance values may be reduced to a common measure. This is made possible by introducing a logically compatible concept of viability (λ) of a performance variable. This concept refers to the survival capacity of a system to face the varying levels of stresses in terms of its basic vitality. Viability of performance variables may analogously be viewed to range along a continuum whose two poles are zero and one. The higher level of measurement is then accomplished by mapping the performance limits to the viability limits. This would then enable us to say, for example, that if the economic growth rate is 15%, its viability measure is one, or if the population growth rate is 4% or over, its λ value is zero.

The logic of polarity however leads us further to the idea of viability segments/zone, depending upon the relative closeness of a λ value to either of the two poles. These zones are defined by the equi-distant points on both sides of the mid-point boundary of λ=0.5. It follows that the zone defined by 0.75-1.00 would indicate maximum viability, and the one defined by 0.0-0.25 would show a highly disturbed state of a very low level of performance. These segmental λ limits are again mapped into performance continuum to yield a linear or non-linear relationship. This process of mapping the segmental limits (as distinct from the terminal limits) is again based on the interpretation of available data. Thus, to take the example of economic growth, the segmental boundaries are given by 0%, 3%, 6%, 10% and 15%. They comprise, besides the maximum and minimum observed rates of growth, the intermediate situation classes as observed in different societies. Similarly, for population growth, λ-zones are defined by 0%, 1.3%, 2.1%, 3% and 4%. Here the top segment represents the demographic situation in industrially advanced 1. The variables selected here are interpreted in terms of their widest implications. PG, for example, is not a mere statistical quantity; it is a crucial indicator of the interactions of Economy Family, Health, and Religion. Riesman, for example, points out that populations with a high birth rate and high death rate tend to be guided by custom; those with a high birth rate and a declining death rate, by conscience; those with a low birth rate and low death rate, by opinion (The Lonely Crowd, Yale University Press). Similarly, economic growth is an indicator of a number of institutions, besides economy, with implications for administrative effectiveness and politico-military pressure.
countries and the bottom one in the Latin American societies. The middle segments represent the situations in Asian and African societies. The measurement frame elaborated so far may now be summarized as follows:

\[
\begin{array}{c|c|c|c|c|c}
\text{EG} & I & II & III & IV & \lambda \\
0\% & 3\% & 6\% & 10\% & 15\%
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c}
\text{PG} & I & II & III & IV & \\
4.0\% & 3\% & 2.1\% & 1.3\% & 0\%
\end{array}
\]

The overall viability level \(Z\) of a society at any period may now be expressed as:

\[
Z = \sum_{j=1}^{\lambda} \frac{\lambda_j X_j}{\Sigma j}.
\]

The additive nature of the function follows from the conceptual and empirical semi-independence of the constituent state variables; i.e., they do not co-vary exponentially. A multiplicative relationship would render the system viability or stability excessively sensitive to even relatively minor changes in the performance variables. The system under study here is not a closely coupled one. It is a large, nonlinear system possessing high inertia.

Since \(Z\) here also varies between the two poles, it follows that for viable and stable systems \(0.5 < Z \leq 1.0\).

The question of the relative importance of variables (i.e., their weights in the system or \(Z\)-estimation) has remained moot so far. If we are able to measure the weights of the variables, the expression for \(Z\) then becomes

\[
Z = \sum_{j=1}^{\lambda} W_j \lambda_j X_j / \sum_{j=1}^{\lambda} W_j.
\]

The values of \(Z\) in terms of this formula have been calculated for the five societies and represented in the form of a \(Z\) curve. (Graphs I - V). The solution to the problem of weight estimations is provided by the concept of a societal flow graph discussed later. The methodology of \(\lambda\) estimation may be summarized as follows.

(1) Identify relevant types of situations in different societies over time and space.

(2) Develop a classification or typology according to the magnitude of their increasing severity or intensity.

(3) Map these situation classes into the four zones of a polar continuum.

(4) Associate the situation under study with such a class and place it within a zone. Select a midpoint value between segment boundaries as a tentative estimate.

(5) Determine the final estimate according to method of successive averages and/or maximum likelihood estimation \(^2\) in the light of the best available information, to reduce placement uncertainty.

The above procedure reduces the estimation error of parameters exponentially in accordance with the following expression:\(^3\)

\[
p_n = p_1 a^{n-1}
\]

where \(p_n\) is error probability after \(n\) successive averages or approximations, \(p_1\) is the initial error probability corresponding to the first midpoint segment estimate, and \(a\) is the fraction by which the error probability is reduced after each successive estimate.

In the situation discussed here, after a variable has been placed in a segment, the first midpoint average may have a maximum error of \(\pm 0.125\) in a range of 0.25. This would connote an initial maximum error probability of 50%, i.e., \(p_1 = 0.5\). Each successive average would reduce the error by 0.062, 0.031, and 0.0155 segmental points; i.e., corresponding to maximum error probabilities of 25%, 12.5%, and 6.25% respectively (\(a^2 - 1 = 0.25\), \(a^3 - 1 = 0.125\), and \(a^4 - 1 = 0.0625\)). This would mean a maximum error probability after second, third, and fourth estimates,

\[
p_2 = (0.50)(0.25) = 0.125
\]

\[
p_3 = (0.50)(0.25)(0.125) = 0.015
\]

\[
p_4 = (0.50)(0.25)(0.125)(0.625) = 0.00098 = 0.001
\]

They may be either positive or negative, and in a process of summation it is likely that some of them may cancel out each other in the aggregate value.

Flow Graph of Societal System

The flow graph of a societal system is a representation of the processes, their directionality, and the multilateral pattern of relationships within a system. It depicts the course of institutional interactions that eventuate in the varying values of state variables over time. It may be viewed as an amplification of the interaction matrix in interrelating the state variables (main variables) through a set of intermediate and auxiliary variables. It is


3. This function is based on the error response probability reduction of learning theory. See, for example, R. Atkinson, et al., An Introduction to Mathematical Learning Theory, Wiley, New York, 1968, pp. 15-16.
Table 1.

RANK ORDERS OF FLOW GRAPH VARIABLES

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Variables</th>
<th>Information Links Meeting</th>
<th>Constraint Rank</th>
<th>Information Links Meeting</th>
<th>Control Rank</th>
<th>Total No. of Links</th>
<th>Salience Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IR (IV)</td>
<td>2</td>
<td>II</td>
<td>5</td>
<td>V</td>
<td>7</td>
<td>VI</td>
</tr>
<tr>
<td>2</td>
<td>GS (MV)</td>
<td>5</td>
<td>V</td>
<td>3</td>
<td>III</td>
<td>6</td>
<td>VI</td>
</tr>
<tr>
<td>3</td>
<td>PMP (MV)</td>
<td>3</td>
<td>III</td>
<td>3</td>
<td>IV</td>
<td>5</td>
<td>V</td>
</tr>
<tr>
<td>4</td>
<td>PU (IV)</td>
<td>2</td>
<td>II</td>
<td>4</td>
<td>III</td>
<td>5</td>
<td>IV</td>
</tr>
<tr>
<td>5</td>
<td>EGA (MV)</td>
<td>2</td>
<td>II</td>
<td>3</td>
<td>II</td>
<td>5</td>
<td>IV</td>
</tr>
<tr>
<td>6</td>
<td>UE (MV)</td>
<td>3</td>
<td>III</td>
<td>2</td>
<td>II</td>
<td>5</td>
<td>IV</td>
</tr>
<tr>
<td>7</td>
<td>ES (MV)</td>
<td>3</td>
<td>III</td>
<td>1</td>
<td>II</td>
<td>4</td>
<td>III</td>
</tr>
<tr>
<td>8</td>
<td>ET (MV)</td>
<td>4</td>
<td>IV</td>
<td>2</td>
<td>II</td>
<td>4</td>
<td>III</td>
</tr>
<tr>
<td>9</td>
<td>PG (MV)</td>
<td>2</td>
<td>II</td>
<td>3</td>
<td>III</td>
<td>4</td>
<td>III</td>
</tr>
<tr>
<td>10</td>
<td>HL (MV)</td>
<td>3</td>
<td>III</td>
<td>3</td>
<td>III</td>
<td>4</td>
<td>III</td>
</tr>
<tr>
<td>11</td>
<td>ED (MV)</td>
<td>1</td>
<td>I</td>
<td>3</td>
<td>III</td>
<td>4</td>
<td>III</td>
</tr>
<tr>
<td>12</td>
<td>AE (MV)</td>
<td>2</td>
<td>II</td>
<td>1</td>
<td>III</td>
<td>3</td>
<td>II</td>
</tr>
<tr>
<td>13</td>
<td>FP (IV)</td>
<td>2</td>
<td>II</td>
<td>1</td>
<td>III</td>
<td>3</td>
<td>II</td>
</tr>
<tr>
<td>14</td>
<td>EAL (MV)</td>
<td>2</td>
<td>II</td>
<td>1</td>
<td>II</td>
<td>3</td>
<td>II</td>
</tr>
<tr>
<td>15</td>
<td>PR (MV)</td>
<td>2</td>
<td>II</td>
<td>1</td>
<td>II</td>
<td>3</td>
<td>II</td>
</tr>
<tr>
<td>16</td>
<td>DE (IV)</td>
<td>1</td>
<td>I</td>
<td>2</td>
<td>II</td>
<td>2</td>
<td>I</td>
</tr>
<tr>
<td>17</td>
<td>FA (AV)</td>
<td>1</td>
<td>I</td>
<td>1</td>
<td>II</td>
<td>2</td>
<td>I</td>
</tr>
<tr>
<td>18</td>
<td>PEG (IV)</td>
<td>1</td>
<td>II</td>
<td>2</td>
<td>II</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>History (AV)</td>
<td>-</td>
<td>II</td>
<td>2</td>
<td>II</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Weather (AV)</td>
<td>-</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

IV = Intermediate Variable  
MV = Main Variable  
AV = Auxiliary Variable

oriented around the processes shaping Government stability. Insofar as Government is the control subsystem of society, values of GS should be identical with those of Z, which represents the regulatedness of system variables or societal viability.

Each of the variables in the flow graph receives or emits a number of information links to other variables. The number and types of links associated with a variable signifies its relative importance or rank weight in the system. The variables who emit and receive the largest number of links are the most important from the control and constraint points of view. Those which receive and emit the maximum number of links are, however, the most salient. The salience rank of a variable represents its relative weight (Wj) in the viability estimation (Z) of the system (see Table No. 1) Control and constraint ranks are important in the context of planning and change objectives.

The picture of society that emerges from the
PREDICTION AND MEASUREMENT IN TOTAL SOCIETY SYSTEMS

flow graph elucidates the nature of society as a complex system. Society is a closed system of interacting processes. These processes illustrate the interaction of societal institutions. Government as the regulatory subsystem of society is confronted with a number of internal and external disturbances (PMP) which strain its regulating capacity and result in a commitment of resources to internal and external security needs (DE). This reduces the quantum of resources available for productive input to the economy (IK). The available resources lead to a maximum possible rate of economic growth (PEG). This maximum is, however, subject to a society's relative efficiency in the utilization of its resources. This efficiency factor, termed administrative effectiveness (AE), is seen to be determined by two major elements: education (ED) and government stability (GS). The higher the educational level of a society, the more efficient would its administration be. Similarly, government stability is vital for the administrative processes to be carried out effectively. The relative role of education and government stability are indicated by their rank weights in the flow graph. Administrative effectiveness acting as a constraint on the maximum rate of growth (i.e., potential economic growth) determines the actual rate of economic growth (EGA). The latter affects unemployment (UE) and price rise (PR) in conjunction with population growth (PG). Unemployment and price rise lead to public unrest (PU) and determine the extent of its stability or instability of postwar France and Italy may also be interpreted in terms of a low leadership factor despite their relative affluence as "developed" industrial societies.

Stability: Estimation Through Flow Graph Simulation

Z-estimation is based on the viability indices of the essential variables sought to be regulated by the system control. It is based on the values of state variables which at any time period define the state of the system. Flow graph, on the other hand, represents the dynamics of system processes whose interaction eventuates in the above system variables. Stability estimation through the simulation of flow graph processes should then, in principle, result in the identical values of Z. The data categories like unemployment, price rise, economic growth, etc., generated by the simulation model, should again display a close correspondence with the actual values obtainable from the secondary sources and utilized in Z-estimation through the computation of their λs. Others like ET, ES, AE, etc. should also be closely comparable with both the modes of computation. A simulation algorithm in this context should be able to depict unambiguously the quantitative nature of relationships between the variables and be consistent with the theoretical framework of information linkages outlined in the flow graph.

The Nature of Simulation Algorithm

The nature of simulation algorithm here is based on Wiener's principle of entrainment of frequencies. According to this principle, the constituent parts of a dynamic system work synchronously so that their frequencies match one another. In postulating this mechanism, Wiener was guided by the case of the electric power generating system in which, through negative feedback, a number of alternators can maintain a sharp frequency despite variations of load. The frequencies pull one another together or "attract" such that slow alternators are speeded up and fast ones slowed down. He noted entrainment in various diverse situations, e.g., the flashing of fireflies in unison, the maintenance of crystalline form in snowflakes, the lumping in the periods of asteroids, the diurnal rhythm in many animals etc. The writer has examined its implications for organizational behaviour. In the context of the system under study here (the flow graph), the principle implies that in its dynamic operation the process variables are functioning together at a matching and compatible pace. In other words, their frequencies within a particular phase (or λ segment) correspond with another. The dynamic basis of change and interaction between variables lies in the mapping of their values in corresponding λ segments.


673
A Comparative Evaluation

The details of the simulation algorithm in terms of the numerous nonlinear functional relationships and their mapping into the compatible procedure where, given an estimate of PMP and the data value of ED and PG, the rest of the variables can be generated. Alternatively, given the value of EGA, values of other variables including PMP can be generated.

The parallelism in the values of Z and GS is demonstrated for all the five societies in the accompanying graphs. The range of variations lies with ±1-2%.

A detailed comparison of some estimates in the two computational modes for the last decade is now given in a tabular form.

Table 2

<table>
<thead>
<tr>
<th>Year</th>
<th>Z-value</th>
<th>GS-value</th>
<th>Actual</th>
<th>Simulated</th>
<th>PR (%)</th>
<th>Actual</th>
<th>Simulated</th>
<th>UE (%)</th>
<th>Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAZIL</td>
<td>1961</td>
<td>0.418</td>
<td>0.415</td>
<td>5.0</td>
<td>3.38</td>
<td>Av</td>
<td>3.50</td>
<td>35.3</td>
<td>Av</td>
</tr>
<tr>
<td></td>
<td>1962</td>
<td>0.384</td>
<td>0.389</td>
<td>1.6</td>
<td>3.17</td>
<td>Av</td>
<td>50.2</td>
<td>-</td>
<td>49.6</td>
</tr>
<tr>
<td></td>
<td>1963</td>
<td>0.399</td>
<td>0.395</td>
<td>3.10</td>
<td>3.15</td>
<td>Av</td>
<td>71.0</td>
<td>-</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>1964</td>
<td>0.406</td>
<td>0.417</td>
<td>3.9</td>
<td>3.36</td>
<td>Av</td>
<td>87.0</td>
<td>-</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>1965</td>
<td>0.404</td>
<td>0.420</td>
<td>2.8</td>
<td>3.44</td>
<td>Av</td>
<td>50.0</td>
<td>-</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>1966</td>
<td>0.420</td>
<td>0.420</td>
<td>5.1</td>
<td>4.34</td>
<td>Av</td>
<td>47.6</td>
<td>-</td>
<td>49.6</td>
</tr>
<tr>
<td></td>
<td>1967</td>
<td>0.483</td>
<td>0.476</td>
<td>4.8</td>
<td>4.84</td>
<td>Av</td>
<td>29.8</td>
<td>-</td>
<td>23.1</td>
</tr>
<tr>
<td></td>
<td>1968</td>
<td>0.555</td>
<td>0.561</td>
<td>8.3</td>
<td>6.63</td>
<td>Av</td>
<td>24.2</td>
<td>-</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>1969</td>
<td>0.574</td>
<td>0.579</td>
<td>8.3</td>
<td>7.66</td>
<td>Av</td>
<td>23.1</td>
<td>-</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>1970</td>
<td>0.683</td>
<td>0.584</td>
<td>9.5</td>
<td>8.07</td>
<td>Av</td>
<td>22.0</td>
<td>-</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td>1971</td>
<td>0.587</td>
<td>0.587</td>
<td>10.0</td>
<td>8.20</td>
<td>Av</td>
<td>22.0</td>
<td>-</td>
<td>21.7</td>
</tr>
</tbody>
</table>

*50% of PR per year is the lower limit of the functional relationship as defined here.

Av-Average

COMMENTS: UE statistics for Brazil do not appear in UN data sources as a set of regular annual figures. A recent survey estimates the extent of unproductive under-employment at 40% of the total labour force. It also puts open unemployment between 10-14% of the economically active population. (Economic Survey of Latin America - 1969, UN, New York, 1970, p. 18.). EG picture in later years is complicated by currency devaluations. The last one occurred in January 1972.

FRANCE

<table>
<thead>
<tr>
<th>Year</th>
<th>Z-value</th>
<th>GS-value</th>
<th>Actual</th>
<th>Simulated</th>
<th>PR (%)</th>
<th>Actual</th>
<th>Simulated</th>
<th>UE (%)</th>
<th>Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961</td>
<td>0.618</td>
<td>0.610</td>
<td>3.6</td>
<td>4.14</td>
<td>Av</td>
<td>3.3</td>
<td>3.8</td>
<td>-</td>
<td>4.0</td>
</tr>
<tr>
<td>1962</td>
<td>0.613</td>
<td>0.595</td>
<td>7.0</td>
<td>3.75</td>
<td>Av</td>
<td>4.1</td>
<td>4.0</td>
<td>-</td>
<td>3.4</td>
</tr>
<tr>
<td>1963</td>
<td>0.604</td>
<td>0.587</td>
<td>4.8</td>
<td>3.58</td>
<td>Av</td>
<td>4.3</td>
<td>3.76</td>
<td>-</td>
<td>3.2</td>
</tr>
<tr>
<td>1964</td>
<td>0.628</td>
<td>0.620</td>
<td>5.4</td>
<td>4.08</td>
<td>Av</td>
<td>3.4</td>
<td>3.76</td>
<td>-</td>
<td>2.8</td>
</tr>
<tr>
<td>1965</td>
<td>0.637</td>
<td>0.653</td>
<td>4.7</td>
<td>4.8</td>
<td>Av</td>
<td>2.5</td>
<td>3.2</td>
<td>-</td>
<td>2.8</td>
</tr>
<tr>
<td>1966</td>
<td>0.648</td>
<td>0.659</td>
<td>4.9</td>
<td>4.96</td>
<td>Av</td>
<td>3.0</td>
<td>3.1</td>
<td>-</td>
<td>2.9</td>
</tr>
<tr>
<td>1967</td>
<td>0.628</td>
<td>0.635</td>
<td>4.7</td>
<td>4.6</td>
<td>Av</td>
<td>2.8</td>
<td>3.35</td>
<td>-</td>
<td>3.5</td>
</tr>
<tr>
<td>1968</td>
<td>0.579</td>
<td>0.573</td>
<td>4.5</td>
<td>3.46</td>
<td>Av</td>
<td>4.7</td>
<td>4.0</td>
<td>-</td>
<td>3.2</td>
</tr>
<tr>
<td>1969</td>
<td>0.641</td>
<td>0.648</td>
<td>4.5</td>
<td>4.69</td>
<td>Av</td>
<td>6.4</td>
<td>3.4</td>
<td>-</td>
<td>2.9</td>
</tr>
<tr>
<td>1970</td>
<td>0.610</td>
<td>0.617</td>
<td>5.8</td>
<td>4.15</td>
<td>Av</td>
<td>5.5</td>
<td>3.8</td>
<td>-</td>
<td>3.2</td>
</tr>
<tr>
<td>1971</td>
<td>0.597</td>
<td>0.601</td>
<td>5.4</td>
<td>3.9</td>
<td>Av</td>
<td>6.8</td>
<td>4.0</td>
<td>-</td>
<td>3.3</td>
</tr>
</tbody>
</table>

COMMENTS: UE data as interpreted here do not appear in UN statistics. The pattern of increase or decrease is, however, in accordance with the aggregate figures given in UN statistics. The unemployment figures generated here are higher than the official figures by ≥1%.

