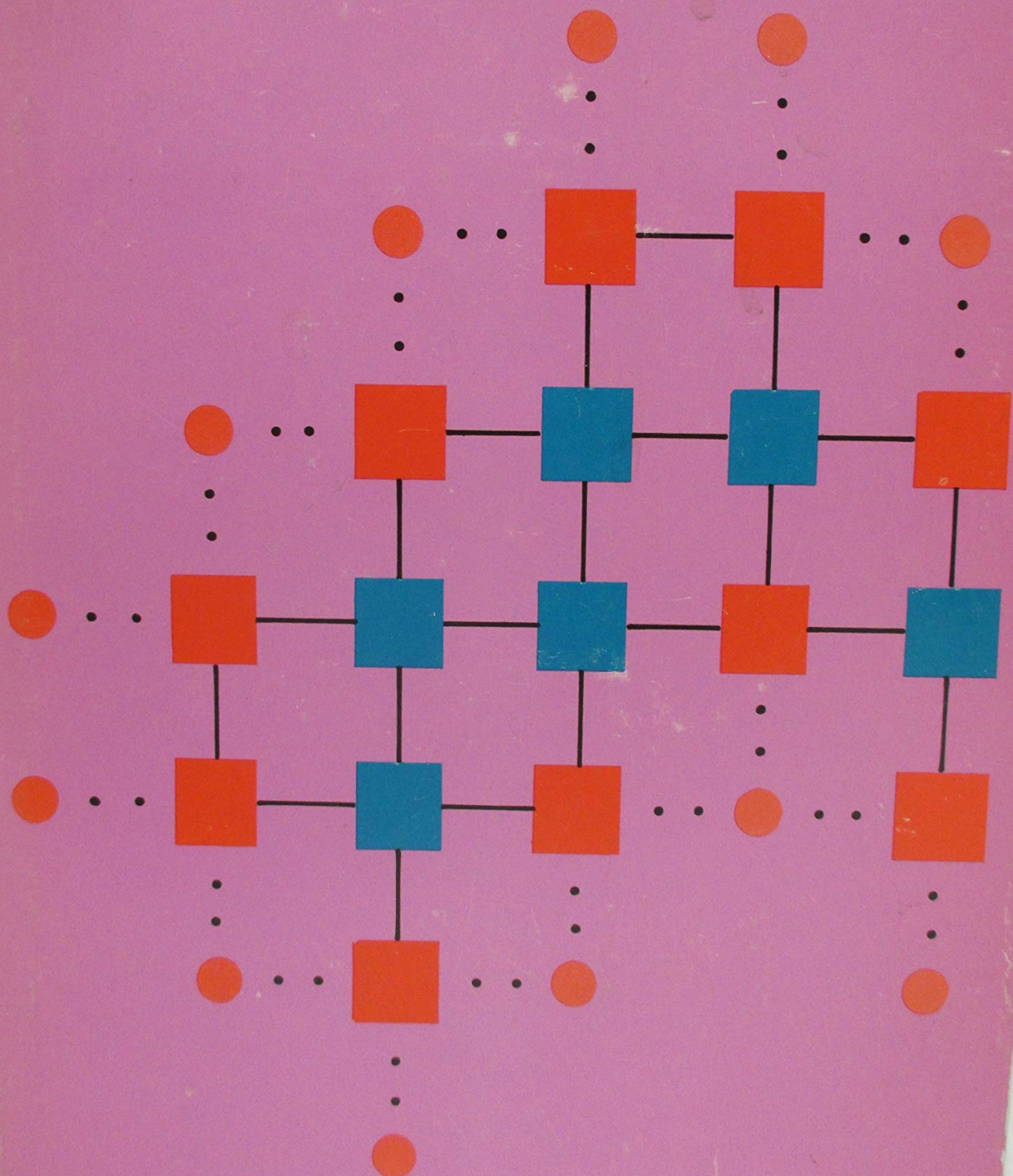# TRENDS IN GENERAL SYSTEMS THEORY

Edited by GEORGE J. KLIR

Trends in General Systems Theory

EDITED BY

GEORGE J. KLIR

School of Advanced Technology
State University of New York at Binghamton

Preface

After nearly two decades of development, general systems theory has sufficiently matured to be taken seriously. Once a new and radical movement in science, well motivated but ill defined, it has overcome many obstacles and has developed into a more moderate but by now a better defined and sophisticated area of human activities. Primarily, it offers scholars, educators, engineers, and artists new and harmonious ways of looking at the world.