INDIA

<table>
<thead>
<tr>
<th>Year</th>
<th>Z-value</th>
<th>GS-value</th>
<th>Actual</th>
<th>Simulated</th>
<th>PR (%)</th>
<th>Actual</th>
<th>Simulated</th>
<th>UE (%)</th>
<th>Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961</td>
<td>0.470</td>
<td>0.478</td>
<td>2.5</td>
<td>2.70</td>
<td>Av</td>
<td>6.0</td>
<td>8.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1962</td>
<td>0.449</td>
<td>0.446</td>
<td>1.9</td>
<td>2.11</td>
<td>Av</td>
<td>6.04</td>
<td>9.10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1963</td>
<td>0.449</td>
<td>0.456</td>
<td>4.0</td>
<td>2.16</td>
<td>Av</td>
<td>8.5</td>
<td>8.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1964</td>
<td>0.464</td>
<td>0.469</td>
<td>6.6</td>
<td>2.34</td>
<td>Av</td>
<td>10.0</td>
<td>8.5</td>
<td>Av</td>
<td>8.5</td>
</tr>
<tr>
<td>1965</td>
<td>0.391</td>
<td>0.402</td>
<td>-5.0</td>
<td>1.81</td>
<td>Av</td>
<td>8.51</td>
<td>9.43</td>
<td>Av</td>
<td>8.9</td>
</tr>
<tr>
<td>1966</td>
<td>0.394</td>
<td>0.374</td>
<td>1.8</td>
<td>1.73</td>
<td>Av</td>
<td>13.8</td>
<td>9.99</td>
<td>Av</td>
<td>12.03</td>
</tr>
<tr>
<td>1967</td>
<td>0.460</td>
<td>0.460</td>
<td>9.00</td>
<td>2.27</td>
<td>Av</td>
<td>11.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1968</td>
<td>0.443</td>
<td>0.450</td>
<td>2.4</td>
<td>2.26</td>
<td>Av</td>
<td>3.03</td>
<td>8.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1969</td>
<td>0.450</td>
<td>0.457</td>
<td>3.3</td>
<td>2.34</td>
<td>Av</td>
<td>7.52</td>
<td>8.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1970</td>
<td>0.455</td>
<td>0.457</td>
<td>4.7</td>
<td>2.39</td>
<td>Av</td>
<td>7.82</td>
<td>8.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1971</td>
<td>0.423</td>
<td>0.438</td>
<td>3.0</td>
<td>2.20</td>
<td>Av</td>
<td>8.02</td>
<td>7.7</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

1. Represents the impact of a technological productivity factor in agriculture, i.e. "green revolution."
2. Manufacturers Index (Total)
3. Consumer Price General Index (Working class)
4. Food Items

COMMENTS: Position regarding UE data here is the same as before. However, the figures here agree with the current estimates of the total number of unemployed put around 30 million; i.e., 10-12% of the working force.
## Table 2 (Continued)

### A COMPARISON OF SOME ESTIMATES IN FIVE SOCIETIES

<table>
<thead>
<tr>
<th>Year</th>
<th>Z-value</th>
<th>GS-value</th>
<th>Actual</th>
<th>Simulated</th>
<th>Actual</th>
<th>Simulated</th>
<th>Actual</th>
<th>Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NIGERIA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1961</td>
<td>0.504</td>
<td>0.512</td>
<td>3.3</td>
<td>3.28</td>
<td>8.0</td>
<td>8.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1962</td>
<td>0.494</td>
<td>0.501</td>
<td>5.7</td>
<td>5.20</td>
<td>6.5</td>
<td>6.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1963</td>
<td>0.509</td>
<td>0.502</td>
<td>4.6</td>
<td>3.12</td>
<td>0.0</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1964</td>
<td>0.475</td>
<td>0.488</td>
<td>2.9</td>
<td>2.86</td>
<td>2.5</td>
<td>2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>0.478</td>
<td>0.482</td>
<td>5.4</td>
<td>2.81</td>
<td>4.2</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td>0.405</td>
<td>0.403</td>
<td>2.6</td>
<td>2.28</td>
<td>18.0</td>
<td>19.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1967</td>
<td>0.329</td>
<td>0.332</td>
<td>0.0*</td>
<td>0.0*</td>
<td>0.0*</td>
<td>9.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968</td>
<td>0.311</td>
<td>0.314</td>
<td>0.0*</td>
<td>0.0*</td>
<td>0.0*</td>
<td>9.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1969</td>
<td>0.348</td>
<td>0.357</td>
<td>3.0*</td>
<td>1.17</td>
<td>20.0*</td>
<td>19.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>0.424</td>
<td>0.435</td>
<td>3.5</td>
<td>2.48</td>
<td>21.0</td>
<td>22.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td>0.488</td>
<td>0.500</td>
<td>6.0</td>
<td>3.46</td>
<td>12.1**</td>
<td>11.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Excludes the three eastern states who seceded.

**Average of General Index up to June '71.

**COMMENTS:** UE data as a proportion of the total work population is not available. There appear to be serious discrepancies in the official UN statistics for price indices and the situation studied by other independent sources (See footnote 6 in text).

| **U.S.A.** | | | | | | | | |
| 1961 | 0.538   | 0.545    | 2.0    | 2.77      | 1.1    | 4.4       | 6.7    | 6.0       |
| 1962 | 0.627   | 0.643    | 6.5    | 4.9       | 1.9    | 2.8       | 5.5    | 4.0       |
| 1963 | 0.600   | 0.616    | 3.9    | 4.38      | 1.2    | 3.3       | 3.7    | 4.16      |
| 1964 | 0.611   | 0.628    | 5.2    | 4.69      | 1.3    | 3.10      | 5.2    | 4.05      |
| 1965 | 0.643   | 0.661    | 5.9    | 5.40      | 2.2    | 2.5       | 4.5    | 3.58      |
| 1966 | 0.641   | 0.663    | 5.5    | 5.55      | 3.9    | 2.55      | 3.8    | 3.64      |
| 1967 | 0.569   | 0.562    | 2.75   | 3.62      | 3.0    | 3.9       | 3.9    | 4.29      |
| 1968 | 0.600   | 0.613    | 4.5    | 4.05      | 4.1    | 3.7       | 3.6    | 4.26      |
| 1969 | 0.572   | 0.577    | 2.5    | 3.71      | 5.3    | 4.1       | 4.1    | 3.5       |
| 1970 | 0.509   | 0.516    | 1.5    | 2.60      | 5.4    | 4.8       | 4.9    | 5.94      |
| 1971 | 0.549   | 0.543    | 4.0    | 2.92      | 4.0    | 4.6       | 6.0    | 5.0       |

**COMMENTS:** A possible explanation of the discrepancy between earlier PR figures may be due to the fact that PR indices may not record the rise in cost of the various services like utilities, transport, and communications.

### SUMMARIZING THE RESULTS OF SIMULATION ALGORITHM AND Z-ESTIMATION

Summarizing the results of simulation algorithm and Z-estimation, we may note the following salient points:

1. Variations between Z values and GS values for all the five societies are within a range ± 1-2%.
2. PMP values in the both set of cases are identical.
3. Variations between the estimated values of AE, ES, and their generated values is also restricted to a range of ±1-2%. This follows directly from the correspondence of GS and Z values.
4. Variations between the values of UE, PR and EG available through secondary sources and the generated values are largely around ±2%. The comparison is, however, valid only in the cases where accurate and comparable data set are available and is vitiated in others.
5. In the case of societies like Brazil, India, and Nigeria where adequate UE and PR data may not be available, the generated values may serve as a reliable guide if the values of other variables like EG and PG are correct.

6. An idea of the possible discrepancies that may exist in this connection is brought about by the example of the price situation in Nigeria in 1967. While UN Statistics mention the price rise as being approximately 8% and 18% for non-food and food items respectively, another source mentions the steep rise in food prices to the extent of 100% and more during post-coup situation. (Schwarz, ibid, p. 203). Moreover, after the 1966-67 Biafran crisis, the prices are shown to decline till 1968. But what of Biafra? Is it to be included in Nigerian society? Computation here include Biafra as part of the Nigerian situation and high PR figures reflect the price situation in an inclusive manner. Biafra was known to be passing through acute financial difficulties during the period.

Similarly, commenting on the disparities between the official statistical picture of the price situation in India and the actual impact felt by the public, a respected Indian newspaper in a recent editorial writes, "Ordinary citizen knows from personal experience that he is paying more for his daily needs and yet experts tell him that the price index shows only a small rise. Why the statistical picture differs so sharply from personal experience is not clear." National Herald, Lucknow, Feb. 24, 1972.
tent correspondences with the observed data obtained from them. This leaves open the postulation of additional mechanisms or factors for any particular society to get a better fit with the data. This leaves open the postulation of additional mechanisms or factors for any particular society to get a better fit with the data.

(7) Economic growth and price rise discrepancies between the secondary data and the generated data here may be interpreted in terms of the impact of the external factors like foreign aid, monetary devaluations, investment of foreign capital, external trade factors, and the changes occurring in world commodity markets, producing an impact on the economic system of the society concerned. Here we treat the societal system as an essentially closed one which is operating in terms of its internal dynamics. External factors are subsumed under the category of politico-military pressure which serves as a forcing function on that system. However, it also incorporates the internal politico-military pressures, as evidenced in situations like insurrection, civil unrest, civil war, etc., which make for a large commitment of financial resources to security and other economically non-productive uses.

Realizability of Predictive Inference

The theoretical structure developed here, therefore, permits the realizability of predictive inferences for total society systems. The main themes may be summarized as follows.

(i) By estimating politico-military pressure through a maximum likelihood procedure, a preview of the system state for a period t + n with reference to present t may be obtained. This knowledge of the future system state would be in terms of the variables like economic growth, unemployment, price rise, ethnic tension, administrative effectiveness, public unrest, expectations of the socii, and stability of government, i.e., the configuration of variables in the flow graph. The supplementary requirements here would be a knowledge of the Z or GS for period t + n-1 and the values of education (in terms of literacy and emphasis on research and development), population growth, and leadership factor LDF.

(ii) Insofar as Government as a societal institution is the control subsystem of society as an adaptive system, the passage of any variable into the "collapse zone" of the viability continuum would signal an impending state of government instability, i.e., a failure of the regulatory element. The strength of the signal (s) would depend upon the salience rank and the number of variables involved.

(iii) The mode of political instability and/or government change may be orderly or disorderly, depending upon the closeness of Z/GS values to the polar limits (see graphs on p. 167). Systems in higher λ zones would display orderly change, owing to their higher innate vitality, while those in the lower zones, i.e., Z < 0.5, may be disorderly. The probability of violent change (revolutions, coups, and chaos) would increase directly with the approach of Z/GS values toward the collapse zone.

(iv) The same would be the case with the duration of political instability. The periods of political uncertainty may be short lived or prolonged, depending upon the Z/GS values. In the case of higher system viability values, Wiener's law would rectify the unstable situation (s), depending upon the (relatively short) period required for the matching of frequencies. Higher λ variables would pull up the lower λ variable (s).

(v) In the case of societies in higher λ zones where most of the variables may be performing viably, the change of government would be exclusively predicated around the variables ES unless it is overwhelmed by sudden external military pressure. A sequential decrease of Z/GS values, i.e., Z_t < Z_t-1 and correspondingly ES(λ)_t < ES(λ)_{t-1} during an electoral period, would provide a basis for prediction regarding the success or otherwise of the ruling party. This would, however, assume that LDF remains constant. In the case of marginal differences, the situation would be one of close electoral contest.

Conclusion

The emerging picture of society here is one pulsated by dynamic processes within an invariant frame provided by Wiener's law. The two major determinants of societal viability that may lead the system toward a trajectory of Z/GS maximization are the high values of leadership factor and the low values of politico-military pressure. Both the determinants conjoin together in propelling the system configuration from one viability zone to another.

ACKNOWLEDGMENTS

Work reported here was made possible by the grant of a travelling fellowship to the author by Nuffield Foundation. Grateful thanks are due to the trustees and authorities of the foundation for the help and facilities provided.

July 5-10, 1971. The writer is grateful to Professor Alexander Douglass, Karl Deutsch, and William Letwin for their constructive suggestions at various stages. They are, however, not responsible for the faults present here.

The paper is a highly condensed version of a book-length work under preparation, entitled The Dynamics of Government Stability and Societal Viability.
The purpose of this note is to provide a technique which may be used to define the concepts of political structure and political system. As suggested by Almond such definitions should be capable of distinguishing between those political elements which may be common to all societies and those which are particular to each. Indeed, the point of view taken here will be essentially to think of the former as comprising the political structure and the latter as determining the particular political system within that structure. This distinction will permit more rigorous and formal definitions than those already familiar to political scientists. It will also provide tests for determining the logical completeness of models and suggest a way of organizing future research.

Perhaps it is best to begin with an analogy drawn from elementary algebra. Consider for the moment two variables, say x and y. Let R_1 denote one relationship between them and R_2 another. For example, R_1 and R_2 may be linear as pictured in the diagram below:

Thus R_1 determines a line in the x-y plane, R_2 determines another line in that plane, and both relationships taken simultaneously determine the unique point of intersection Q. If there were a third relationship R_3, then there would be no point simultaneously satisfying R_1, R_2, and R_3: the set of relationships would be overdetermined. If there were only one relationship R_1 then the model would be underdetermined: a unique point on the line R_1 could not be specified. Thus, in order to have a determinate model (i.e., one which resolves itself into a unique point), it is necessary that the number of relationships be exactly equal to the number of variables. If the number of relationships and variables of a model are equal, the model will be said to pass the primary test for logical completeness.

Of course the fact that the number of relationships equals the number of variables does not guarantee that the model is determinant. R_1 and R_2 could turn out to be two ways of writing the same thing; they could describe two parallel lines which do not intersect; or they could be nonlinear curves in the plane which twist and turn but never cross. There are well-known mathematical conditions which, if met, ensure the existence of unique intersection points. The model will be said to satisfy the secondary test for completeness whenever such conditions are met.

One more comment about this analogy: R_1 and R_2 may themselves depend on parameters. Each time a parameter changes, a relationship will "shift." Thus, for example, a change in parameters could cause R_2 to change into R_3. In this case the intersection point would move from Q to T. Hence the intersection point depends on the values of the parameters as well as on the nature of R_1 and R_2. Specifying R_1 and R_2, then, determines a new relationship which relates each set of parameter values to a point of intersection.

In what follows an attempt is made to apply this technique to the problem at hand. Political variables and relationships between them are specified. It makes no difference that the variables cannot be quantified as in the above analogy because the property of "capability of quantification" is never used in the analysis. Political structure is defined as the set of relationships between the variables while a political system is thought of as an "intersection point" which depends on "parameter values."

It is axiomatic that there is no general agreement among political scientists as to which variables and relationships should be used to define political structure. Nor is there the place to resolve any such arguments. The variables and relationships between them are specified. It makes no difference that the variables cannot be quantified as in the above analogy because the property of "capability of quantification" is never used in the analysis. Political structure is defined as the set of relationships between the variables while a political system is thought of as an "intersection point" which depends on "parameter values."

*The author would like to acknowledge his debt to Clifford Kaufman and Lucy Behrman for many helpful comments and suggestions.
2. The existing literature on this subject is summarized by Almond, op. cit., pp. 5-7.
3. For example, let R_1 be the equation 3x + 4y = 1 and R_2 be 6x + 8y = 2. Then R_1 and R_2 describe the same line in the x-y plane.
4. In note 3 the parameters of R_1 are 3, 4, and 1.
relationships presented below are intended only as an illustration of the technique of definition proposed. Many of them have been taken from Apter.\(^5\) In order to simplify the discussion they will be presented in symbolic form. The variables are:

\begin{align*}
  v & \text{- values and goals of society.} \\
  c & \text{- coercion used by government to maintain authority and operate efficiently.} \\
  z & \text{- types and quantities of information flowing between government and those governed.} \\
  d & \text{- patterns and processes for making authoritative decisions.} \\
  a & \text{- ways in which government may be held accountable to a group or groups outside itself.} \\
  I & \text{- ideology.} \\
  P & \text{- government policy in all situations.}
\end{align*}

There is one parameter, \(n\), which denotes the needs, values and perceptions of reality of all individuals in society.

Seven relationships are assumed to hold. The first states that authoritative decision-making is related to the values of society, coercion, and information. This may be written as:

\[ d = f^1(v,c,z), \]

where \(f^1\) is a symbolic representation of the relationship itself. For the moment it is not necessary to specify the precise nature of \(f^1\); the important fact is that it exists. The remaining six relationships, which may be read verbally as the first, are:

\begin{align*}
  a & = f^2(v,c,z), \\
  c & = f^3(z), \\
  P & = f^4(v,c,z,I), \\
  v & = f^5(P,I), \\
  z & = f^6(P,I,v), \\
  I & = f^7(n,v).
\end{align*}

Note that the semicolon in the last relationship is used only to separate the parameter \(n\) from the variable \(v\).

This is a system of seven relationships in seven variables and one parameter. The political structure consists of \(f^1, \ldots, f^7\). For each value of the parameter \(n\), unique values of the variables frequently may be determined. The latter constitute the political system which, to use Apter's terminology, may range anywhere from the reconciliation to the mobilization type. The model passes the primary test for logical completeness since the number of relationships and variables are the same.

This technique of definition and the tests for completeness have been known to economists for some time. Léon Walras\(^6\) is usually given credit for being the first to apply them to economic analysis. He defined the "perfectly competitive economic model" as a system of simultaneous demand and supply relationships which reflect respectively the assumptions that for every set of input and output prices, (a) the rational consumer buys those commodities and sells those resources which maximize his satisfaction within the limitations of his budget, and (b) the rational firm produces those outputs and hires those inputs which maximize its profits. The concept of general equilibrium was then defined as a point which satisfies all relationships simultaneously (thus demand equals supply in every market). Economists have spent considerable time and effort studying the properties of this model. When stated in terms of simultaneous relationships it becomes immediately clear, for example, that due to the presence of budget constraints, in an economy with \(n\) markets if \(n-1\) of them are in equilibrium, then the last must, of necessity, also be in equilibrium. Therefore the logically complete formulation consists of the relationships and variables in \(n-1\) markets only. This is, indeed, a valuable lesson concerning the advantages and power of the technique proposed here.

Further investigations into the secondary tests of completeness have led economists into fruitful studies of the conditions under which general equilibrium exists and is unique. Thus restrictions are now known which must be placed on the parameters and relationships in order to guarantee that equilibrium is possible in the model. This is crucially important since observations of economic phenomenon in the real world are thought of as either being equilibrium points or located on a dynamic time-path which approaches equilibrium. It follows that if these restrictions are not assumed, then the model may have no bearing on reality since equilibrium may not exist within it. Finally, aside from the secondary tests, economists have also pursued inquiries into the stability and optimality (welfare aspects) of equilibria. The importance of these investigations is again obvious.

There is no reason why research in political science cannot be organized along similar lines. Political structure may be defined as a set of simultaneous relationships and political system as an equilibrium point. Questions concerning the existence, uniqueness, stability and optimality of the system may now be posed within the context.


\(6\). Léon Walras, Éléments d’Économie Politique Pure, Lausanne, 1874; trans. by W. Juffe, Homewood, Ill.: Irwin, 1954.
of a fixed postulated structure. As such they are relevant and important for, among other things, they help determine the logical completeness and usefulness of the model.

In conclusion, it may be useful to state specifically the kinds of questions this approach raises. First, what are the principal variables and relationships determining political structure out of which observable political systems may be derived as "intersection points"? How do political structures (i.e., the relationship) vary among societies and how do these variations affect intersection points? Is there one political structure out of which all observable political systems can be derived? What is the precise nature of the relationships and what restrictions must be placed on them in order that they be complete and useful? Finally, which values of the parameters in a given model determine which political system? Once the answer to the last is known it may become possible by observing changes in parameter values to predict the direction in which a political system is going to change.
F.

APPLICATION OF SYSTEM THEORY TO EMPIRES
This paper, in a sense, represents an attempt to put some old wine into a new bottle. The old wine is that heady system of imperialism as it was known and practiced in the middle of the seventeenth century. The new bottle is really a flask for fractionated distillation. By using the concepts and tools of system analysis to reexamine mercantile imperialism, I hope to isolate some of its more important factors, and hence to add to our understanding of the more significant systems and processes in international politics.

A few prefatory comments are in order at this point. First of all, by the phrase "mercantile system of imperialism" I am not referring to any single theory of imperialism, whether formulated by a contemporary of Hobbes or such twentieth-century political thinkers as Schumpeter, Lenin, or Norman Angell. Rather, a search of seventeenth-century history reveals persistent patterns and processes adding up to an analytical system that might be called "mercantile imperialism." This paper will look at the operation of that system—its principal actors, the variables governing the relationships of the actors, the structure of the system, and the sources of internal disequilibria.

By speaking of a "system" of mercantile imperialism, I do not want to imply that such a system did in fact exist in all of its ramifications during the seventeenth century or at any other single point in time. What I am suggesting is that the system is a paradigm for the operation of all imperialist systems that came into being after the great discoveries and explorations of the late fifteenth and early sixteenth centuries and that continued to exist, in some cases, down into the twentieth century. It was during the mid-seventeenth century, however, that the realities of state behavior came particularly close to matching the paradigm of mercantile imperialism.

By way of a final prefatory note, the system that I shall outline is based most immediately upon the British version of imperialism; but with only the most minor adjustments it could serve equally well as a model of French, Dutch, Spanish, or Portuguese practice during the same era. Similarly, the case study of imperial disintegration that I shall analyze—the breach of the imperial tie between Great Britain and the American colonies—has its parallels in the nineteenth and twentieth centuries.

THE SEVENTEENTH-CENTURY IMPERIAL SYSTEM

The environment of mercantile imperialism in the seventeenth century was the European state system. Europe in the 1650s, after the Peace of Westphalia, consisted in a number of more or less independent actors: absolute monarchies, monarchies with constitutional limitations, confederations, republics, principalities, and even duchies. The two actors with claims to universal roles within the system—the Holy Roman Empire and the Catholic Church—had lost their effectiveness in more than half of the system. Only one non-European actor, the Ottoman Empire, played a role in the European state system. But even this actor was perceived as an outside threat to the system rather than as a full-fledged member, subject to the same rules of behavior and principles of treatment as were the European actors. The actual structure of the European state system need not concern us here. I might merely note that, although the system was in a state of flux, it tended toward what might loosely be termed a "balance of power" instead of hegemony by a single actor. In large part it was system-dominant rather than subsystem-dominant.

As early as the fifteenth century a number of actors were going outside the European state system—trading, colonizing, exploiting—in efforts to enhance their position within the system. During the course of the next two centuries they created new systems, called "empires," with new actors and rules of conduct. In part the rules governing the behavior of actors within the European state system were transferred to the individual imperial systems. But, strictly speaking, each metropolitan country was able to determine its own relationship with its colonies, as well as the structure of its own imperial system. Hence these empires were subsystem-dominant rather than system-dominant.

If we analyze the individual imperial systems created by the European actors, we find a number of structural and behavioral regularities. Together these regularities—actors, variables governing the relationships of the actors, structures, and sources of instability—comprise the mercantile IMPERIAL SYSTEM.
The Actors in the IMPERIAL SYSTEM

The mercantile IMPERIAL SYSTEM comprises two subsystems—a DOMINANT SUBSYSTEM on the one hand and one or more DEPENDENT SUBSYSTEMS on the other. The DOMINANT SUBSYSTEM is the Metropolitan Country. We may consider the Metropolitan Country as a single actor, although clearly it, too, has subsystems. In the political field alone such subsystems include the chief executive officer of the government (for example, the monarch or the prime minister), a parliament or some other forum for the representation of the nobility and commoners, a cabinet or other advisory board, a board of trade or minister responsible for imperial matters, and enforcement and judicial officers. For the most part we shall ignore the subsystems of the DOMINANT SUBSYSTEM, concentrating upon the latter as a single actor in the IMPERIAL SYSTEM.

The DEPENDENT SUBSYSTEMS are the colonies founded or otherwise acquired by the Metropolitan Country. Geographically, the colonies may be contiguous to the Metropolitan Country (as the eastern provinces were to Germany in the twelfth century, or as the frontier settlements were to the American Republic in the nineteenth century) or—and this is the more likely possibility in the seventeenth century—they may be non-contiguous or overseas colonies. Like the Metropolitan Country, the DEPENDENT SUBSYSTEMS have their own social, political, economic, and other subsystems. Except in one important respect, however, we shall treat the DEPENDENT SUBSYSTEMS as single actors in the IMPERIAL SYSTEM. The exception to this general rule will be the occasional necessity to differentiate between two residential groups within the DEPENDENT SUBSYSTEMS: the colonists and the colonial administrators coming from the Metropolitan Country; and the indigenous or native population in the colonial territories.

The Variables Governing Relationships Among the Actors

The first variable that will be considered in this paper is the direction of the flow of policy decisions among the actors. By the phrase "policy decisions" I mean control over the key values operative in the IMPERIAL SYSTEM: the distribution of power and income, the establishment of principles of legitimacy, the use of force, the allocation of scarce resources, the distribution of honors and other symbols of social status. In considering the relationship between two actors in the IMPERIAL SYSTEM, we shall essentially be asking the question: who influences whom, over what ranges of action, using what means, and with what effect?

The second variable is the direction of communication transaction flows among the actors. This variable includes patterns of attention and migration, a division of labor, the flow of trade, the flow of mail, and other means of interpersonal communication—in short, the entire set of facilities for and habits of communication. The important question here is the extent to which the individual actors in the IMPERIAL SYSTEM direct their attention to other actors within the SYSTEM, as opposed to the extent to which they direct their attention to actors outside the SYSTEM.

Third, what is the direction of loyalties and group identification within the IMPERIAL SYSTEM? In large part this is the question of a sense of community: is it SYSTEM-wide, or does each actor have its own sense of community, not identifying psychologically with any other actor in the SYSTEM? A sense of community comprises, first of all, an awareness—both by members of the group itself and by actors outside the group—that the group does in fact exist as a separate entity, distinct from other such groups, and that its members are in some ways interdependent. When a group awareness is present, members as well as outsiders will use specific collective terms in referring to the group. Second, there is an internalization of group interests: a recognition that certain events or processes are of common interest to the members of the group; a certain degree of probability that the members will be able to coordinate their behavior in a way that will promote their common interests (and even to coordinate their activities to extend beyond those of immediate necessity to the group); and the presence of certain structures or processes to perform certain functions in the group interest. What we are asking here is whether or not the DEPENDENT SUBSYSTEMS identify themselves, their interests, and their fates with those of the DOMINANT SUBSYSTEM. And, conversely, we want to find out whether the latter reciprocates this sense of community. In its use of collective symbols, for example, does the Metropolitan Country seek to differentiate itself from its colonies? Or does it seek to tighten the bonds of loyalty and community?

The Structure of the IMPERIAL SYSTEM

Since the DOMINANT SUBSYSTEM is in effect the creator of the IMPERIAL SYSTEM, that is, since it provided the impetus for the establishment of the IMPERIAL SYSTEM, the DOMINANT SUBSYSTEM determines what the IMPERIAL SYSTEM will be. The Metropolitan Country is primarily self-interested; its attention and communication patterns focus primarily upon itself, and little or no attention is paid to the IMPERIAL SYSTEM as a system. Nor does the Metropolitan Country pay much allegiance to or otherwise identify itself with the IMPERIAL SYSTEM. There is no perceived duty on the part of the DOMINANT SUBSYSTEM to maintain the stability of the IMPERIAL SYSTEM, except insofar as it is to the
The Relationship Between the DOMINANT SUBSYSTEM and the DEPENDENT SUBSYSTEM

It is the goal of the Metropolitan Country to make its colonies totally dependent upon it for their social, psychological, political, and economic rewards as well as for their models of behavior and taste. At the same time, the Metropolitan Country attempts to use the colonies to enhance its own welfare (and, presumably, its own power position in the contemporary international system, that is, the European state system) without, however, becoming a prisoner of the colonies by being too dependent upon their resources or trade.

The primary means to make the DEPENDENT SUBSYSTEMS actually dependent upon the DOMINANT SUBSYSTEM come under the heading of "mercantilism." Among the more important mercantilist policies are: excluding foreign shipping from trade with the colonies; enumerating commodities which the colonies may obtain from or ship to the Metropolitan Country only; encouraging the colonies to produce commodities needed (whether for consumption, finishing, or re-export) by the Metropolitan Country; discouraging the colonies from producing commodities exported by the Metropolitan Country.

In such a relationship policy decisions flow from the DOMINANT SUBSYSTEM to the DEPENDENT SUBSYSTEMS. The latter must look to the DOMINANT SUBSYSTEM itself rather than appealing to some broader concept of imperialism in order to secure changes in the policies and attitudes of the Metropolitan Country. Although their communication patterns (and particularly trading patterns) bind the colonies individually to the Metropolitan Country, the reverse is not true. The Metropolitan Country not only has a larger number of opportunities to enter into communication and trade transactions with actors outside the IMPERIAL SYSTEM, but is likely to utilize these opportunities as a way of precluding too great a dependence upon the colonies. Finally, the colonies pay their primary allegiance to the Metropolitan Country, although it is by no means clear that the Metropolitan Country reciprocates this demonstration of affection and loyalty.

The Relationship Between a DEPENDENT SUBSYSTEM and Another DEPENDENT SUBSYSTEM

A mercantilist policy operates to keep the colonies isolated not only from the colonies of other metropolitan countries (as well as those countries themselves), but also from one another. It is, therefore, unlikely that there are any strictly intercolonial policy decisions to be made. Minimal intercolonial communication and trade ties force the individual colonies to look to the Metropolitan Country for their rewards and for models of behavior, fashions in clothing and table manners, linguistic conventions and styles of speaking, standards of excellence in literature and the arts, methods of agriculture and production, the art of government, and so forth. This being the case, colonial administrators and colonists alike are extremely apt to identify themselves psychologically with the Metropolitan Country, and to pay it their primary allegiance. Identification with and loyalty to the individual colony is only secondary.

Thus the relationship between the DOMINANT SUBSYSTEM and the DEPENDENT SUBSYSTEMS is star-shaped, that is, the lines of decision making, communication, and loyalty run between the Metropolitan Country and each individual colony, with very few lines running between the colonies themselves. In fact, the colonies may be mutually jealous, fearing that other colonies will receive preferential treatment from the Metropolitan Country.

Destabilizing Elements in the IMPERIAL SYSTEM

The star-shaped, subsystem-dominant IMPERIAL SYSTEM that I have just described pictures mercantile imperialism as being essentially in equilibrium with itself. In fact, however, several sources of instability are inherent in the IMPERIAL SYSTEM, producing what might be termed a dynamic rather than a static model of mercantile imperialism and, unless checked, may lead to the breakdown of the model.

One important dysfunctional element in the IMPERIAL SYSTEM lies in the opportunities available to the individual colonies to end their isolation by communication and trade with one another. Such opportunities are particularly visible when the colonies are contiguous or, at least, closer to one another than to the Metropolitan Country in an era when means of transportation and communication are slow and generally irregular. And it is not unreasonable to expect that, once the different
DEPENDENT SUBSYSTEMS do communicate, they will find points of commonness in spite of their diversity, possibly even grievances against the DOMINANT SUBSYSTEM. Again, they might find that trade among themselves would serve their individual and joint interests better than trade patterns connecting them individually to the Metropolitan Country.

Similarly, in the absence of rigid controls by the Metropolitan Country over the activities of its colonies (especially in the area of commerce), the DEPENDENT SUBSYSTEMS may seek new relationships with actors outside the IMPERIAL SYSTEM. If the colonies are successful in their attempt to establish outside contacts, the effect is to give the colonies a measure of control over their own policy decisions. They may choose, for example, to trade with countries with whom their terms of trade are better rather than to send their produce dutifully to the Metropolitan Country. The cost of stopping such practices could place tremendous burdens upon the DOMINANT SUBSYSTEM. Efforts by the Metropolitan Country to place its colonies under more rigid controls would probably meet at the very least by resistance and evasion, and possibly even by rebellion.

The preoccupation of the DOMINANT SUBSYSTEM with its own affairs is another source of instability in the IMPERIAL SYSTEM. A lack of responsiveness to the needs and desires of the DEPENDENT SUBSYSTEMS, or a simple disinterest in their development, may lead to an unawareness of disaffection and the possibility of growing hostility. Such a policy of drift encourages the colonies to seek ties with one another and with actors outside the IMPERIAL SYSTEM. And, by the time that the Metropolitan Country does turn its attention to the colonies, it may already be too late to prevent the disintegration of the IMPERIAL SYSTEM as a whole. In overcompensating for its past self-preoccupation, the DOMINANT SUBSYSTEM may accomplish little more than to drive a wedge between itself and the DEPENDENT SUBSYSTEMS.

Turning to the colonists and the colonial administrators in the DEPENDENT SUBSYSTEMS, we find another possible source of instabilities. On the one hand, the mass of colonists often comprises the disaffected elements of the Metropolitan Country's population. If migration is entirely voluntary, the ranks of the colonists may be filled with communicants of persecuted minorities, politicians in disfavor, social and economic failures, criminals and revolutionaries escaping possible punishment. Then, too, there will be an abundance of fortune-seekers, spinners of dreams of glory that they hope to fulfill in a new world. If migration is not voluntary, the ships may well be filled with such social undesirables as debtors and convicted criminals. In neither case are the colonists likely to be bubbling with affection for the Metropolitan Country, imbued with the spirit of the IMPERIAL SYSTEM.

On the other hand, colonial administrators often contribute to the instability of the IMPERIAL SYSTEM. It is not uncommon for the Metropolitan Country to banish its unwanted or obstreperous politicians to the colonies, there to sulk or to build their own satrapies of power as they see fit. However shortsighted a policy it may be, retaining the best officials and politicians for service in the Metropolitan Country is understandable. But even worse for the stability of the IMPERIAL SYSTEM is the petty tyrant, the incompetent and bumbling member of the lesser nobility, perhaps, who sees the power and authority of a colonial governor as his chief goals in life. Instead of spending his time conspiring to regain his position in the political subsystem of the Metropolitan Country, thereby ignoring his administrative duties and subjecting the IMPERIAL SYSTEM to a policy of drift, he gratifies his own ego by the overexuberant and even despotic exercise of his authority, driving the colonists into a search for new patterns of communication and loyalty.

In focusing upon the colonists and colonial administrators as destabilizing elements, I do not mean to suggest that every one of them fits into the categories discussed in the previous paragraphs. Indeed, far from it, for most might well be loyal, conscientious, and dedicated to the service of the Metropolitan Country. What I am suggesting is that the IMPERIAL SYSTEM has an inherent bias in favor of instabilities of this sort. It would be the exceptional Metropolitan Country that would send its best citizens and administrators to the colonies. Such a conscious and alert policy was not prevalent in the seventeenth-century IMPERIAL SYSTEM.

Still another dysfunctional element is the relationship between the colonial administrators and the colonists on the one hand and the indigenous population of the colonies on the other. There are a number of approaches that the DOMINANT SUBSYSTEMS (and, by implication, the colonial administrators and colonists) may take toward the native population: eliminating them (unintentionally, through the introduction of new and fatal diseases, or deliberately, as Hitler sought to exterminate large portions of the Eastern European population in the early 1940s); enslaving them, either outrightly or, as the British attempted in central Africa, by forcing them to pay taxes (for "civilizing" improvements such as roads and a government) with money that could be earned only by working as unskilled laborers for the colonial administrators; ignoring them or treating them as nuisances (much as the colonists in America dealt with the Indians), meanwhile keeping them generally ignorant and uneducated in the ways of self-government; putting them on the path toward social mobilization and at least local self-government;
treated them as equals with whom even inter-marriage is considered acceptable. But each of these approaches creates its own problems for the colonial and native populations of the colonies; and none contributes to the stability of the IMPERIAL SYSTEM over the long run.

THE COURSE OF EMPIRE

With such dysfunctional elements at work, it is not surprising that seventeenth-century imperialism—the state of reality most closely approximating the paradigm of the mercantile IMPERIAL SYSTEM that I have described—degenerated more or less rapidly into newer and generally more complex types of systems. It might be useful at this point to characterize briefly some of the ways in which these seventeenth-century empires changed.

Attempts to Maintain the IMPERIAL SYSTEM

In some cases metropolitan countries have found it necessary to use every and any means at their disposal to prevent the disintegration of their own imperial systems. Such policies include: jealously-guarded dominance over the policy-making for the colonies; short tours of duty and rapid rotation for colonial administrators to prevent them from forming ties of community in the colonies; the encouragement of one-crop economies and other measures designed to make the colonies economically dependent upon the mother country; strict controls over internal migration in the colonies, particularly with respect to the indigenous population; conscriptive labor forces; the retention of low literacy and mobility rates for the native population, as well as a near-subsistence level of income; close supervision over the colonists migrating from the mother country.

Such an inflexible policy viewing imperial relationships as static rather than as dynamic, and holding that the stability of the relationships rests primarily upon the firm application of measures of control over the colonies, may be successful at certain times in some places. The history of Portuguese policies toward Angola and Mozambique provides the world of the 1960s with cases in point. But over the long run, I would suggest it is an attitude that seems most likely to produce the structural tensions that eventually break out into open rebellion and nights of the "long knife."

Shifts to System-Dominant Pluralism

It is possible that, as the DEPENDENT SUBSYSTEMS grow in population, levels of technology, and skills of government, they will become co-equal with the DOMINANT SUBSYSTEM in decision making for the IMPERIAL SYSTEM as a whole. Carried to an extreme, as in the Commonwealth today, each of the actors within the IMPERIAL SYSTEM would have control over its own policy decisions, with none having the ability or right to impose its will upon the others. The communications net within the SYSTEM would be more or less symmetrical, with each actor directing more of its attention to other actors within the IMPERIAL SYSTEM than to outside actors. With respect to trade and patterns of production, however, a symmetrical relationship need not predominate. Indeed, it might prove most advantageous to all if a system-wide division of labor were instituted. Whatever the structure of the relationship, the cardinal principle will be that each subsystem must benefit more by remaining a member of the IMPERIAL SYSTEM than it would if it dissociated itself from the SYSTEM. As long as this were the case, each of the subsystems would place a high value on the maintenance of the SYSTEM. In consequence, each actor would direct a fair share of its allegiance to the SYSTEM, even though its primary loyalty might be self-directed. The formerly DOMINANT subsystem might hold a position as primus inter pares, possibly even retaining the ties of affection previously linking it to the DEPENDENT SUBSYSTEMS, but it would no longer direct the decision-making processes of the IMPERIAL SYSTEM.

The Political Integration of the IMPERIAL SYSTEM

The structure of the IMPERIAL SYSTEM may degenerate from a dominance-dependence relationship among the actors to one amalgamating the actors upon the basis of equality into a single, integrated system (such as a federation or a confederation). In such a process we might anticipate that the political subsystem of the Metropolitan Country would undergo the least amount of structural change, and that the former colonies would adapt themselves to accommodate the structure of the Metropolitan Country. It is, of course, also possible that mutual structural changes will ensue. Or the value of a given colony may be so great—if, for example, large deposits of scarce resources are to be found in the colony—that the Metropolitan Country would be willing to pay a price in terms of structural changes or other concessions to entice the colony into a system-dominant union.