So far, various branches of general systems theory have been evolving spontaneously and individually, without any significant coordination. It seems that a stage of evolution has been reached where some sort of integration might be beneficial. It is the major purpose of this book to provide a basis for such cooperation, by showing "the state of the art" and by predicting future developments.

To preserve authenticity, I have asked some of the most important originators of various aspects of general systems theory to describe their own motivations, views, approaches, contributions, and trends in this area. I have preserved differences in their terminology and notation. For easier reading, I have included a preview to the entire volume, as well as editor's comments and glossaries to some of the individual chapters.

The book contains four major parts:

- Part I includes some information about the history and basic aspects of general systems theory such as the system isomorphism, the problem of system complexity, and the role of computers.
- Part II discusses some important contemporary system problems. These are, essentially, problems brought from the social sciences.
- Part III describes the Mesarovic and Wymore theories, as well as my own approach, outlined by Robert A. Orchard in Chapter 7.
- Part IV contains studies of some aspects of formal systems theories. Included are the role of topological concepts, metamathematical aspects, problems of computability, and a discussion of some new or modified mathematical concepts important to general systems theory.

Although Parts I and II are relatively easy to read and do not require much mathematical knowledge, Part III is mathematically oriented, especially

Chapters 8 and 9. But the most difficult may prove to be some sections of Part IV.

The book was prepared with the intent to make it suitable as a textbook for graduate or upper-division courses or seminars in general systems theory. For that reason, individual chapters are supplemented by problems and comments in which I specify the necessary preliminaries.

It appears that the book contains enough material for a two-semester course. Chapters 1-6 and Sections 7.1-7.8, 8.1, 9.1, and 11.1, together with the Preview and the Epilogue, might form a good basis for the first semester. The remaining sections, which are more mathematically oriented, seem more suitable for the second semester. Various subsets of the book can be used for a seminar work.

Since the book is a survey of present trends in general systems theory, I believe that it will be helpful not only as a text but also as a basis for planning general systems theory curricula and, as already mentioned, as a medium for initiating cooperation in this exciting young field.

I thank all the authors who contributed to this volume. It was a great pleasure and a privilege to serve as their editor. My thanks also go to the School of Advanced Technology, State University of New York at Binghamton, which provided me with the best possible environment for my editorial work. In particular, I am deeply indebted to my colleagues, Professor Joseph V. Cornacchio and Mr. Sirajul Islam, for their very valuable comments, and to Miss Lucy Gabriel for the many hours she spent in helping me with the editorial work of the whole volume.

<div style="text-align: right">

GEORGE J. KLIR

</div>

Binghamton, New York
September 1971

Contents

Trends in General Systems Theory

Preview
The Polyphonic General Systems Theory

GEORGE J. KLIR

School of Advanced Technology, State University of New York at Binghamton

General systems theory in the broadest sense refers to a collection of general concepts, principles, tools, problems, methods, and techniques associated with systems. Although the name "system" may have different meanings under different circumstances and for different people, it ordinarily stands for an arrangement of certain components so interrelated as to form a whole. Diverse types of components and their interrelations represent different systems.

Although the notion of *system* is an old one, as pointed out by Ludwig von Bertalanffy in Chapter 1, the concept of a *general system* and the idea of general systems theory are relatively new. They were suggested by von Bertalanffy just before World War II, but were given publicity only after the *Society for the Advancement of General Systems Theory* (later called the *Society for General Systems Research*) was formed in 1954.