A number of conditions render it more or less difficult to form an amalgamated union between the DOMINANT SUBSYSTEM and one or more DEPENDENT SUBSYSTEMS. Political integration may be easier if the actors are contiguous to one another. Examples of such cases include twelfth-century Germany, which successfully absorbed the eastern provinces, and nineteenth-century America, which incorporated the newly-opened western territories into the Union as states. That the same process is conceivable for non-contiguous areas is demonstrated by the accession of Alaska and Hawaii to the American Union in 1959. Most imperial powers, however, have found it difficult if not completely impossible to integrate and assimilate their overseas colonies.
If there are sharp ethnic or racial differences between the native population of the colonies on the one hand and the population of the Metropolitan Country on the other, the problems of assimilation might be quite complex. A successful effort on the part of the Metropolitan Country to inculcate a common group loyalty in the native population or, conversely, the success of the latter in infusing their own values into the societal system of the Metropolitan Country could facilitate the process of assimilation. Clearly, however, the assimilation of an ethnically- or racially-differentiated native population into that of the DOMINANT SUBSYSTEM depends upon the rate of social mobilization among the natives, and upon the efforts of the Metropolitan Country to accelerate the process of social mobilization.

The Disintegration of the IMPERIAL SYSTEM

A fourth type of structural degeneration is the outright disintegration of the SYSTEM. This might result from any one of the three policies just mentioned, particularly if the policy were to be inflexibly or, perhaps, prematurely applied and implemented. But disintegration might also result from a policy of drift or any of the other dysfunctional elements in the SYSTEM. It is even conceivable that the DOMINANT SUBSYSTEM is the actor most anxious to do away with the SYSTEM. This might be the case if the Metropolitan Country valued the structural integrity and solvency of its own subsystems more highly than the maintenance of the SYSTEM as a whole. President De Gaulle's ultimate disposition of the Algerian question seems to be an example of such an attitude.

Regardless of the cause, however, the disintegration of the IMPERIAL SYSTEM ends in the creation of new and separate systems. Each of the DEPENDENT SUBSYSTEMS may become a system in itself; a few of them may retain their ties of dependence to the Metropolitan Country; several of them may regroup themselves into a new system; others may align themselves with actors outside the IMPERIAL SYSTEM. The former colonies may establish friendly or hostile relations with the Metropolitan Country. They may retain their ties of trade and communication with the Metropolitan Country, or seek new markets for their produce and attention. In short, the aftermath of empire may be, and indeed has been, quite varied.

THE DISINTEGRATION OF EMPIRES:
A CASE STUDY

In the ensuing paragraphs I shall examine in more detail a case study of the mercantile IMPERIAL SYSTEM—the relationship of the American colonies to Great Britain—and its disintegration into two separate and, for several decades at least, mutually antagonistic systems. The first question that must be answered is whether or not the realities of the seventeenth-century Anglo-American empire in fact match the most essential aspects of the paradigm that I have been discussing.

The Anglo-American Empire as an IMPERIAL SYSTEM

The seventeenth-century American colonies existed in a state of semi-isolation, separated from one another in many cases by stretches of uninhabited wilderness and more generally by inadequate means of intercolonial transportation and communication. Contacts with the mother country were often easier to maintain, and perhaps more fruitful, than those with neighboring colonies. To the extent that there was any coordinating among separate colonial administrations, it was the result, not of the colonists' cooperation, but of the efforts of His Majesty's Government in England. Even as late as the middle of the eighteenth century the colonists were unable to organize an effective intercolonial defense against marauding Indians on the western frontiers; and some voices expressed fears of armed conflict among certain colonies.

In the area of policy decisions, the seven­teenth-century colonies underwent several changes. "All but a very few of the English colonies in America were begun as private enterprises. Some were started as joint-stock trading companies, others by feudal proprietors, and still others by groups of private individuals unauthorized by the king." (1) Although decision making was diffused during their early years, "gradually private ownership gave way to royal control; charters were surrendered or annulled; unsanctioned settlements were absorbed by their duly constituted neighbors. Finally only a handful of colonies lay outside the direct administration of the crown." (1) The colonies generally retained control over purely local matters, but the political subsystem of the mother country determined the most important policies distributing power and income, allocating resources, directing the use of force, establishing principles of legitimacy, distributing honors and other symbols of social status. The flow of decisions went from the mother country to the colonies, and not the other way around.

The nature of the communication ties between Great Britain and the colonies cannot be so sharply delineated. One reason for this is the unavailability of certain types of information, particularly quantitative political, social, and economic data. We do know some basic facts, however, from which we can extrapolate the patterns of communication transactions between Great Britain and her colonies in America.

In the first place, a very high percentage of the colonists were of English extraction. In 1790
about 61 per cent of the more than three million
white Americans had family names that could be
identified as English (and another 18 per cent had
Scotch or Irish family names) (2). If we accept
these population estimates as accurate, and re­
member that European political struggles and
shortages of food during the middle of the eight­
teenth century drove significant numbers of Scots,
Irish, and Palatine Germans to American shores,
then we might well assume that the colonial popu­
ation of the previous century (estimated at 50,000
in 1650 and at 251,000 fifty years later (3)) was
even more homogenous and more predominantly
English in origin.

English life and manners provided the colo­
nists with a model of social behavior that they
then sought to adapt to their existence in the wil­
derness. With their common English (or at least
British) heritage, the seventeenth-century colonists
looked to the Mother Country for their fashions
and standards of excellence, their customs and
language, their methods of agriculture, production,
and local self-government. Wealthier colonial fam­
ilies sent their children to England to be schooled;
and study at an English university was widely con­
sidered a sine qua non for would-be laywers and
medical practitioners, as well as for many men
of the church. London—and not Boston or Phila­
delphia, New York or Williamsburg—was the bright
star in the colonists' firmament. It was England
that the American colonist wanted to visit once
more before he died. So great was this sense of
attachment in some instances that writers in South
Carolina occasionally referred to England as home
even in the early 1760s (4).

Available data on the trade patterns of the
seventeenth-century American colonies suggest a
picture consistent with classic mercantile doctrine
in its main lines if not in all of its details. Mar­
garet Shove Morriss tells us, for example, that
during the years from 1689 to 1701, an estimated
75 ships per year passed between England and
Maryland (with about half of them going to or com­
ing from London), that "the trade to foreign coun­
ctries, on the whole, was extremely small;" and
that about 25 ships engaged in the intercolonial
trade stopped at Maryland's ports each year (5).
And, regarding the last four decades of the seven­
teenth century, Bernard Bailyn writes that "the
heart of New England's commerce was mercantil­
istically sound." Bailyn adds that, by the middle
of the 1670s, the New England merchants recog­
nized that their "economic activities were now at
the mercy of diplomatic decisions made by Eu­
ropean statesmen motivated by considerations utterly
unrelated to New England or its merchants." (6).
This is not to say that the mercantile relation­
ship of the colonies to the mother country during
this era was strictly regulated; for, due among
other things to England's involvement in European
wars, the mother country had neither the time nor
the material resources needed to control every
aspect of the colonists' trade. But, in the main, the
imperial relationship suggested previously (page 93,
figure 1) was ascendant in the seventeenth century.

Drift, Dynamic Ecology, and Shifting Colonial Per­
ceptions

Lest it seem that the imperial relationship
between Great Britain and the American colonies
was approaching a steady state of equilibrium by
the beginning of the 1700s, several sources of in­
stability must be noted.

The first of these was an English policy
toward the colonies that can only be described as
"drifting." By 1680 His Majesty's Government in
London had decided upon its principles and poli­
cies of colonial rule, as well as upon the means
of implementation and administration. In brief,
the principle was government by instruction, with
the colonial governor solely responsible for carry­

ing out the instructions sent to him by the Board
of Trade. There is much to commend such a re­
lationship, perhaps, but it was also conducive to
practicing an eighteenth-century version of the
"law of least effort." After observing that "the
commissions and instructions to the governors
underwent no fundamental change" during the hun­
dred years after 1680, Leonard W. Labaree points
out that "at bottom the system was fixed, static,
and unchanging, an expression of what the Board
of Trade loved to call 'the true principles of a
provincial constitution.'" (7) Such "true princi­
ples" did not, however, take into account changes
in the colonies and the imperial relationship, and
the changing needs, perceptions, and demands of
the colonists.

Thus, with what they thought was a resolu­
tion of the colonial issue, the British went back
to their own political problems and court intrigues,
and to the complications of the power struggle in
the European state system. As long as the colonial
relationship continued to function adequately, they
saw no need to concern themselves overmuch with
colonial affairs; and they were loathe, given the
dearth of able administrators in the British Isles,
to send good men out to the provinces.

While the Royal Government was following a
policy of drift, the colonists were developing ways
of life and patterns of self-government that began
to vary widely from the English model. One such
pattern was the increasing power of the colonial
assemblies, stemming from their right to initiate
legislation and their control of the purse. Although
the Board of Trade sent fairly uniform instructions
to the colonial governors, it left them to find their
own means of implementing the instructions. Thus
each governor had to fight his own battle with the
assembly in his colony. He had to convince the
legislators, first of all, that certain measures pro­
posed by the Crown were necessary and, second,
that the assembly should appropriate the requisite
monies. As Samuel Eliot Morison and Henry Steele
Commager have written,
colonial politics are largely the story of struggles between the assemblies and the royal or proprietary governors. In such a conflict between colonists with a lively sense of their rights and interests, and the representatives of a central government with not so keen a sense of imperial needs or English interests, and which did not wish to be bothered, the colonists generally won. The system worked well enough, for the British Government by veto or disallowance was able to prevent things, such as abuse of paper money, that it did not like; but it was unable to get positive things done, such as full co-operation in time of war. (9)

The net result of these processes was the increasing power of the colonial assemblies to make policy decisions over an ever wider range of activities, and hence a diffusion of the decision-making function in the IMPERIAL SYSTEM.

Meanwhile the ecology of the colonists' perceptions of themselves, the mother country, and the Anglo-American empire as a whole—that is, the relationship of the perceptions to the environment in which they occurred—was undergoing gradual but vital changes. On the one hand, Great Britain was assuming a less important role in the lives of the colonists while, on the other, the number and scope of intercolonial communication transactions of all sorts were growing.

Illustrative of these developments is the shift in colonial trading patterns. The number of ships plying between the ports of New York, Philadelphia, Hampton, and Charleston in America and harbors in the British Isles doubled from 1734 to 1772 (increasing in number from 264 to 556); but the number of ships engaged in the coasting trade quadrupled (from 402 to 1750) during the same period. Comparable figures for the Port of Boston are even more dramatic: the number of ships sailing from Boston to the mother country rose from an average of 48 per year in the 1714-1717 period to 59 per year in the four years from 1769 to 1772 (an increase of 23 per cent); the number of coastal vessels jumped from 117 to 451 (an increase of 286 per cent!). In terms of the total annual tonnage shipped from Boston, 19 per cent (3,985 tons) went to Great Britain or Ireland in 1714-1717, while 16 per cent (6,171 tons) of the yearly tonnage did so in 1769-1772. The share of the total tonnage shipped each year from Boston to other colonial ports rose from 17 per cent (3,583 tons) in 1714-1717 to 43 per cent (16,766 tons) in 1769-1772 (9). In short, while the shipping facilities of the colonies expanded generally during the course of the eighteenth century, intercolonial trade grew at a much more rapid rate than did trade with the mother country.

Other indicators of colonial communication patterns during the first six decades of the eighteenth century point to a similar increase in intercolonial communication (10). The expanding American population began to fill in the gaps separating the urban clusters scattered along the Atlantic seaboard, establishing a fairly continuous line of settlement from Penobscot Bay in the north to Savannah in the south. And with the expansion of the population came the construction of postroads, ferries, and other means to facilitate intercolonial travel and communication. Intercolonial mobility, in turn, made increasingly possible the exchange of ideas among the colonists: religious organizations and movements (such as the "Great Awakening") spread throughout America; the newspapers increased in number, size, and scope of coverage, drawing to an ever greater extent upon intercolonial news sources; colonial printers, such as Benjamin Franklin, and other colonial merchants normally had extensive familial and business connections in several colonies; while lawyers, doctors, and men of science often traveled to or corresponded with their colleagues in other parts of America.

There were, to be sure, factors working against increased intercolonial cooperation and communication. Diverse and often unstable colonial currencies, as well as the virtual absence of intercolonial credit facilities, hampered trade and commercial relations. Conflicting territorial claims led to harsh words and occasionally even bloodshed. Religious factionalism and regional jealousies sowed the seeds of mutual antagonism. In spite of the importance of these divisive factors, however, they did not prevent the triumph of mutual interests among the colonists over their mutual hostilities when the time of decision arrived.

Significant shifts in the colonists' focus of attention and patterns of self-perception accompanied the political drift and changing ecology of the eighteenth century. Data derived from a quantitative analysis of place-name symbols (such as "London," "Williamsburg," or "Italy") appearing in the colonial press from 1735 to 1775 suggest several interesting trends (11).

While maintaining a fairly constant interest in the mother country, the colonists shifted their attention from European wars and other events outside the Anglo-American empire toward an interest in American news. Although symbols of place names located in the British Isles generally occupied about one-fifth of the newspapers' symbol space throughout the entire 41 years, their share declined sharply relative to the space given over to American symbols. From 1735 to 1762 about one in seven symbols referred to place names in the American colonies; during the next thirteen years almost one in three symbols did so. The years 1774 and 1775 found the colonial printers devoting more than half their symbol space to news of America.

A more detailed analysis of American symbols in the colonial press (12) suggests a second trend: a shift away from purely local interests to an awareness of events affecting the colonies as a whole, with attention paid by the newspapers to
symbols of colonies other than their own remaining essentially unchanged over the long run. News of the home colony was important throughout the 41 years from 1735 to 1775. It declined in importance, however, relative to intercolonial news. The salience of a collective concept in the colonial press, that is, the use of symbols referring to all of the American colonies as a single unit, was low until 1763. After that date such collective symbols comprised about one-quarter of the total number of American symbols in the newspapers.

The IMPERIAL SYSTEM in Transition

A changing perception of America's place in the Anglo-American empire—a change that had actually begun to develop in the late 1750s—assumed dramatic proportions in the early 1760s. I have already noted the increased emphasis in the press upon place-name symbols identifying all of the colonies as a single unit during these years. The question that we want to examine now is, what was the content of these symbols? In answering this question I shall consider only symbols in articles with American datelines and, for the sake of convenience, I shall divide the collective symbols into three groups: first, those specifically identifying the colonies as British or royal domain; second, symbols (such as "the colonies" or "the provinces") that emphasize the imperial relationship and omit any specific mention of the mother country; and, finally, those that implicitly or explicitly identify the area as American, separate from any British community.

By looking at the symbols linking the colonies to the empire as a ratio of symbols identifying the area as British, an interesting and, I believe, significant finding emerges. For every ten British-oriented symbols appearing in the press during the two decades from 1735 to 1754, there were nine empire-oriented symbols. This ratio almost reversed itself during the next five years. The number of empire-oriented symbols appearing for every ten British-oriented symbols during the remaining colonial years were: fifteen from 1755 to 1759; thirty-four from 1760 to 1764; forty-six from 1765 to 1769; and sixty-three from 1770 to 1775. Thus, over a twenty-year period, the ratio of empire-oriented to British-oriented symbols multiplied sevenfold. In effect, through their use of self-referent symbols, the colonists (or at least their newspapers) were substituting an allegiance to the IMPERIAL SYSTEM for their former ties of loyalty to the DOMINANT SUBSYSTEM, that is, Great Britain.

Such a finding accords well with what we already know about the changing political philosophy in the colonies during the 1760s. The colonists had begun to emphasize the heritage and rights that were common to both Englishmen in the British Isles and Englishmen in the American colonies; and they had begun to express their resentment toward perceived infringements of those "common" rights. The response to the Sugar Act of 1764 and the Stamp Act of 1765 are cases in point. As early as 1764 James Otis of Massachusetts referred to Harrington's assertion that "Empire follows the balance of property," and argued:

The Colonists being men, have a right to be considered as equally entitled to all the rights of nature with the Europeans [that is, the English], and they are not to be restrained in the exercise of any of these rights, but for the evident good of the whole community... (13)

There is more than just a statement of natural law in Otis' remarks. There is the implication that the "whole community" of the empire is somehow superior to the English in the mother country as well as to the colonists in America.

Within the next decade the idea of an IMPERIAL SYSTEM superior to its subsystems became common coin in the arguments of colonists. In 1774 James Wilson of Pennsylvania wrote of the doctrine of imperial equality that "the Commons of Great Britain have no dominion over their equals and fellow-subjects in America; they can confer no right to their delegates to bind those equals and fellow-subjects by laws." Was Wilson arguing in favor of the outright independence of America? To the contrary, he wrote of the Americans that "They are the subjects of the King of Great Britain. They own him allegiance." Of the "strict connection between the inhabitants of Great Britain and those of America," he wrote:

... They are fellow-subjects; they are under allegiance to the same prince; and this union of allegiance naturally produces a union of hearts. It is also productive of a union of measures through the whole British dominions. To the King is intrusted the direction and management of the great machine of government... He has a negative on the different legislatures throughout his dominions, so that he can prevent any regnancy in their different laws. (14)

The king, then, is representative of the IMPERIAL SYSTEM as a whole; and, while Parliament is the chief legislator for Great Britain, it is no more than the co-equal of the colonial assemblies within the IMPERIAL SYSTEM.

Concurrent with this changed attitude toward the structure of the IMPERIAL SYSTEM was an emerging group loyalty among the colonists themselves. We have already seen signs of this in their changing attention patterns. Throughout the 1760s and early 1770s the colonists found it increasingly possible to work together—in the Stamp Act Congress of 1765, in the committees of correspondence, and eventually in the Continental Congress—as well as to think of themselves as a community of fate. John Dickinson of Pennsylvania wrote in 1768, for example, that "every man
amongst us who in any manner would encourage either dissension, diffidence, or indifference between these colonies is an enemy to himself and to his country." (15) And, by July of 1775, the delegates to the Continental Congress could assert: "our union is perfect." (16)

By the middle of the 1760s, then, the colonists had evolved a perception of the IMPERIAL SYSTEM differing considerably from that of the seventeenth-century mercantilist. The changing context of colonial society was in part responsible for the newer perceptions. First of all, the colonies were no longer isolated from one another, almost solely dependent upon the mother country for their sources of rewards and models of behavior. Decision making for the SYSTEM as a whole was diffused, with the colonial assemblies enjoying an ever greater measure of freedom to determine their own policies, and with mounting opposition to attempts by Parliament to legislate for the colonies. Second, the lines of communication were stronger among the colonies themselves than between the colonies and the mother country. Third—and this is perhaps the crucial point—the colonists were dividing their loyalties. To an increasing extent, on the one hand, they were redirecting their ties of allegiance, away from the DOMINANT SUBSYSTEM, that is, Great Britain, and toward the IMPERIAL SYSTEM itself. There was to be sure a continued recognition of the superior role of Great Britain in the IMPERIAL SYSTEM, but the relationship that the colonists perceived was one of primus inter pares, and not one of dominance and submission. On the other hand, however, there was growing group loyalty among the colonists themselves, pointing the way to identification with an exclusively AMERICAN SYSTEM.

In brief, the 1760s found the colonists consciously or unconsciously pushing the IMPERIAL SYSTEM toward degeneration into what I earlier termed "system-dominant pluralism," with the British monarch as the titular head of the system. If the British government had perceived the Anglo-American empire in the same light as did the colonists, the result might have been the emergence of a British Commonwealth in the eighteenth century. But the British did not concur in the Americans' perception of a dynamic IMPERIAL SYSTEM. It was the clash between the British perception of static imperialism and the American perception of dynamic imperialism that was to lead to the eventual breakdown of the Anglo-American empire.