A need for a better understanding of biological, psychological, and social phenomena initiated an interest in the study of systems with strong (non-negligible) interactions between their components, as well as between each system and its environment. This new area of study contrasted with the "*classical*" (*Newtonian*) *method* in science, which regarded an object of scientific investigation as a collection of isolated parts and tried to derive the properties of the whole object directly from the properties of its parts without considering possible interactions between the parts.

A new scientific approach, often referred to as the *systems approach*, was thus suggested in the 1930's as superior to the classical approach in some disciplines of science, primarily in biology, psychology (psychiatry), and the social sciences. Since then, more and more evidence has been found that certain properties of systems do not depend on the specific nature of the individual system, that is, they are valid for systems of different natures as far as the traditional classification of science (physical, biological, social) is concerned.

Some of these properties were first understood as various kinds of simple *system similarities* (e.g., geometric, kinematic, thermodynamic). Two systems were considered similar if both the variables of one system were of the same physical nature as those of the other, and the values of these variables were proportional at corresponding times. Later, the meaning of similarity was generalized to include systems whose variables were of different physical natures. This kind of similarity, now usually referred to as the *analogy* between systems, is based on a similarity in the algebraic or differential equations describing the systems involved. For instance, certain electric circuits are considered analogous to certain mechanical, acoustical, thermal, and other types of systems provided they are described by similar equations.

Various principles of system similarity were finally incorporated in a formal theory referred to as the *theory of similarity* or *similitude* [18, 35]. It then became possible for a discipline to utilize methods developed elsewhere. For instance, sophisticated procedures developed for the analysis of complex electric circuits became directly applicable to mechanical, magnetic, acoustical, and thermal systems whose methodology was far less advanced. This ultimately led to the creation of a new discipline—the *theory of generalized circuits* [6, 38]. Proper understanding of the principles of analogy also stimulated the development, production, and use of *analog computers* [35].

The generalization of geometric similarity (recognized many centuries ago) to other types of system similarities, constituted the first step in the development of the concept of a general system. We can easily verify that the relation of similarity, in the sense described above, is *reflexive, symmetric,* and *transitive.* As such, it is an ordinary *equivalence relation* which partitions all systems of a particular discipline into *equivalence classes.* Each equivalence class can then be represented by a single system—a *representant of the class.* All results obtained by investigating this representant can be modified, using only rules of the theory of similarity, to apply to any system belonging to the same equivalence class.

The generalization of system similarity to system analogy was the second step in the development of a general system concept. Here, suddenly, several disciplines were involved. The concept of analogy introduced again an equivalence relation between systems and thus partitioned systems into equivalence classes. This time, however, the same equivalence class contained systems from different disciplines. Consequently, results obtained by investigating certain properties of one system (a representant of the equivalence class) could be transferred to other disciplines.

The similarity in the forms of algebraic or differential equations is a kind of *mathematical isomorphism.* When this is generalized to include any relation, whether expressible by equations or not, then the concept of a *general system* acquires its proper meaning. It is a contentless (mathematical) *representant*

(*model*) *of a particular equivalence class*, obtained when an isomorphic relation (which is always an equivalence relation) is applied to certain traits of systems.

Thus mathematical isomorphism, which is discussed in detail by Anatol Rapoport in Chapter 2, is crucial to any form of general systems theory. A study of various aspects of isomorphism, its modifications, and its generalized form—*homomorphism*, within individual conceptual frameworks, is very important for developing areas of general systems methodology with clearly specified applications. Let us note that the homomorphic relation is reflexive and transitive but not symmetric. This means that a similarity based on a homomorphic relation between systems classifies but not partitions them. Nevertheless, some problems concerning all systems in a class can be solved in terms of its representant—a *homomorphic model*.

Strictly speaking, *general systems theory* (in the widest usage of this term) is not a theory in the formal sense (an axiomatic theory), although it embodies some formal theories—the theory of finite-state (sequential) machines (automata) [1, 7, 14, 15, 20, 34, 45], the theory of probabilistic (stochastic) machines (automata) [2, 7, 33, 45], the mathematical theory of formal languages [1, 7, 16, 21], the theory of Turing machines (Chapter 11), the Mesarovic theory (Chapter 8), the Wymore theory (Chapter 9), and others. In addition to these formal theories, general systems theory is often considered to contain various concepts, hypotheses, methodological principles, computer techniques, and other particulars which do not belong to any formal theory.