Thus the Anglo-American Empire of the 1760s was in transition. A century ago it had closely approximated the paradigm of the mercantile IMPERIAL SYSTEM. What it would become rested upon the attitudes and actions of both the British and the Americans.

**The End of Empire: The Clash of the 1770s**

As it turned out, the British persisted in maintaining their image of the mercantile IMPERIAL SYSTEM during the ensuing years. Overreacting, perhaps, to long decades of drift, Britain set about to bring its colonial relationships in order. The Townshend Acts of 1767 reorganized the customs service in the colonies, providing for more effective enforcement of existing regulations, and instituting new duties on colonial imports from Britain. The duties themselves were to be used for paying the salaries of the colonial governors, thereby relieving them from their dependence upon the assemblies. Instructions to colonial customs officers in 1769 revamped the existing trade regulations and sought to close some of the loopholes that had proved so advantageous to colonial merchants. The British response to the "Boston tea party" of 1773 was a policy of toughness: the Port of Boston was closed; extensive changes in Massachusetts' form of government were ordered; provision was made for transporting criminals to England for trial. But the policy of toughness backfired. Although it conceivably could have some effect in maintaining the SYSTEM during an earlier era, it merely succeeded in pushing the colonists to greater extremes in the 1770s.

Nor were the Americans ready to retreat from their minimum demands for dynamic imperialism. Expressions of the idea of Anglo-American co-equality in decision making under the aegis of the Crown appeared regularly throughout the early 1770s. In what might be termed the last major colonial effort to save the Anglo-American empire from disintegration, Joseph Galloway of Pennsylvania sought to reconcile the differing British and American perceptions by institutionalizing the American idea of empire within the British framework. In September of 1774 Galloway boldly suggested to the delegates of the first Continental Congress "That a British and American Legislature,
for regulating the administration of the general affairs of America, be proposed and established in America, including all the . . . colonies; within and under which government, each colony shall retain its present constitution and powers of regulating and governing its own internal police in all cases whatsoever." (17) But his efforts went for naught. By a majority of one vote the colonies rejected Galloway's plan of union, and in its place they drafted a series of far-reaching resolutions petitioning King George III and the English population to listen to the light of colonial reason and calling upon the colonists to refrain from trading with the mother country should the so-called "coercive acts" not be repealed.

By the end of 1774 the idea of complete independence from Great Britain seems to have taken hold of the colonists' thoughts. Anglo-American co-equality under an imperial monarch was no longer an acceptable solution for the Americans who wanted even greater freedom of action in policy making.

It is possible, I would suggest, to trace the trend toward thoughts of independence by looking at the colonists' changing perceptions of their country and themselves. The first of these changes, as indicated by the use of self-referent symbols in the colonial press, was a growing tendency to distinguish the American colonies from both Britain and the empire. For every ten British- or empire-oriented symbols appearing in American-datelined items during the twenty-five years from 1735 to 1759, less than six identified the colonies as "American." This ratio shifted sharply, however, after 1760. During the last sixteen years of the colonial era, sixteen symbols identifying the colonies as American were printed in the newspapers' columns for every ten symbols identifying them as a part of Britain or the empire.

It is interesting to note that this wave of what might be called increasing American separatism followed closely upon an earlier wave that sought to dissociate the colonies from Great Britain while associating them with the empire. In the usage of symbols at least, the transition from empire to independence took place in three steps: from alignment with the DOMINANT SUBSYSTEM to identification with the IMPERIAL SYSTEM itself to, finally, dissociation from both the DOMINANT SUBSYSTEM and the SYSTEM. Let me again stress the fact that these transitional steps took place in a context of increasing emphasis upon terms symbolically uniting all of the American colonies into a single political unit.

The second major change in the colonists' perceptions during the late 1760s and early 1770s pertained to their views of themselves. Before 1763 the use of symbols identifying all of the colonists as a single group was rare indeed: they comprised about seven-tenths of one per cent of the total number of symbols referring to American place names. Moreover, the newspapers were virtually unanimous in associating the colonial population with Britain.

The turning point came in 1763. From that year until 1769 the average issue of a newspaper contained eight times as many symbols collectively identifying the colonists as did the average issue from 1735 to 1762; and the use of such symbols increased by another 50 per cent during the six years from 1770 to 1775.

Once the collective concept had become salient, however, the colonists were quick to use self-referent symbols differentiating themselves entirely from their brethren in the mother country. Collective symbols referring to the population did not follow the pattern previously set by collective symbols referring to the territory—shifting from British-oriented symbols to empire-oriented symbols and finally to symbols of American separation. The middle stage in this transitional process never materialized. Rather, the transition from perceptions of the colonists as British to perceptions of them as Americans took place in one step. For every ten British- and empire-oriented symbols appearing in the press from 1763 to 1769, there were four American-oriented symbols; and from 1770 to 1775 there were eleven American-oriented symbols.

What I am suggesting is that the recognition of American nationality came as a step-level function for the colonists. The years immediately after the conclusion of the French and Indian War saw the rapid acceleration of a number of processes under way in colonial America since the turn of the eighteenth century: a drifting British imperial policy broken by sporadic attempts to recapture Britain's position of dominance in the Anglo-American empire; colonial efforts to widen the decision-making powers of their assemblies; increasing intercolonial trade and communication transactions of all sorts, in a context of fairly constant or even diminishing contacts with the mother country; the development of an interest in American events as well as a growing perception of the colonies as a single unit. The 1760s were years of transition, years in which people who called themselves "British-Americans" or "colonists" sought to reconcile their own desire for a greater measure of freedom with a sense of loyalty demanding the maintenance of the Anglo-American empire. At this stage, substantial British concessions in the area of policy making might even have prevented or postponed the disintegration of this empire.

By about 1770, however, the inhabitants of the thirteen colonies no longer considered themselves to be "British-Americans" or even "colonists." They were Americans, living on American—not British or imperial—soil. They were no longer trying to reconcile two differing perceptions of the IMPERIAL SYSTEM so much as they were
becoming aware of their basic opposition to that IMPERIAL SYSTEM.

The path that the Anglo-American empire followed during the 1770s is familiar to us all: the increasingly hostile postures of both the British and the Americans; the feverish activity of colonial patriots; the efforts of the British government to restore the dominance-submission relationship envisioned in the ideal IMPERIAL SYSTEM; the outbreak of fighting at Lexington and Concord; and, ultimately, the Declaration of Independence.

FIGURE 3: THE AFTERMATH OF IMPERIAL DISINTEGRATION

By 1776 the Anglo-American empire had disintegrated into two independent systems. Great Britain and the United Colonies had separate decision-making processes; trade and other communication transactions between the two countries had all but ceased; and the ties of loyalty and identification were self-contained in each system. Within the AMERICAN SYSTEM a confederal political structure was created—a structure with decentralized decision making in some important areas, but nonetheless adequate enough to conduct and win the War of Independence.

It is particularly interesting that the colonies, having declared their independence from the Anglo-American empire, were willing to unite in an American political community. In large part, I would suggest, this willingness stemmed from a strong sense of community or group loyalty—a group loyalty resting upon decades of mutually beneficial transactions of all sorts as well as the development of mutually compatible perceptions and attitudes. It is of course true that the events of the 1760s and the 1770s played a role in making the colonists conscious of their mutual interests and problems. But it is also true that rapidly expanding facilities for and habits of intercolonial communication paved the way for a situation in which the colonists could perceive and react to these events in a mutually compatible manner. On top of this interplay of communication transactions and external developments was built a superstructure of formal governmental structures and processes that, in their own turn, contributed to the emerging sense of American nationality.

SYSTEMS AND EMPIRES

The chief value of a paradigm for students of international politics lies in its explanatory power, in its helpfulness in trying to understand the structure or process that it seeks to picture. Such an "understanding" may take place on several levels of analysis. We might, for example, merely be interested in finding a paradigm of some system (mercantile imperialism, perhaps, or the twentieth-century international system) that will give us an intuitively sound notion of its operation during a given period of history or under certain conditions. If this paradigm will help us to understand or at least to analyze specific examples of such systems in practice, and particularly if it will help us to explain deviant cases, that is, examples of the system that do not conform to the paradigm, then it will be of even more value. Finally, we might be interested primarily in finding a paradigm that will enable us to predict the future development of the system in question.

In seeking to understand such a system as modern or seventeenth-century imperialism, a main problem is finding means to make its key variables operational. By that I mean, how can we find (quantitative) measures of the flow of communication and attention within a system? Or, how can we find (quantitative) measures of patterns of identification? A start has been made in this direction, concentrating especially upon examples of modern empires and other state systems. To operationalize key variables in the Anglo-American empire of the seventeenth and eighteenth centuries, however, poses a different and more complex task. Many of the quantitative political and social data that we need for analyzing significant trends in colonial America are neither accurate nor complete enough to be of much value. In this respect, as I tried to point out in this paper, some of the newer research concepts and tools of the social sciences offer possibilities at once fruitful and exciting. We may hope that the application of these ideas and methods to such systems as imperialism will enhance our knowledge and understanding of their operation.
REFERENCES


4. See The South-Carolina Gazette, 31 October 1761.

5. MORRISS, MARGARET SHOVE, Colonial Trade of Maryland, 1689-1715 (Baltimore: The Johns Hopkins Press, 1914), pp. 86-87, 108-109, 110-113. The last set of figures has been revised to exclude ships listed as sailing only between Maryland and England or the West Indies.


9. The remainder went to the Caribbean and other colonies in the New World (58% in 1714-1717 and 38% in 1769-1772) and to Europe and Africa (6% in 1714-1717 and 3% in 1769-1772). Computed from data given in U.S. Bureau of the Census, Historical Statistics of the United States, op. cit., Series Z 56-75, pp. 759-760. The earlier figure for the port of Hampton is 1733 rather than 1734.


11. The entire contents of four issues per year of the Massachusetts Gazette (from 1735 to 1775), the Boston Gazette, and Country Journal (from 1782 to 1775), the New York Weekly Journal (from 1735 to 1751), the New York Mercury (from 1752 to 1775), the Pennsylvania Gazette (from 1735 to 1775), and the Virginia Gazette (from 1736 to 1775) were systematically analyzed. For further details, see Richard L. Merritt, Symbols of American Community, 1735-1775 (publication forthcoming).

12. In tabulating the distribution of American symbols only, both direct place-name symbols ("Virginia" or "Boston") as well as indirect references to such place names ("this colony" or "this city") were included. In the comparison of the total number of American place-name symbols with the number of British place-name symbols in the previous paragraph, such indirect symbols were not tabulated.


15. DICKINSON, JOHN, Letters from a Farmer in Pennsylvania to the Inhabitants of the British Colonies, letter xii; cited in Morison, Sources and Documents, op. cit., p. 53.

16. A Declaration by the Representatives of the United Colonies of North America, now met in Congress at Philadelphia, setting forth the causes and necessity of their taking up arms; cited in Morison, Sources and Documents, op. cit., p. 144.

G.

APPLICATION OF SYSTEMS THEORY TO WORLD PROBLEMS
I. INTRODUCTION

Some aspects of arms races have a structure similar to that of economic models of competition. This paper explores this similarity by using reaction curves to analyze and advance previous theoretical studies of arms races. In particular the classical ratio goal and Richardson models are extended by explicit consideration of alternative strategic assumptions and constraints on weapons holdings.

The limitations of reaction curve analysis must, however, be constantly borne in mind. The greatest shortcoming is undoubtedly the use of a single policy variable, namely aggregate stocks of weapons. The use of weapons stocks as a policy variable ignores certain factors that are crucially important in arms races: the diversification of weapons as a response to uncertainty and the possibilities of weapons developments and improvements. Other shortcomings of reaction curve analysis are the limitation to two contending nations and the assumption that the only influence on weapons stocks are the opponent's weapons stocks. Because of these shortcomings and other limitations the analysis and conclusions are presented not for direct application but rather as indicative of a path of analysis for, and conclusions that might be applicable to, the much more complicated actual situation. The analysis simplifies the situation without entirely losing its meaning; the conclusions hopefully will provoke further study.

II. CLASSICAL MODELS

The ratio goal and Richardson models are classical models of arms races. Both models make the simplifying assumption of two contenders, each having a choice of a single policy variable, levels of weapons stocks. These models and their extensions can therefore be analyzed by reaction curves.

Reaction curves in the theory of duopoly (two sellers of a good) indicate the optimum choice of output of one seller given the output of his competitor. The reaction curves used here to analyze arms races similarly indicate the optimum choice of weapons stocks given the weapons stocks of the opponent. Such reaction curves are similar to the lines of constant power of Ash and the lines of objective security of Burns.2

The classical examples of arms races, battleship building and army mobilization, were often studied by means of ratio goals.3 To the extent that the outcome of an encounter of fleets or armies depended on size ratios it was reasonable to assume that arms races were based on ratios of battleship tonnage or numbers of divisions. The attempt to obtain some ratio goal more favorable than parity (equality of weapons) on both sides, together with the assumption of some protective weapons, that is, a nonzero optimal weapons stock at zero opponent's weapons, are shown in Fig. 1

![Fig. 1 — Reaction curve: ratio goal](image)

where $W_A$ is the stock of weapons held by A and $W_B$ is the stock of weapons held by B. $AA'$ is the reaction curve for A, showing the weapons A desires to hold ($W_A$) as a function of the weapons...
MICHAEL D. INTRILIGATOR

This reaction curve shows that if B holds no weapons \((W_B = 0)\) there is a positive stock of weapons that A desires and that if B increases its holdings, A also increases its holdings. The ratio goal is shown as the dotted line through the origin. In attempting to meet the ratio goal, the reaction curve \(AA'\) comes closer and closer to this dotted line. The reaction curve shows how much A would like to hold (as a function of weapons held by B), not the actual combination of weapon stocks \((W_A, W_B)\). If, however, the actual combination differs from the desired combination, A will attempt to move the actual combination to the reaction curve. A succeeds in moving the combination by increasing or decreasing \(W_A\), hence the arrows show direction of movement.

The reaction curve \(AA'\) for A is combined with a similar reaction curve \(BB'\) for B in Fig. 2.

It is assumed that A and B both want protective weapons and have ratio goals. In addition to the dotted ratio goal for B there is shown a dotted parity line, along which holdings of weapons are equal \((W_A = W_B)\). The left and right arrows of Fig. 1, which show how A varies its actual weapons levels to attempt to reach the reaction curve, are supplemented in Fig. 2 by up and down arrows which show how B similarly varies its actual weapons levels. The result of combining these reaction curves is clearly an endless arms race spiralling to higher and higher weapons stocks, since A and B together will move an actual point above \(BB'\) or to the right of \(AA'\) into the central area in which weapons stocks on both sides increase.

The classical battleship building arms race would, however, yield a finite equilibrium configuration if the A ratio goal were above the B ratio goal, as shown in Fig. 3. At the equilibrium point \(D\) both A and B are simultaneously satisfied with their holdings of weapons. By the direction of the arrows shown it is clear that \(D\) is a stable equilibrium in that small movements away from \(D\) will give rise to forces restoring \(D\). The condition necessary to ensure an equilibrium such as \(D\) is that the product of the ratio goals be less than one. Assuming both sides have the same ratio goal the necessary condition is that each ratio goal be less than one, implying counter-intuitively a ratio goal for both sides that gives weapons superiority to the opponent.

The Richardson model\(^4\) represents an arms race by a set of differential equations, indicating for each country the effect of stocks of weapons in both countries on weapons production. Letting \(W_A = \frac{dW_A}{dt}\) and \(W_B = \frac{dW_B}{dt}\) represent the production of weapons in countries A and B respectively, the Richardson model is:

\[
\begin{align*}
\dot{W}_A &= a_1 W_B - a_2 W_A + a_3 \\
\dot{W}_B &= b_1 W_A - b_2 W_B + b_3
\end{align*}
\]

Each of the coefficients has an interpretation: \(a_1\) (and \(b_1\)) are "defense coefficients," indicating the influence of the opponent's weapon stocks on weapons production; \(a_2\) is a "fatigue and expense coefficient," indicating the burden of supporting weapons, and \(a_3\) is a "grievance term," including all factors other than weapons stocks responsible for weapons production. The first coefficient then represents military, the second economic, and the third political-social bases for arms races. All coefficients are assumed positive.

The Richardson model can be represented by means of reaction curves given by the combinations of weapons stocks \((W_A, W_B)\) yielding no change in weapons stocks:

\[
\begin{align*}
\dot{W}_A &= a_1 W_B - a_2 W_A + a_3 = 0 \\
\dot{W}_B &= b_1 W_A - b_2 W_B + b_3 = 0.
\end{align*}
\]

These lines of no change in weapons stocks are

---

SOME SIMPLE MODELS OF ARMS RACES

shown in Fig. 4. The line $\dot{W}_A = 0$ corresponds to the reaction curve $AA'$ and the line $\dot{W}_B = 0$ corresponds to the reaction curve $BB'$. The interpretation of these lines as reaction curves is clear from the arrows indicating direction of movement. An equilibrium point like that shown in Fig. 4 is reached if the slope of $AA'$ exceeds the slope of $BB'$, that is, if $a_1/a_2 \cdot b_1/b_2 < 1$. If such an equilibrium point exists it is stable, as is seen from the direction of the arrows. An unstable equilibrium might exist, however, if the grievance terms are negative. This situation of good will leads reasonably to a downward spiral of weapons for slight movements below equilibrium but ironically to an upward spiral of weapons for slight movements above equilibrium.

$$\begin{align*} \dot{W}_A &= 0 \\ A' \\ W_B &= 0 \\ B' \\ W_A \\ A \\ B \end{align*}$$

Fig. 4 — Arms race with equilibrium: Richardson model

The Richardson model with equilibrium is similar to the ratio goal model with equilibrium, as seen by a comparison of Figs. 3 and 4. Both models require that the product of ratios be less than unity. If symmetry is assumed in that the ratio goals or coefficients for both sides are the same, the ratio goal model requires for equilibrium that the ratio goal be less than one, whereas the Richardson model requires for equilibrium that the defense coefficient be smaller than the fatigue and expense coefficient. Similarly the Richardson model without an equilibrium ($a_1/a_2 \cdot b_1/b_2 < 1$) is similar to the ratio goal model without equilibrium (Fig. 2).

III. STRATEGY

Classical models of arms races have many shortcomings, including the use of a single policy variable and neglect of strategic uses of weapons and constraints on weapons stocks. The extension to several policy variables will not be undertaken here, but the implications of alternative strategies and constraints on weapons stocks will be considered.

Strategy must be considered explicitly for nuclear weapons. The naval encounters or military maneuvers of the past, in which ratios of weapons stocks were a major determinant of the outcome, have no counterpart in nuclear warfare. Nuclear warfare and hence nuclear arms races depend fundamentally on the assumed uses of nuclear weapons. In a deterrent strategy, for example, nuclear weapons could be used to deter the opponent from striking by threatening to inflict upon him unacceptable damage. In an arms depriving strategy, on the other hand, nuclear weapons could be used to disarm the opponent by threatening to destroy his nuclear weapons.* These alternative strategies lead to reaction curves that differ from those above because of scale effects of weapons according to which increasing the numbers of weapons proportionately for both sides aids a deterrer and harms an arms depriever. According to Schelling:

For anything like equal numbers on both sides, the likelihood of successfully wiping out the other side's missiles ("weapons") becomes less and less as the missiles on both sides increase. And the tolerance of the system increases, too. For small numbers on both sides, a ratio of 2 or 3 to 1 may provide dominance to the larger side, a chance of striking first and leaving the other side a small absolute number for striking back. But, if the initial numbers on both sides are higher, it may take a ratio of 10 to 1 rather than 2 or 3 to 1 to have a good chance of striking with impunity; neither side needs to panic if it falls behind a little bit, and neither has any great hope that it could draw far enough ahead to have the kind of dominance it would need.5

Scale effects therefore imply that the reaction curve of a deterrer will rise less than proportionately to the increase in opponent's weapons, and the reaction curve of an arms depriever will rise more than proportionately to the increase in opponent's weapons. Reaction curves for A and B acting as either deterrer or arms depriever are shown in Fig. 5. It is assumed that both a deterrer and an arms depriever would like to hold protective weapons when the opponent holds no weapons.