At present, there is a general trend to formalize so as to diminish conceptual confusion. As a rule, however, the process of formalization narrows the original semantic meaning of the entities concerned. It is this "poverty of fully formalized concepts," in the words of my friend Eugene Kindler, which is a grave disadvantage of the formalization in its present form despite its many advantages. This is a problem to which Preston C. Hammer addresses himself in his criticism of the common interpretation of some very basic concepts in mathematics (Chapter 12).

One way of encompassing various facets of "semantically rich" concepts, associated with systems in their formalized forms based on available mathematical tools, consists in developing several formal theories of systems, each of which reflects certain aspects of reality. These theories, though mutually different, are not necessarily disjoint. Together, they offer far richer coverage of the semantic content of various system concepts than each of them could provide individually. This is, essentially, the approach that has been followed so far.

Another way consists in preserving as much of the semantic content as possible in the process of formalization. Clearly, such a way is heavily dependent on new developments in mathematics. Involved are both a

modification (extension, generalization) of existing mathematical concepts, as discussed by Hammer in Chapter 12, and a creation of new mathematical concepts, principles, and tools. A good example of such a new mathematical concept is the Zadeh idea of *fuzzy sets* [10, 17, 43, 44]. The need for a modification and an extension of existing mathematical concepts is recognized increasingly and represents an important trend in general systems theory at large.

Although a reform in mathematics would make this discipline a more adequate tool for the formalization of general systems theory, there would be no hope of significant improvement if we did not have powerful computers equipped with sophisticated programming facilities. The role of computers in general systems theory, which is well described by Gerald M. Weinberg in Chapter 4, can hardly be overestimated. For system theorists the computer is a tool as basic and essential as the microscope is for biologists. Either of these tools enhances enormously the ability of the human being in a particular area. Unfortunately, our present computers can be compared with the microscopes used by Robert Koch; computers comparable to the electron microscope are far in the future.

Although computers help tremendously in solving various problems concerning complex systems, there is a definite limit to the *practically manageable complexity of systems*. This limit moves toward increasing complexity of systems in proportion to improvements of existing computer facilities and methods. Nevertheless, as pointed out by Bremermann [9], there is a theoretical upper bound to manageable complexity.

By simple physical considerations based on quantum theory, Bremermann makes the following conjecture [9]: "No data processing system, whether artificial or living, can process more than 2×10^{47} bits per second per gram of its mass." Then he calculates the total number of bits processed by a hypothetical computer the size of the Earth within a time period equal to the estimated age of the Earth. Since the mass and the age of the Earth are estimated to be less than 6×10^{27} grams and 10^{10} years, respectively, and each year contains approximately $\pi \times 10^7$ seconds, this imaginary computer cannot process more than 10^{93} bits.

The *Bremermann limit* seems at first sight rather discouraging for general systems theorists even though it is quite conservative (less conservative assumptions would lead to a smaller number than 10^{93}). Indeed, many problems dealing with medium-size systems fall far beyond the Bremermann limit in their computational complexity. For instance, let us consider the problem of implementing, by a single type of universal element (module), a given function which maps a set of n logic (two-valued) input variables, with values of 0 and 1, to a set of n logic output variables. Assume that the module has m input logic variables and one output logic variable. Assume further that our objec-

tive is a design with the smallest possible number of copies of the module. This is a very practical problem in the area of computer design.