Strategy and scale effects lead to several conclusions. First, reaction curves cannot be analyzed by ratio goals or Richardson models because of scale effects. Second, if both countries

*Arms depriving is a counterforce strategy.
are arms deprivers there is no equilibrium point as in the case of the ratio goal or Richardson models with no equilibrium point, which results in an endlessly spiralling arms race. Third, if both countries are deterrers there is an equilibrium point (E) of mutual deterrence similar to the ratio goal or Richardson models with equilibrium point. This equilibrium point is stable in that any deviations about the point establish self-correcting forces restoring the equilibrium. Fourth, the situation of an arms depriver facing a deter­rer can lead to stable equilibrium points (F or G). Fifth, the most explosive situations are those indicated by shading in Fig. 5, in which one side has sufficient weapons to disarm its opponent, and the opponent has insufficient weapons to deter. Richardson's conclusion that increasing weapons stocks tend to increase the probability of war is qualified, if not reversed, by the fact that weapons serve as a deterrent. If both sides hold sufficiently large amounts of weapons neither side need fear that the other will initiate war with a disarming strike.

IV. CONSTRAINTS

Classical models of arms races neglected constraints on weapons stocks as well as strategy. Reaction curves indicate the tastes of policymakers but do not take into consideration the scarcity of resources and the existence of institutions that act as constraints on the policy­maker (budgets for example). To tastes, embodied in reaction curves, must be added resources and institutions, embodied in constraint curves. If the constraints are based on resources they are probably independent of the opponent's weapons stocks. If the constraints are based on institutions, however, they probably increase with the opponent's weapons stocks. For example, if the opponent adds more weapons, budgets are increased, pushing out the constraint curve.

Constraints are important only to the extent that they actually constrain. The stable equilibrium of mutual deterrence at E in Fig. 5 would be reached if E were within the constraints of both sides, in which case the constraints are not binding and tastes alone generate an equilibrium. Constraints will, however, become binding in the case of no equilibrium point, for example, in the case of two arms deprivers in Fig. 5. This case is shown in Fig. 6 in which A and B both act as arms deprivers with reaction curves of AA' and BB' respectively. The constraint curves C_A A' and C_B B' have been added to the reaction curves for A and B respectively.

Several results emerge from Fig. 6. First, constraints have done what the tastes of two arms deprivers could not do: establish an equilibrium point. To the extent that each side attempts an arms depriving policy they inevitably push weapons stocks to the maximum permitted by the constraints, point H. H is a stable equilibrium to the extent that it is reached via the process of interacting weapons acquisitions and would be returned to if there were any deviation from H. It is not an equilibrium in the sense that both sides are satisfied; neither side has been able to obtain sufficient weapons to disarm the opponent. Second, H is desirable in that neither side need fear being disarmed by the opponent. Third, this property is true not only of H but also of all points above and to the right of I (the shaded area). Any point in the shaded area has the property that, given the constraints, the opponent could never attain adequate weapons for an arms depriving strategy. To this extent it appears plausible that after the contenders have reached H and realized that neither side could attain adequate weapons to disarm the opponent, in their mutual self-interest they would

---

6. Kenneth Boulding in *Conflict and Defense*, Harper, New York, 1962, refers to these constraint curves as "maximum home strength boundaries" based on "resources, will-to-strength, and so on" (pp. 234-241). However, he inexplicably considered only the cases in which constraints are either independent of or decrease with opponent's weapons stocks.
each disarm voluntarily down to I. I and H are alike in that they ensure that one side will not be the victim of an arms depriving strategy of the opponent, but I is a point at which both sides hold less weapons than H and hence the costs of holding weapons, the probability of accidentally firing a weapon, and so on, are reduced.7 Of course both sides would be reluctant to let weapons stocks fall below I. If A chose a $W_A$ to the left of I then B could, within the limits of its constraints, attain the reaction curve and thus have sufficient weapons to disarm A. Similarly B would never choose a $W_B$ below I. In particular, the point of general and complete disarmament, the origin, is dangerous in that either country could easily attain enough weapons to use an arms depriving strategy successfully.

The case of two deterers with binding constraints is illustrated in Fig. 7. The attempt to build up weapons to deter the opponent leads to an arms race that terminates at J at which there is a stable but uneasy equilibrium. The equilibrium is uneasy because it indicates that neither side has enough weapons to deter the opponent so that J is a point of fears, anxiety, threats, and bluffs. Again there is little to recommend J over K: both indicate that neither side has enough weapons to deter the opponent and that the opponent cannot attain deterrent capability within the confines of the constraints. The entire area between J and K is, however, one of mutual fears and distrust.

The case of a deterrer (A) and arms depriv- er (B) is illustrated in Fig. 8. If both constraints are operative, L is reached at which A cannot deter B and B cannot disarm A. If the constraint on A is nonoperative (for example, shifted out beyond A’s reaction curve) then M is reached at which A deters B but B doesn’t disarm A. Similarly if the constraint on B is nonoperative, then N is reached at which A doesn’t deter B and B disarms A.

![Fig. 7 — Arms race: deterers with constraints](image)

V. CONCLUSION

Some of the limitations of the classical treatments of arms races, the ratio goal and Richardson models, have been overcome by rephrasing the models by means of reaction curves and extending the models to allow for strategy and constraints. Allowing for deterring and arms depriving strategies leads to a point of stable mutual deterrence and to a delineation of explosive situations. Allowing for constraints on weapons stocks leads to non-explosive stable equilibria. There are still, however, many severe limitations to this type of analysis. The diversity of weapons, and, further, the many dimensions of policy, have been ignored here. Their treatment will have to await a further analysis of arms races.

7. The shift from H to I based on effective communication to each side of the constraints relevant to the opponent is similar to the shift in non-zero-sum games, such as the Prisoner’s Dilemma, from certain payoff configurations to other strictly dominating (for both) configurations. See R. D. Luce and H. Raiffa, *Games and Decisions*, John Wiley & Sons, New York, 1957.
A GAME FOR STUDYING THE PROBLEMS OF ARMS CONTROL*

R. H. Davis, P. B. Carpenter, and C. W. Missler

ABSTRACT

A game for the study of problems in the area of arms control is described. Two or more teams of "experts" possess unique military force structures and aggregated supporting economies. In managing their military forces and supporting economies a nation's leaders (game "experts") have only limited information about their opponents which is obtained by means of formal and informal intelligence channels. Given a particular disarmament agreement, the effects of various treaty provisions on the decisions of participants are studied.

INTRODUCTION

One of the ways to acquire familiarity with a problem is by "real-life" experience, but for a variety of reasons this is not always possible nor is it always desirable. For many centuries men have been devising games as substitutes for more direct experience with problems of military strategy and tactics. In more recent times, extensive efforts have been made to apply war gaming techniques to other areas. The American Management Association developed the first business game (4) in 1956, and during the past six years many similar business games have been invented (1). Attempts to apply gaming techniques to broader political questions pre-date business games. Both the Germans and Japanese are known to have conducted political games prior to World War II (2,5), and there is reason to believe that these games had an impact on their subsequent actions. In recent years political games have been reported by a number of authors, particularly Goldhamer and Speier (2) and Guetzkow (3). The present paper takes as its point of departure the notion that many arms control problems can be studied using simulation techniques similar to those employed in war games, business games, and political games. The central question which we seek to answer is: "How might gaming techniques be reworked and applied to arms control problems?" The answer to this question is couched in the form of a game which we call Arms Control Simulation (ACSIM).

The development of a suitable simulation technique for studying arms control could have important implications for improving our understanding of the problem. Such a simulation vehicle could (1) permit the empirical analysis of suggested disarmament measures; (2) be used to generate hypotheses about the inspection provisions of treaties; (3) provide training for personnel concerned with arms control; and (4) demonstrate to interested parties some of the mechanics of arms control implementation and planning, and possible consequences of given courses of action.

STRUCTURE OF THE SIMULATION VEHICLE AND GAME

The structure of the simulation vehicle and game assumes that the political and economic control of the environment has been divided between two groups, each of which possesses a military force and a supporting economy. The assumptions of only two opponents simplifies the description of the game, but it should become apparent as the technique is outlined that the game can be expanded to include multi-nation alliances and "nth" country problems.

The "real world" context assumes the existence of an arms control agreement ("Disarmament Proposal" in Figure 1) to be studied in the laboratory. In real life, each of the two nations has available to it certain tangible resources, such as air bases, missile launch pads, industrial complexes, aircraft carriers, infantry divisions, etc. These resources can be managed by the nation's decision makers in compliance with or in violation of the disarmament treaty. In managing their activities, a nation's leaders have only limited information about their opponents which is obtained by means of various restricted channels of intelligence—some formal, some informal.

In the laboratory setting, decision-making functions are performed by two teams of experts (Red and Blue). It is important to emphasize that the simulation vehicle involves the use of both experts and a computer. Experts interact continuously with the computer to make crucial high-level decisions which would be made in real life by

*Reprinted by permission from the SP series (SP-779) of the System Development Corporation, Santa Monica, California. The concepts described in this article have evolved out of innumerable discussions with interested groups and individuals. The authors would like to express their appreciation to all of those who have exchanged ideas with us, especially Harold Aaron, Lloyd Belt, William Cannon, Ronald Gilliam, and Thomas Rowan.
Fig. 1. The "real world" context on which a disarmament agreement is to be imposed, showing information and control channels.

umans. The computer simulates the environment with which the humans interact.

The resources available to each team are defined by a comprehensive collection of aggregated descriptive facts, a DATA BASE. Each team manipulates its data base by means of computerized models. The descriptive elements of the data base serve as stimuli for a variety of retrieval programs which generate the intelligence data for the opposing team.

A game runs for 25 to 30 cycles, each representing one or more months of real time. Each cycle consists of four distinct activities: (1) both sides request intelligence about the other; (2) both sides request changes in their own data base; (3) both sides receive intelligence about the other; and (4) both sides receive feedback on the updating of their own data base.

A Control Staff acts as a monitor and performs "referee" functions, resolving conflicts in the decision rules and making assessments by means of computerized evaluation aids.

The resources of each team are divided into two general areas: a military establishment and a supporting economy. These resources can be manipulated to provide varying amounts of support to military and civilian establishments, as suggested by Figure 2.

The military establishment includes not only the operational forces, but also research, development, and certain production activities. Each of these activities requires a cyclic commitment of resources which ultimately results in changes in potential military effectiveness. By means of assessment models, values are derived which describe the over-all capabilities of each team in terms of its strategic offensive strength and its vulnerability to strike, under conditions of both conventional and limited war and for both non-nuclear and nuclear weapons employment.

A portion of the resources of the supporting economies is available during each cycle of the game. These resources can be committed against any of the military activities described above or may be returned to the "economy" in the form of direct industrial expansion and long-term research projects. These resources can also be committed to such areas as foreign aid, which in turn affects the long-term growth of the economy and the overall military assessments. The total measure of the economy is formulated in an "Aggregated Production Index" which is a measure of the consolidated productive investment made by the team. This index determines the amount of resources available for commitment during each game cycle. The raising of this index serves as an alternative
A GAME FOR STUDYING THE PROBLEMS OF ARMS CONTROL

Fig. 2. Resources available to the teams for their civilian and military establishments and possible channels of resource allocation.

Fig. 2

The laboratory vehicle used to simulate the decision-making environment is composed of four elements for each team (Figure 3):

1. A DATA BASE, the comprehensive collection of facts defining "reality" for each instant of exercise time;
2. MANAGEMENT MODELS, for the manipulation of the data base and for the costing and forecasting of changes;
3. INTELLIGENCE MODELS, which retrieve appropriate descriptive fragments from the opponent's data base, and generate the various types of intelligence reports; and
4. ASSESSMENT MODELS, to aid the Control Staff in taking stability measures and other analytical data.

The data base for each team is composed of two general types of data: Descriptive Data Fragments, which serve as sensor stimuli for the generation of intelligence reports, and Aggregated Data. The data base (Figure 4) can be conceptualized as a matrix organized along three axes: the Environmental Resources axis, the Sensor axis, and Game Time (T). Descriptive data fragments, within the data base, serve two principal purposes: (a) they are the basic units manipulated by the team decision makers in posturing their resources pursuant to a given strategy; and (b) they serve as stimuli for the generation of intelligence reports to the opposing team. These descriptive data fragments include unit records for the components of the military forces, various strategic production activities, critical stock-piles, and key research programs. These fragments also serve as potential discrete sensor stimuli. A series of unique designations are included in these descriptive fragments which describe (1) the total military capabilities, (2) the requirements in resources for maintenance and modification, and (3) various sensor susceptibility factors which are employed by retrieval models for the generation of intelligence reports. For areas of particular importance to the game, data fragments are detailed. For example, they may deal with production networks, assembly facilities, testing sites, depots,
Fig. 3. The analogue of the "real world" context of Figure 1, illustrating the structure of the game.

Fig. 4. The three axes of the data base illustrating the storage of descriptive fragments over time and by type of sensor.
and the like. In the case of less critical areas, these fragments are more fully aggregated.

Civilian production and other non-military areas which need not serve as discrete sensor stimuli are aggregated on a grosser scale than the descriptive fragments of military forces previously described. These records, however, still include fixed and variable costs, manpower breakdowns, critical materials, and other resource units. The records also include other relevant parameters affecting the manipulation of the fragmented portions of the data base and the resulting effects on the economy.

Each team is provided with management models with which to manipulate its data base. These models function as automated decision rules which enable teams to make changes in their force posture, to allocate their resources against various strategies, and thus to pursue particular goals. These models also assist in forecasting commitments, costing changes, and reporting progress.

The particular disarmament treaty under consideration at any given time outlines the specific "sensors" which are available to the two teams for "sampling" the data base of the other. The term "sensor" here refers to data collection means which play a perceptive role for the two teams, such as aerial inspection, seismographic stations, or roving inspection teams. Sensors are the intelligence data source for the teams. One of a team's goals is to exploit its sensors so as to forecast mobilization trends and detect activities which breach the agreed disarmament treaty.

Intelligence Reports to each team are developed by computerized models which retrieve the raw data fragments, subject the raw data to appropriate degradation and filtering, and then synthesize and edit the finished reports. These models also serve a corollary function by establishing and maintaining the team's target files—a form of "derived data base" on the opposing team—and thus aid in the summary processing of the accumulated intelligence as the exercise progresses.

It should be noted that although the structure of the simulation vehicle is similar for both teams, the actual data bases and their associated models are asymmetric: each must be tailored independently to simulate its respective country.

In attempting to simulate an environment, it is necessary to define limiting boundary conditions which will reconcile the experimental objectives with the resources available. The major rationale for the simulation vehicle is to provide a reasonable base for decision making about increases or decreases in armaments. Accordingly, it is necessary to have a relatively complete data base with respect to military information. Less sophistication with greater aggregation is tolerable for economic data, since they only serve to support the military data used in decision making.

The Assessment Models (Figure 3) are employed to assist the experimenters to tabulate the various goal measures of both teams: the Aggregated Production Index and the various lumped parameters describing the collective military capabilities. These measures are used to follow the progress of both teams and to evaluate the degrees of instability suggested by their relative strengths. These models also assess the yields available to the aggregated economies on succeeding game cycles.

The game is administered in monthly cycles, each month being denoted by the advancing of the "data base clock" one discrete step. Groups of three cycles make up a "quarter" which includes a period of review, evaluation, and planning. Each team, in accordance with its long-term strategy, outlines its manipulations of force posture and allocates its resources by quarters, and then implements detailed changes by monthly change requests. Intelligence requests are also submitted on a monthly basis. Each monthly cycle includes, therefore, the following basic tasks:

1. Review of recent reports and subsequent specification of data base changes;
2. Specification of Intelligence Requests;
3. Review of Progress Reports and other data base summaries;
4. Review and evaluation of latest Intelligence Reports.

CRITICAL EXPERIMENTAL VARIABLES

The simulation vehicle discussed in the previous section is a highly versatile tool which permits a unique approach to a wide range of questions about the effects of disarmament proposals. There are associated with each of these questions a large number of potential experimental variables, such as the type, amount, and quality of information available to the two teams or the disarmament strategies adopted by them.

Although it would be virtually impossible to manipulate or control all critical disarmament variables simultaneously, it is important to identify and control selected variables. For experimental purposes a dynamic case study method is used; this method involves presenting to expert participants a unique configuration of a selected set of disarmament variables during any given run of the game. Examples of critical disarmament variables which might be studied in a game of this type are:

1. Intelligence
2. Phasing of disarmament
3. Disarmament levels
4. Strategies of participants
5. Others (e.g., blueprints versus no blueprints).
Intelligence refers to information necessary for the process of verification. Intelligence may vary as to type; for example, it may be covert information of a specific or general nature, or it may be overt random intelligence. The quality of intelligence, its frequency, or even the sensor by which it is obtained may vary within each intelligence type. Intelligence can also vary as to the amount available and the ratio of the amount of information available to the two contending sides.

Disarmament levels refer to the degree of arms reduction programmed to occur. Obviously this varies as to abruptness or gradualness; abrupt disarmament could occur either early or late in any stage of disarmament.

Data collected are used to perform two kinds of analyses: (1) Chronologic, and (2) Debriefing. Chronologic analyses involve the development of historical, narrative descriptions of each Dynamic Case Study. These descriptions not only involve reporting the sequential unfolding of events, but also include efforts to uncover possible cause and effect relationships, to discover important interactions, and to isolate particularly sensitive variables. In addition to the Chronologic analysis, a Debriefing follows each Dynamic Case Study. During these debriefing sessions, expert participants are presented with performance data collected in the course of the case study, including data about the behavior of their opponents and changes occurring in both data bases on a cycle-by-cycle basis. Experts analyze these data and their relationship to the variables introduced in the study for hypotheses about cause and effect relationships, possible interactions, previously unanticipated problem areas, etc.

Unfortunately, the real world is too complex to be duplicated exactly. Because of this complexity, an element of artificiality characterizes all simulation studies. For example, the data bases are inevitably incomplete. While military resources may be represented in the data bases rather completely, they will nonetheless be aggregated to some degree (perhaps in the case of the Army to the division level) whereas economic chains will be incomplete and highly aggregated. Just as the data bases are incomplete, assumptions about the interactions of people, weapons, production facilities, etc., also must be incomplete.

The existence of data bases of the type described, even though incomplete, furnishes a new and valuable tool for studies of the type described and represents a significant advance over simple desk-top analysis where the data bases are only implicit and necessarily much more incomplete. While assumptions as to how the world functions may also be incomplete, they must be stated clearly for research purposes and be out in the open for examination and criticism. The research environment is dynamic. Conjecture about "what would happen if . . . " can be replaced by analysis of "what happened when . . . ."

DATA COLLECTION AND ANALYSE

Two primary classes of performance data are collected during the run of the game: (1) data about the behavior of expert participants; and (2) records of changes in the data bases of the two sides as a consequence of the play of the game. Comparisons are then drawn between case studies as well as within case studies on a cycle-by-cycle basis.

REFERENCES


INTRODUCTION

As social conflict in the modern world becomes increasingly menacing, our incentive to control it likewise increases. But conflict control requires not only the will but the knowledge, and our knowledge remains painfully inadequate. In no sector is the need greater and the knowledge base more flimsy than in international politics. As I see it, this scientific inadequacy stems from several general conditions.

First, the amount of scientific research, in which we test our ideas against the benchmarks of evidence, has been close to negligible. This is partly a function of the reluctance to treat international politics in general, and international conflict in particular, as legitimate objects of scientific inquiry. First of all, there is the still widespread belief that all social phenomena are so unique and discontinuous that the search for general laws is a waste of time. In addition, the conventional definitions of loyalty usually require a predisposition to believe in the righteousness of your own nation's cause, making rigorous analysis of foreign policy problems quite unnecessary; and scholars have been no more immune to this definition than have their fellow citizens.

Another factor, one which is by no means independent of the others, may well be the nature of the models we use. While a good model is no guarantee of success, a poor one virtually assures that we will, at the least, ask the wrong questions. While the others also deserve further examination, it is to the problem of a model of international conflict that this paper is addressed. Before suggesting a particular one, however, a few preliminary comments are in order.

SOME CRITERIA OF A USEFUL MODEL

What do we mean by a "good" model in the social sciences? There are several epistemological requirements, as well as some substantive ones, to which we will turn in a moment. First, it should reflect whatever we think we know at the time and should be neither ignorant of, nor incompatible with, such knowledge. Second, it should be built around propositions and conceptual relationships that can be put to the test. Third, it should be adaptable to relevant new knowledge and not need a drastic overhaul with each discovery, as long as that discovery is logically and empirically compatible with the model. A fourth and related requirement is that, in accord with the general systems outlook, the model should be able to exploit relevant knowledge from other disciplines and other levels of social organization. Fifth, it should provide an accurate—if simplified—representation of the referent world. Finally, it should be built around as few variables as are necessary to eventually account for the phenomenon to which it is addressed.

These criteria (as well as the definition of model which they imply) are, of course, not beyond dispute, and the social sciences today are alive with epistemological and methodological controversy. But given the nature and generality of our concern in this paper, they seem to offer one set of appropriate benchmarks.

Shifting now from the affirmative to the negative, what are the inadequacies of the models generally in use for the analysis of international conflict? In addition to their failure to often satisfy the above methodological requirements, most of the typologies and models now found in the literature suffer from a number of more substantive inadequacies. Among these, two would seem to be critical. One major flaw is that most extant models are not explicitly longitudinal, and I mean this in two senses. One, there is little systematic effort to compare a large number of international conflicts over time, in order to ascertain recurrent regularities or secular trends. Two, there is little explicit concern with the extent to which the configuration of any given conflict may be both a consequence of those which preceded it, and a predictor of those which follow it. That is, international conflicts are seldom treated as either comparable (in the operational, systematic sense of the word) or sequentially interdependent.

The second, and perhaps fatal, flaw lies in the general tendency to focus on only one level of analysis, rather than treat the interactions that occur across the several relevant levels. That is, the model often puts most of the explanatory burden on a) human nature; b) certain "vested interests" or economic classes, as in the Marxian mode; c) specific types of "aggressive" nations; or d) the structure and culture of the international system. Despite the rhetorical attractiveness of many such

*This is an expanded version of an earlier paper entitled "A Cybernetic Interpretation of International Conflict" in Rizzo and Gray, eds., Unity and Diversity, New York: Braziller, 1970.
I suspect that they will lead us nowhere.

As I see it, a model that might carry us to a fuller grasp of the dynamics of international conflict must not only satisfy the epistemological criteria outlined above, but must also be both longitudinal and multi-level in its focus. The present paper represents a modest effort to approximate such a model, to explore its contemporary policy implications, and to suggest some possible ways in which the ominous consequences of its dynamics might be mitigated, if not eliminated. Lest there be any misunderstanding, however, there is no suggestion that it satisfies all of the key requirements of a scientific model; it is, at the least, pre-operational.

INTERLOCKING POLITICAL SYSTEMS: NATIONAL AND INTERNATIONAL

In any examination of global politics, there are at least five levels of analysis available to the observer: First, there are the three which are most frequently utilized in the disciplines as mentioned above: the individual, the national state, and the international (or global) system. For the period from approximately the Treaty of Utrecht (1713) up to perhaps World War II, these three system levels turn out to be more or less sufficient, but for a longer historical view, into the future as well as into the past, two additional levels of analysis seem essential. Let me refer to them in general terms as the sub-national and the extra-national. With the individual, plus these four classes of social system, one can put together a fairly complete description—cross-sectional or longitudinal—of global politics for any epoch (Singer 1969). In this section and the next, I will focus primarily on the national and global levels and will re-introduce the sub-national and extra-national levels later in my argument.

One of the more important but less obvious characteristics of modern international politics is the fact that the same sets of individuals play the dominant role in both national and international politics. These are of course the national political elites—those individuals who comprise what is variously called the government, the regime, the administration, or, less frequently, the court. Within the national state there may well be other elites with a fair degree of autonomy who dominate provincial or local politics, but who are normally subordinate to those who comprise the national regime. On the other hand, however, there does not yet exist any legitimate authority above the hundred-odd national regimes. Given the extraordinary durability of the doctrine of national sovereignty, most influence in international politics is exerted in a horizontal direction—nation vis-à-vis nation—and almost none in the downward vertical direction. There are of course many international organizations and even some supra-national ones, but they remain largely the creatures of their nation members; hence we speak of the global system as "sub-system dominant" (Kaplan 1957).

One consequence of this state of affairs is that national elites constitute the major actors in both national and international politics. Moreover (and of central concern to us here) the demands of these two systems are often quite incompatible. Behavior which leads to success in one environment may often lead to disaster in the other, and vice versa. The balance of this paper will be addressed to (a) the nature of the conflicting incentives, temptations, and constraints which are generated by both sets of systems; (b) the resulting inadequacy of their homeostatic mechanisms in terms of their impact on both the escalation and control of the conflict process; (c) some possible short-run modifications of a self-correcting nature that might reduce the magnitude of these conflicts which are so inevitable a part of international politics.

What makes a certain level of such conflict almost inevitable? In the global system, given the absence of legitimate supra-national authority, national elites have relied on the ultimate threat of military power as a means of defending "national interests" against possible interference by other nations (Singer 1965). This traditional reliance on force as the final arbiter has, in turn, inhibited the growth of an alternative basis for inter-nation harmony: a widely accepted normative code which provides for peaceful settlement of the inevitable conflicts and clashes of interest. In the absence of both coercive authority and normative consensus, and in the presence of many material and psychic scarcities, the only remaining basis for cooperative behavior is a utilitarian one—a payoff matrix which rewards short-run restraint and accommodative strategies.

And there is the rub. If two nations become involved in a conflict, the general options are two. The most natural, and probably the most frequent, response is to stand firm on the original conflict-inducing position, or perhaps to even increase the original demands. Within most well-integrated national societies, this response tends to be applauded, and the limited opinion survey data suggest that it generally enhances the popularity of the regime. Moreover, this behavioral response tends to reinforce the existing norms of world politics ("this is the way things are done") and hence the probability that other nations will handle subsequent conflicts in the same general manner. 1

1. The analysis suggested here does not of course necessarily apply to every conceivable inter-nation conflict. While most such conflicts are, in my judgment, matters of routine incompatibilities between and among traditionally defined national interests, some do indeed raise legitimate issues of justice and morality. Unfortunately, we have not yet developed any generally accepted criteria for distinguishing between the two types of case, and even if we could, nationalistic appeals would often overwhelm them.
But this is a fairly standard and stylized opening round routine, and not particularly pregnant with danger. The critical question is whether the protagonists now succeed in "backing off" sufficiently so that routine diplomatic procedures can be brought into play, or whether one or both parties continue to press their claims in the original and more vigorous fashion.

The other general option is to recognize the opening moves for what they are and to then initiate and reciprocate moves of a more conciliatory nature. But the probabilities are all too high that the competence, courage, or patriotism of one or both sets of elites will then be challenged by a "hard-line" domestic opposition, be it a legitimate political party in a democratic system or a less institutionalized faction in a more autocratic system. Moreover, the efficacy of that challenge from the "outs" will generally be high, due largely to the prior actions of the "ins." That is, political elites cannot man an army and finance a military machine without some sort of psychological mobilization (Deutsch 1953). In persuading an appreciable sector of their society that preparedness is necessary, they inevitably create a climate which must be relatively responsive to jingoistic appeals from the opposition. As a matter of fact, had some minimum psychic and material preparedness not existed prior to the conflict, there might well have been no conflict; had the nation been militarily weak or psychologically unprepared, the competitor would probably have had its way without any diplomatic conflict.\(^3\)

Having suggested the general linkages between the national and the international systems, creating largely incompatible sets of demands on the national elites, let me now describe the feedback processes in greater detail. My purpose here is to indicate more precisely where the self-aggravating tendencies are greatest, and then to suggest some possible feedback mechanisms whose effects might tend more in the self-correcting, and less in the self-aggravating, direction.

**SOME SELF-AGGRAVATING LINKS IN THE FEEDBACK PROCESS**

Given the limitations of space here, the most feasible procedure is to bypass any thorough description of the structural, cultural, and physical setting within which foreign policy decisions are made and executed, and concentrate rather on those few variables which are critical to the scheme outlined here. As I suggested earlier, one of the reasons for our failure to understand and more fully control inter-nation conflict is the tendency to treat such conflicts as discrete and separable events. By viewing them rather as part of an oft-recurring feedback process, we might better appreciate that the way in which any single conflict is handled is both a consequence of prior such experiences and a predictor of the way subsequent ones will be handled. The position taken here is that intra-national and inter-national events all impinge on one another in a cyclical and ongoing process within which the self-aggravating propensities frequently exceed the self-correcting ones by an unacceptably large amount (Deutsch 1963, Milsum 1968). As I see it, there are four points at which the self-aggravating effects of positive feedback are particularly critical during the inter-nation conflict. Let us discuss them, one at a time, noting how the traditional Western notions of confrontation between a government and its citizenry are viewed here as highly symbiotic relationships.

Regime and Opposition\(^4\)

The first point is found at the apex of the foreign policy hierarchy within the nations themselves. The political elites, often unwittingly, "paint themselves into a corner" in order to accomplish two short-run objectives when engaged in diplomatic conflict. One objective is to demonstrate to the foreign adversary that they have both the intent and the capability to stand firm; the other is to head off any potential domestic attack based on the inadequacy of that intent and capability. In order to satisfy both these objectives, however, the elites will ordinarily resort to the kind of rhetoric which does little more than "raise the ante" all around. The intended message to the adversary may be merely one of firm determination, but since it will be heard at home as well, it cannot be too conciliatory; as a matter of fact, by making a commitment audible to the domestic audience, the decision makers may hope to make their foreign policy threats more credible, given the domestic costs, real or apparent, of capitulation.

Assuming for the moment that the early verbal behavior has demonstrated the appropriate degree of firmness abroad and at home, what are the likely consequences? The adversary's regime, of course, "will not be intimidated," and so responds in public messages to the several relevant audi-

---

2. Among the studies which deal with public response to foreign policy moves, at least in the U. S., are Deutsch and Murriti (1965), and Roseau (1965 and 1967).

3. An important (and hopeful) exception to this generalization would be those cases in which the psychic mobilization has been too hysterical, has gone on too long, or has otherwise rested on a less than credible base; in such cases, the domestic opposition, or some sector of it, might possibly push the "ins" toward a less bellicose position.

4. For the sake of simplicity here, I not only assume that there is a viable opposition in most nations but that the political spectrum is largely based on a two-faction, quasi-pluralistic division, with one or the other in power at a given time. These are, of course, drastic simplifications, but do not affect the argument at hand.
Military and Psychological Mobilization

In order to examine the second point at which positive feedback can get us into serious trouble, we can focus on another set of factors. Let us assume, reasonably enough, that both nations in the conflict are moderately well-armed by contemporary (but non-nuclear) standards, but that one enjoys a discernible superiority over the other in the relevant military categories, and that neither can turn to close allies for diplomatic or military support. The regime of the disadvantaged protagonist, having permitted the conflict to pick up some momentum, now has the choice of 1) bluffing, 2) retreating, or 3) delaying while improving its military position. The first can lead to a sharpening of the conflict and a more humiliating retreat later (or even a stumbling into war); the second makes it vulnerable to political attack at home. Thus there is always some temptation to try to close the manpower and weapons gap in order to bargain from a position of parity or even of greater strength. If this route is taken, the regime will first need to launch a program of psychological mobilization, without which neither the volunteers and conscripts nor the funds for weapon acquisitions might be forthcoming. In the process of mobilizing public and sub-elite support for these preparedness activities, however, two new conditions are generally created. First, the adversary is not likely to sit idly by, watching its superiority disappear; its regime therefore embarks on a similar set of programs. Second, both publics must become more persuaded of the need to resist the menace to their nation's security, and as a consequence, offer a more fertile ground for any militant domestic opposition. Given the almost irresistible temptation to exploit this state of affairs, the net effect is to raise hostility levels in both nations, and therefore to raise the expectations as to what would constitute a satisfactory settlement, negotiated or otherwise. Since these rising expectations tend to be fairly symmetrical, neither regime is in as good a position to compromise as it was during the first round of the conflict. The probability of further escalation, diplomatic rupture, or war itself is now appreciably greater.

Amplification Via the Media

Let me now turn to a third source of danger in the cyclical conflict processes that seem to characterize so much of international politics. To this point, the role of the media has had little attention, yet mass communications would seem to play a particularly central role in helping along the self-aggravating process. Again, the differences between a highly autocratic and a relatively democratic nation are seldom as profound as contemporary elites—communist, anti-communist, and other—profess to believe. At almost any point along the autocratic-democratic continuum, the political elites need the media, and the media need the political elites, be they regime or opposition. The regime relies on the media to help mobilize the population, to bargain with and ridicule the domestic opposition, and even to communicate with other nations.

While "managing the news" may be simpler to arrange when the party in power exercises formal control over its media, any effective and stable régime has little difficulty in doing so. First of all, the words and actions of the elite are, by definition, newsworthy, and therefore widely transmitted. Secondly, members of the regime have information available which can be of great help to the reporter or commentator to whom it is made available. Thus, by judicious release or righteous restraint, government officials can all too readily help or hinder the careers of many media employees. Thirdly, as regimes become more conscious of the need—and possibilities—of domestic propaganda, they begin to recruit media people into their very ranks as "public information" officers. Many newsmen are thereby involved in competition for these often attractive bureaucratic positions, and one way to stay in the running is to describe the appropriate agency's activities in a way that can turn to close allies for diplomatic or military support. The regime relies on the media to help mobilize the population, to bargain with and ridicule the domestic opposition, and even to communicate with other nations.

5. A similar interpretation is offered by Thorstein Veblen (1919, p. 3) in looking at the extent to which the Spanish-American and other wars were forced upon the respective governments: "The more that comes to light of the intimate history of that episode, the more evident does it become that the popular war sentiment to which the administration yielded had been somewhat sedulously 'mobilized' with a view to such yielding ... So also in the case of the Boer war ... And so again in the current European war ... here again it is a matter of notoriety that the popular sentiment had long been sedulously nursed and mobilized to that effect, so that the populace was assiduously kept in spiritual readiness for such an event,"

6. As I write these lines, the radio newscast tells us that "American aircraft carried out fifty-two sorties in the Hanoi-Haiphong area" but that the "Viet Cong terrorists continued their campaign of intimidation against unarmed civilians in an effort to disrupt the forthcoming democratic elections!" Propaganda in the name of patriotism is of course a virtue.
generally favorable fashion. While access to, and control over, the media may not be quite as simple for the "outs" as for the "ins," factions or parties in legal opposition are not without the sorts of media amplifiers they need to berate the regime for being "soft on . . . . ," devoid of courage, or incapable of defending the nation's honor. In some nations, each political party has its own newspaper, magazine, or radio station; in others, the possibility of the opposition coming to power can make the media somewhat more responsive than might be expected.

I am not, in this section, arraigning the media of most nations on charges of "selling out," although the charge would be far from groundless. Rather, despite the existence of a vigorous and independent sector in the media services of many nations, the general impression is that the incentives work to make these institutions a major factor in amplifying inter-nation conflicts and contributing to the positive feedback, escalation, process.7

The Redistribution of Domestic Power

The fourth and final factor to be considered in this analysis is the effect which a nation's participation in an escalating conflict can have upon the distribution of social, economic, and political power within the society. Without accepting those conspiratorial models that see generals and "military-makers" actively fomenting rivalry, conflict, and war, one must be extraordinarily naive to expect no systematic biases in the foreign policy preferences of those who comprise the military-industrial-labor-academic complex.8 Even more than with newsmen, questions of ambiguity will regularly tend to be resolved in the hard-line direction by many military officers, corporate executives, labor leaders, government bureaucrats, defense intellectuals, and technical consultants, as well as by the standard phalanx of patriotic organizations. Given the state of our knowledge about international politics, most foreign policy problems are indeed matters of opinion, rather than of knowledge or fact; and in matters of opinion, the point of view which gets the benefit of the doubt can be expected to win out most of the time.

The problem here of course is that in most nations the major positions of power—as well as the public plaudits—go to those who are in the ideological mainstream; this seems to hold even if the mainstream of the moment is allegedly pragmatic and non-ideological, as in the United States of today and (probably) the Soviet Union of tomorrow. Having acquired power, prestige, and credibility by advocating, or acquiescing in, the modal foreign policy positions, these middle elites are seldom likely to shift too far in their views. And they are particularly reluctant to shift toward a position which could be interpreted (or misinterpreted) as giving aid and comfort to the enemy, whoever the enemy of the moment may be.9

Furthermore, as the intensity of the inter-nation conflict increases, the higher becomes the value of the professional and extra-curricular services of these middle elites. On top of this, as their individual influence and status increases, the size of their sector also increases. When the armed forces expand, officer promotions accelerate, and when more weaponry is being designed and produced, more engineers and technicians are promoted and recruited; even academics in the social and physical sciences find that foreign policy conflicts lead to increased opportunities for money, status, and influence in the modern world. The high energy physicist or the professor of biology has his role to play in the preparedness program, just as the political scientist or anthropologist finds himself consulting on log-rolling tactics in international organizations, military strategy, or counter-insurgency. If for no other purpose than to give intellectual legitimacy to the conventional wisdom, academics are almost as likely to be co-opted into the foreign policy mainstream as are the more obvious members of the military-industrial complex.

My point here is that it does not take a so-called totalitarian regime to mobilize key sectors of the society. The basic properties of the sovereign national state in the industrial age are such that this mobilization occurs with little effort. No secret police, no dictatorial government, not even any veiled threats are required to generate the joint "conspiracies" of silent acquiescence and noisy affirmation, once a nation becomes embroiled in a conflict of any intensity, or a preparedness program of any magnitude. For the past century or so, the self-correcting mechanisms have gradually withered, despite the assumptions of economic liberalism and classical democratic theory. In the absence of vigorous countervailing forces within the nations or in the larger global community, the self-correcting mechanisms of international politics are feeble indeed, with the consequence that all too many of the inevitable conflicts among nations are free to grow into costly rivalries and, occasionally, into tragic wars. In the next two sections of this

7. For a cross-national range of interpretations of the media's role in foreign policy, see Cohen (1963), Kruglak (1963), Reuten (1966), Nimmo (1964), and Halo (1984).
8. The Vietnam War has led to a number of excellent U.S. studies which illustrate the naturalness of the process by which those who stand to gain from the escalation or continuation of the conflict drift into behavior which contributes to such escalation; see Barnet (1970), Lewin (1967), and Janeway (1968).
9. The fact that some members of the middle elite (including political office-holders) do eventually get off the bandwagon in no way contradicts the model. They seldom do so until the conflict has reached its apogee and shows little tendency to decline; recent examples might be Germany in 1944, France in 1957, and the U.S. in 1970.
paper, I will try to suggest certain limited procedures whereby these four basic self-aggravating mechanisms might be partially weakened or controlled.