In the general case, none of the output variables is equal either to a constant (0 or 1) or to an input variable. Then we can proceed as follows. One copy of the model is identified with each of the output variables of the given mapping. The input variables of each of these copies are functions of the input variables of the designed system. These functions must satisfy a decomposition of the respective output function of the designed system with regard to the function represented by the module. The decomposition, which is an ambiguous operation, can be expressed by a Boolean equation with m dependent variables (inputs of the module) and n independent variables (inputs of the designed system). All functions acceptable to the inputs of the module can be determined by solving the Boolean equation. We want to select a function which enables us to implement the given mapping with the smallest possible number of modules. It can easily be found [25, 29] that the maximum number of solutions is

$$(2^m)^{2^n} = 2^{m \cdot 2^n}.$$

Since we have to solve one Boolean equation for each one of n outputs, we can express the maximum number N of functions satisfying the decomposition by the formula

$$N = n \cdot 2^{m \cdot 2^n}.$$

It is well known that universal logic modules exist only for $m \geq 2$. Let us consider the most favorable case, $m = 2$, for which

$$N = n \cdot 2^{2^{n+1}}.$$

When evaluating this formula for several values of n, we obtain the following table:

n	1	2	3	4	5	6
N	16	512	2×10^5	1.7×10^{10}	9.2×10^{19}	2×10^{39}

n	7	8	9	10
N	8.1×10^{77}	10^{155}	1.6×10^{309}	3.2×10^{617}

We see that the maximum number N of solutions of the Boolean equations is beyond the Bremermann limit for $n \geq 8$. Now, assuming that the actual number of solutions is only a small portion, say one millionth, of the maximum number, we still get

$$8 \times 10^{154-6} = 8 \times 10^{148}$$

for $n = 8$, which is again beyond the Bremermann limit. Let us note that, if we used more than one type of module or a module with more than two inputs, the number of possibilities would become even larger.

Thus the problem of implementing a given mapping from eight logic variables to eight other logic variables by the smallest possible number of copies of a given set of modules is, regardless of the modules used, practically unsolvable, though it is certainly solvable theoretically (in terms of the theory of computability summarized by Lars Löfgren in Chapter 11). Still, this type of problem is of considerable importance to computer designers.

In order to solve problems like the one suggested, we must reduce our requirements. For instance, we do not require that the implementation contain the smallest possible number of modules, we select modules with certain convenient properties, and we make other concessions. Generally, we try to simplify the problem sufficiently so that it becomes practically solvable with the aid of computers.

Our example was taken from engineering. However, similar difficulties arise in science, and again we are forced to simplify in order to manage our conceptual systems. The simplification is more acceptable in some scientific disciplines than in others. For instance, as pointed out by Gerald M. Weinberg in Chapter 4, superimposition of pairwise interactions works fairly well in mechanics, but it is unthinkable in biology, psychology, or the social sciences.

The so-called systems approach has been developed in science for the purpose of considering all possible interactions of the elements of a system when the behavior of the system is to be derived. This contrasts with the "classical" approach, which investigated individual interactions isolated from each other and then simply superimposed one upon the other. As pointed out by W. Ross Ashby in Chapter 3 and elsewhere [5], the systems approach, though highly desirable, often leads to problems that are practically unsolvable. In such cases, simplification is necessary. This means excluding some of the interactions, which leads in turn to the study of the amount of interaction between the elements of a system. Ashby has been involved in this study for some time, and his views are expressed in Chapter 3.

Hence an important trend in general systems theory consists in developing methods that enable us to construct conceptual systems where interactions between the elements are sufficiently but not completely incorporated. It is hard to disagree with Ashby when he says, " The future of system theory seems to lie in the study of systems that are sufficiently connected to be real systems yet by no means totally connected." Weinberg goes even further in Chapter 4 when he talks about the *science of simplification* and relates it very strongly to general systems theory. Some aspects of this " science of simplification " appear to be treated in the recently evolved *constraint theory* [13].

Several approaches to the formalization of general systems theory have been

developing within the last decade or so. Each of them was initiated with a certain purpose in mind and has been built on the basis of a chosen conceptual framework. Three approaches are outlined in this volume: the axiomatic theory known as the *Mesarovic approach* (Chapter 8); the *Wymore approach*, which he himself calls "a wattled theory of systems" (Chapter 9); and my own approach, described by Robert A. Orchard (Chapter 7).

The Mesarovic approach can be characterized as a highly abstract axiomatic theory. It is built hierarchically from the highest point of abstraction, at which general systems are understood as arbitrary relations, each one defined on a collection of some abstract sets. More axioms are added when systems with more specific properties are studied.