SELF-INITIATED SELF-CORRECTING MECHANISMS

Is the interaction between and among nations in global politics as dismal as I have painted it here? Is the relative potency of our self-correcting mechanisms this much less than that of the self-aggravating ones? Considering the paucity of scientific, data-based research on global politics, and the absence of much evidence at either the micro- or macro-level, it is a bold man indeed who will take so dim a view and embrace so pessimistic a formulation. The picture may, admittedly, be overdrawn for the sake of emphasis, and it may even be that, as a science of global politics develops, this characterization of the nation-state system will turn out to have been seriously incomplete or inaccurate. Be that as it may, responsible scholars must act on the basis of the little that is known, even while working to enlarge that knowledge base, and the interpretation offered here will therefore have to suffice for the moment. The word "act" is used quite literally, since I intend in this section to shift from the descriptive mode to that of prescription. Having described how these aspects of the global system look to me, I suggest now a set of modifications that might conceivably reduce the probability of any given conflict erupting onto war, and of any given war converting great parts of humanity into a nuclear rubble. With so much at stake, it is embarrassing to propose so little, but the approach offered here may possibly generate some self-amplifying processes of its own.

Assuming that this formulation is essentially correct, and reiterating that a great deal more rigorous research is called for, I would single out the communication and norm-setting nodes in the national societies as one of the more high-priority points of intervention. Until decision makers become aware of the many ways in which their own behavior exacerabates conflicts, and converts the possibility of win-win outcomes into zero-sum ones, the chances are they will continue to act in the traditional manner, and often find themselves, unexpectedly, in situations from which extrication is costly or impossible.\(^\text{10}\) Journalists and commentators, for example, could pay more attention to the effects of such moves on the inter-nation conflict itself, and less to the effects on the regime's popularity vis-à-vis its domestic opponents. The various private or semi-independent groups that exist to influence foreign policy, or the public's attitudes toward it, could devote as much of their attention to the regime's conflict management techniques—and the opposition's acquiescence in, or exploitation of, such techniques—as they do to pursuing their own particular and narrow goals or applauding "our side" in world politics.

Perhaps more critical, but demanding much more in the way of short-run self-sacrifice, is the need for the political "outs" to play a less opportunistic game. Support for a "vigorous defense of the national interest" may win the opposition a word of thanks from the regime, and criticism of "a policy of appeasement" may win it some support from a large sector of the public, but neither of these tactics is likely to make the regime's diplomacy any more successful.

Nor will they have led to any improvement in the future. This is an utterly critical phenomenon in all social processes, yet it is very rarely acted upon, or even appreciated, and may deserve more than this passing allusion. Consider for a moment the relationship between two classes of phenomena in any social system: beliefs (embracing preferences, predictions, and perceptions) and behavior (including verbal, decisional, and physical). Every public action, especially if taken by a highly visible reference figure, exercises some impact, however minor and however indirect, on the beliefs and attitudes of those who observe or hear about the act. It may lead to the strengthening and reinforcement of some attitudes among some people, and to the weakening or modification of some attitudes among others. Given this dynamic interdependence, it behooves us to pay attention to the possible consequences of every foreign policy action that occurs. Each act of the opposition—no matter how weak its power or how cynical the public—helps shape those attitudes which will, in turn, shape the behavior of many of the participants in the foreign policy process. If, for example, the "ins" and the "outs" are seen to agree on the rights and wrongs of a foreign policy conflict, many citizens will conclude that there is no other reasonable position. And if they disagree to the extent that the regime is accused of appeasement, the regime will either modify its policy in a more militant direction, or try to appear as if the policy were indeed at least as militant as that advocated by the opposition "outs." Either way, citizen attitudes will be strengthened in a more nationalistic

\(^{10}\) In almost every government agency or corporation, there is a controller (or comptroller) whose major assignment is to watch budgetary income and outgo and to issue warnings when they tend to get out of balance, or look as if they might do so. Perhaps every foreign ministry ought to have an analogous officer whose sole responsibility is to watch for those trends which signal a potential loss of diplomatic maneuver.
and short-ranged direction. Unfortunately, many of those so influenced will be reference figures who themselves are "opinion influentials."

The importance of these positive feedback mechanisms is relevant not only during conflicts and crises, but before and after. If a conflict is finally resolved in a more or less satisfactory fashion, the contending regimes are likely to emphasize the diplomatic "victory" they have achieved by their firmness in the face of the adversary. Once again, this may enhance their prestige for a few weeks or months, but the main effect is to increase the popular expectation that all subsequent conflicts will end in victory-through-firmness. If it ends unsatisfactorily for one side, the norm is a refusal to acquiesce in the "unjust" outcome, and a pledge to redress the nation's grievances at the earliest opportunity. In either case, the prognosis for peaceful resolution of future conflicts is not favorable. Likewise, there is a great and naive myth in many more or less democratic societies that elections have one purpose and one purpose alone: to decide which party or faction shall be in power. Thus the campaign strategists first try to ascertain the dominant views of the various voting blocs and then proceed to pander to these views. With few exceptions, then, election campaigns — because they receive fairly wide and sustained publicity — tend to serve as a powerful reinforcement for existing views on many domestic and some occasional foreign policy issues. And, as I have already mentioned with some frequency, these are not views which make it easy for decision makers to pursue peace abroad and honor at home.

For the information channels to play a useful part in reducing the dominance of positive over negative feedback mechanisms in inter-national conflict, several groups will have to contribute. Scholars need to identify which points are most critical in different classes of conflict and which behavioral patterns account for most of the self-aggravating and self-correcting tendencies. Journalists and other media people need to take a more detached and critical view, accepting the important difference between their professional roles and those of politicians. Politicians need to appreciate the trade-off between short-run tactical gains vis-à-vis the domestic opponent and the middle-run liabilities that accrue when, in negotiating with foreign elites, they find little room for maneuver.

At first blush, these look as if they might indeed be steps which individuals and groups in each nation could take on a unilateral basis (Osgood 1962). If they could be taken unilaterally, one might feel somewhat more optimistic, but the fact is that too much progress along these lines in any single nation could put that nation at a modest disadvantage vis-à-vis other nations in the global system. After all, each of these steps implies, almost by definition, some reduction not only in the level of political and psychological mobilization within the affected nation, but also a longer-range trend toward public resistance to the standard mobilization appeals. When the attentive public in a nation becomes more sophisticated, far-sighted, and tolerant of compromise with foreign powers, its regime must enter diplomatic bargaining at a disadvantage. As a matter of fact, a favorite ploy in such bargaining is to inform the adversary that one's own public (or legislature, or press, etc.) just would not accept a particular settlement, and certain concessions must therefore (and regretfully) be requested in order to get an agreement which could be "sold" at home. If this is indeed an accurate portrayal, the only way to start the trend toward more realistic diplomacy is for the initiative to be taken by those nations which are clearly in the strongest bargaining position in any such negotiations. On the other hand, a generally accepted dictum in diplomacy is that the stronger power need not negotiate and the weaker one dare not, and this has all too often been the rule applied.

SOME NEGOTIABLE MECHANISMS

Let me shift now from some of the reforms which might be taken—albeit with some risks—within individual nations, to some which might conceivably be negotiated between and among nations, the better to mitigate the oft-disastrous consequences of positive feedback in international conflict. These suggestions will strike many as utopian and unrealistic, and if realism is defined as
conventional, they clearly are vulnerable to this charge. But, as many observers, from E. H. Carr (1945) onward have pointed out, 19th century diplomacy is pregnant with disaster in the environment of 20th century technology and ideology. If these interim measures of a palliative nature (or measures of a comparable sort) are not instituted, those who think and act in national state terms will find their system shattered even sooner than some predict. In my judgment, the national state is already beginning to falter in its ability to solve a variety of problems (domestic and international) and is beginning to lose the support and loyalty of its citizens. New forms of social organization are already in the wind, and though their precise form remains far from discernible at the moment, my guess is that one or another of these alternative forms will in time supersede the nations as major bases of human organization. The major question probably is whether these transformations will come about in a relatively non-violent fashion, or as a consequence of a major military catastrophe.

**Moderating the Media**

Be that as it may, let me outline in the barest detail some possible arrangements which might be negotiated in either an overt or tacit fashion between and among certain members of the international system. Focusing first on the role of information and/or propaganda, it was noted earlier that much of the self-aggravating tendency in international conflicts may be attributed to the attitudes and expectations that are generated and reinforced within important sectors of most societies. Suppose for the moment that we could readily devise a measure whereby, via content analysis, it could be ascertained how many bits of self-correcting and self-aggravating information are being directed at the public in a given nation (Singer 1963). Given the rather rapid advances in computerized content analysis (Stone, et al. 1966, and Holsti 1969) and the respectable fund of experimental data which social psychologists have accumulated, there is no reason to believe that we could not design a system which produces a weekly or bi-weekly index of the ratio between exacerbating and ameliorating information addressed to each population from within. Presented in matrix form and published widely and frequently, the domestic output within each nation regarding all other nations could show (a) which other nations are receiving most of the attention in each nation’s media; (b) how much of that attention is of a hostile, neutral, or friendly nature.

When the attention level reaches a certain threshold, the international agency in charge of the enterprise could begin to make and publish more frequent analyses, and if the ratio of self-aggravating to self-correcting output approached some previously agreed magnitude, the offending government and their media would be sent a prompt warning. If the ratio is exceeded, certain penalties of an appropriate nature might be imposed, although their precise characteristics need not be discussed here. More important than any sanctions, however, is the mere fact that national governments would themselves be altered regarding the extent to which they may be losing their freedom of action and painting themselves into that diplomatic-political corner alluded to earlier.

Now, some will argue that such a system requires some degree of government manipulation and management of the news, and I can only agree. This might be a serious liability if the news were not now managed and manipulated in most societies, but, persuaded as I am that the virtue of most mass circulation newspapers and other media was lost at a very early age, I cannot work up much concern over threats to the innocence of a “free press.” The relationship between media and regime is sufficiently symbiotic in most nations that this innovation need hardly cause a ripple.

**Inhibiting Mobilization Rates**

The first World War is often cited as the classic example of a disaster which might have been avoided had not the contending regimes been in such haste to mobilize their armies. But military mobilization was merely the last phase in a feedback cycle which had begun years before, and many historians marvel at the fact that the war’s onset was delayed as long as it was. Almost from the close of the Franco-Prussian War in 1871, the major powers began to build not only their alliances but their war machines. While new war plans were being drawn up, slow but steady mobilization of human and material resources got under way. Between 1907 and 1914, for example, the annual defense budgets for the Triple Entente rose from 152 to 239 million pounds sterling, and those for the Triple Alliance rose from 84 to 151 (Richardson 1960, p. 87).

All of this of course provided little more than the volatile setting for the six weeks embracing Ferdinand’s assassination (June 28), the Austrian ultimatum (July 23), the Serbian and Austrian mobilization of active and reserve forces (July 25), Russian mobilization (July 29 and 30), and the catastrophic events of July 31, which saw the German ultimatum to Russia and the French and German general mobilization. Had there been some regulatory mechanism at work during the arms race period and during the 1914 crisis period, the war might well have been avoided. That is, whether the pace of escalation is slow and steady

13. In an earlier paper (Singer 1969), this projection is discussed at some length, but the evidence by which it may be vindicated or rejected has not yet been assembled. For an interpretation of the opposing evidence, see Deutsch (1969).
(as in most arms races and other conflicts which require long lead-times) or rapid and erratic (as in most crises), there is a very high probability that the positive feedback tendencies will gradually increase in their potency, while the negative and self-correcting ones will diminish in their effectiveness. Left unchecked, such processes must culminate in war.

Suppose, however, that the leading nations were to negotiate an agreement according to which every increase in military capability were accompanied by an increase in certain socially desirable non-military capabilities. For example, every expenditure for weapon development or production might be matched by an equal expenditure for agriculture or housing, and every training program for military personnel might be matched by a comparable program for social science training, and so forth. While military preparedness in today's world is often enhanced by a wide range of activities, with farming improvements freeing more manpower for the army, and education in social or physical science tending to have militarily useful by-products, this is not the major problem. Of course, the stipulated activities and allocations should be of maximum constructive usefulness and minimum destructive usefulness, but even when they are not, the regulatory mechanism will have a self-correcting effect due to the bounds imposed by every society's resources. Human resources are strictly limited, and fiscal-material resources can only be increased at modest rates; thus, whatever is allocated to essentially peaceful activities is to some extent not available to military activities. Moreover, in order to make the inhibitions against rapid preparedness even stronger, the signatory regimes could agree on a steeply graduated scale of matching allocations; modest expenditures might be matched on a one-for-one basis, but above a certain threshold the compensatory allocation might be set at the square or the cube of the military allocation. In the same vein, per capita preparedness expenditures could provide the basis for contributions to the United Nations and its specialized agencies, also on a steeply progressive scale.

Considering the two kinds of mobilization—psychological and material—which need to be retarded and inhibited if international conflicts are to terminate in war less frequently, several regulatory mechanisms have been suggested. Others come to mind, but discussion of them will have to wait. It will surely be suggested by some readers that, if the sorts of agreements of unilateral initiatives outlines here could be undertaken, we probably would not need them in the first place. This argument overlooks a fundamental concept in social science, which, when put in the common vernacular, is known as "timing." If we recognize that the relationship between any two social entities tends to fluctuate widely between friendship and hostility over any length of time, the virtue of the approach suggested here becomes more evident. During periods of low tension, for example, nations can negotiate agreements whose contemplation at another time would be labeled as treason. If we can take advantage of such opportunities, national regimes can establish—before conflict develops—procedures which will give them or their successors considerably greater flexibility and more options during a conflict.

A closely related virtue is the fact that few arrangements of this type could be expected to function unless an extra-national or supra-national agency of appreciable authority were responsible for the bookkeeping. The information which such an agency puts out must be highly credible, and such credibility depends on accurate data inputs, the development of valid indicators, and rigorous data analysis. This proposal is, then, very similar to that which Quincy Wright (1957) made in the inaugural issue of the Journal of Conflict Resolution. Recognizing the need for and the feasibility of "a quantification of political and psychological conditions and trends" over a decade ago, he recommended the establishment of a privately-endowed, non-governmental World Intelligence Center which would publish periodic reports on the "climate of international relations." As important as the information itself might be, a useful by-product might be some modest increase in the skepticism with which attentive citizens greet much of the information disseminated by their own governments and national media.

**Negotiation by Proxy**

In a paper several years ago, I tried to outline another procedure whereby the conflict resolution responsibilities of foreign policy elites might be partially de-politicized (Singer 1965). Let me recapitulate here in a few paragraphs. One way in which to reduce the positive feedback effects might be to recapture some of the virtues of eighteenth- and nineteenth-century diplomacy and put this delicate activity back in the hands of relatively anonymous professionals. As a number of labor unions and industrial firms do, so might national governments hire law firms whose staff would actually conduct most of the negotiations.

---

14. The past decade has seen the establishment of several peace research centers around the world, usually associated with such universities as London, Oslo, and Michigan. The recently created International Peace Research Institute in Stockholm, while financed largely by the Swedish Government, may approximate the Wright scheme on the physical science side by publishing reports on disarmament inspection methods, etc. The Pugwash Committee is also, at this writing, preparing a memorandum on the creation of such an agency. Also relevant is the annual Report on the World Social Situation, published by the United Nations Department of Economics and Social Affairs.
The "mercenary negotiators" would, first of all, be thoroughly briefed on the background of the conflict and would also engage in independent research in order to have a balanced and accurate understanding of the issues. They would then be informed as to the range of outcomes that are envisaged, from most satisfactory to least satisfactory, with the quid pro quo associated with each such outcome. Having been so briefed, they would then negotiate with their sponsors a contingent fee schedule such that the most satisfactory outcome within a fixed period of time would produce a maximum fee, whereas no agreement at all would bring little more than compensation for their basic costs. In addition to providing some insulation from domestic politics, such an arrangement would require the contracting regimes to specify exactly what it was after and exactly what concessions it was willing to make in pursuit of those goals. This, of course, is not always done under present conditions.

The negotiators, could but need not be, nationals of the government they represent, and need not necessarily be lawyers. Quite possibly a United Nations diplomatic academy (or an agency such as UNITAR) could take on the responsibility of training and perhaps accrediting people from a variety of professional and national backgrounds. While this approach may have something in common with traditional practices of conciliation and mediation, it remains clearly a matter of domestic jurisdiction. It is not proposed here that negotiation be compulsory or that any third party intervene in any fashion. All that need be agreed is that, for a given conflict, both parties will resort to mercenaries, the better to remove the conflict from the temptations of domestic maneuvering and global posturing.

A mechanism such as this admittedly raises a number of new problems, ranging from the danger of collusion between the lawyers representing the governments involved, to resistance to a secrecy which violates the basic tenets of democratic practice in some nations. Other difficulties also come to mind, but in my judgment it is an innovation which deserves further investigation and perhaps limited experimentation between and among those nations whose conflicts are mild and whose populations have little distrust of one another. If it were to prove successful, the practice might spread, and even though it represents a move away from the direction in which global politics seem to be headed, it might help us to survive the dangerous transitions that lie ahead.15

To sum up this section, three possible mechanisms for the more successful resolution of international conflict have been outlined. While the assumption is that pre-conflict agreement on such hopefully self-correcting arrangements could be reached, it should be noted that any of the three could be, in principle, undertaken unilaterally, with the hope that others might, in turn, also adopt them.

CONCLUSION: HOMEOSTATIS AND SYSTEM TRANSFORMATION

In this paper I have tried to describe those relationships and behavioral propensities which account for a great deal of the positive feedback in the international system, and therefore help to convert many minor disputes into major conflicts. Now, some will urge that the mechanisms which I propose in order to strengthen some of the negative feedback tendencies are little more than short-run palliatives, and that the system is basically inadequate in its present design. The charge is probably a fair one, but it is also somewhat beside the point; this is the nature of the system as it now stands, and our immediate concern is to devise those homeostatic mechanisms which will keep the fluctuations in conflict within the safe range. Moreover, there may well be some "natural" tendencies toward more self-correcting behavior even in the absence of such innovations. For example, war has by and large not been a particularly effective conflict-resolving technique in this century, and the nature of the cost-benefit ratio is increasingly appreciated (Singer and Small 1970). Certainly, nuclear weapons and missile delivery vehicles can do nothing but increase that ratio—even if they do contribute somewhat to the tendency toward pre-emptive attack. Likewise, the drive for overseas possessions, once a major source of international conflict, has become less and less attractive. Colonialism, at least in its older form, just did not pay (Clark 1936). A third element of built-in stability is the trend I noted earlier: an increasing disenchantment with national states and nationalism, even in the "third world"; many of earth's inhabitants are beginning to look for alternative and more efficient forms of human organization. If this alleged trend continues, many regimes will find it increasingly difficult to mobilize support for traditional foreign policies.

Despite these favorable possibilities, the need for supplementary control mechanisms nevertheless remains. We must of course explicitly differentiate between mechanisms which are designed to perform a largely homeostatic function, such as those outlined here, and those designed to initiate a self-reinforcing feedback process which might lead to fairly radical system transformation. Even though we have alluded to some potential tendencies in this direction, we cannot go beyond them.

15. Law firms (sometimes foreign) occasionally handle governments' cases before international legal tribunals, and the lessons from those experiences might be applicable to these essentially non-legal conflicts.
ESCALATION AND CONTROL IN INTERNATIONAL CONFLICT

My major concern, then, has been to formulate a model of international politics which highlights, however crudely, those factors that are most relevant to the escalation and control of the conflict process, and to suggest some short-run adjustments that might keep such conflict within tolerable bounds. The next step—even as the present scheme is being refined and put to the test—is to focus on those elements which might actually initiate and/or accelerate the system transformation process. If such a more complete (and epistemologically satisfactory) model can be developed, it may not only help us to keep the inherent self-aggravating tendencies within safe limits, but also give us what may be a final opportunity to design a global system which is fit for human habitation.

REFERENCES


WRIGHT, QUINCY. "Project for a World Intelligence Center," J. Conflict Resolution, 1, 1 (March 1957), 93-97.
META-LAYER | six commodore, self-tracker.

If nominalism is univocal essence, mind can frame no single concept or image corresponding to a univocal or actual term.