Mesarovic uses two ways of specifying behaviors of general systems whose variables are classified as inputs and outputs:

(i) *Input-output* (*terminal, causal*) *specification*, in which the behavior is explicitly specified as a binary relation defined on a Cartesian product of two disjoint families of abstract sets.

(ii) *Goal-seeking* (*teleological, decision-making*) *specification*, in which the same binary relation as introduced in (i) is described implicitly in terms of a goal-seeking process.

Although more specific aspects of the Mesarovic approach are outlined in Chapter 8, two of its distinguished features should be stressed here:

1. A strong involvement in the elaboration of the theory of *general hierarchical systems*, where the goal-seeking description of behavior plays an important role [32, 40]. The importance of studying hierarchical systems can be properly appreciated after reading Chapter 5.

2. An applicability to *metamathematical problems* associated with the consistency and completeness of axiomatic theories [31].

The Wymore "*wattled theory of systems*," to use its originator's own words, "has been developed to subsume both the theory of discrete automata and continuous systems defined by differential equations." Wymore's definition of system is based essentially on a state-transition structure. As such, it is quite similar to various definitions of finite-state machines (Moore or Mealy machines, kth-order finite automata), but extends the latter to continuous functions requiring neither a finite number of states nor a finite number of stimuli. The theory is applicable, therefore, to hybrid systems, which contain both continuous and discrete variables, as well as to systems defined on infinite sets. For instance, various Turing machines (Chapter 11), whose tapes are potentially infinite in both directions, can easily be described in terms of the Wymore theory.

In addition to his definition of a system, Wymore formalizes the notion of a *coupling* of these systems. This extends the theory to collections of coupled

systems and makes the problems of analysis and synthesis meaningful. Finally, he uses the concept of system homomorphism (or isomorphism as a special case) to formalize the principle of *modeling* and *simulation*. He requires the model to exhibit the same input-output performance as the original.

Wymore has intentionally developed his conceptual framework independently of any specific form of mathematical representation. This has allowed him enough freedom to choose a suitable representation for a given purpose. For example, a real-number representation of a finite-state machine [22] is preferable to an integer representation if the machine is coupled with a continuous system or simulated on an analog computer.

My own approach to general systems theory, as introduced in my book [25], is described and further developed by Robert A. Orchard in Chapter 7. Whereas both the Mesarovic and Wymore approaches are of a *deductive* nature, mine might be called *inductive*. Rather than defining the concept of a system axiomatically, as Mesarovic and Wymore do, I identify *system traits* before I define a system as such. The identification is based on our intuitive feeling for systems and system-type problems in various disciplines (natural sciences, social sciences, engineering, mathematics, the arts). Those traits are compiled which are independent of a specific nature of variables involved (behavior, states, transitions, elements, couplings, resolution level, etc.). The compiled traits are then classified and formalized. Restricting ourselves to traits that satisfy certain natural requirements (primary traits), we arrive at five basic definitions of systems. Each of the basic definitions can be supplemented by additional traits, or several of them can be used together to define a system.

My approach thus leads to a spectrum of system definitions, each of which is associated with a particular class of system-type problems. *Primary traits* are those given in the problem; secondary traits are those which are to be found. All the sets of *secondary traits* which are correct solutions of a problem represent an equivalence class with regard to that problem. Similarly, all the system-type problems which use the same definition of system and require various secondary traits of the system to be determined create an equivalence class. Hence the definitions of systems classify system-type problems and so establish a broad basis for a general systems methodology.

According to the conceptual framework introduced in [25], the system changes if any of the primary traits participating in its definition changes. Orchard suggests and formalizes in Chapter 7 an extension in which a *time sequence of systems* is considered as one system under consideration. This extension, which Orchard calls the sixth basic system definition, allows the study of all kinds of *evolutionary processes* (self-organization, self-reproduction, etc.). Orchard's contribution not only is a significant addition to my

approach but also represents an important trend in the whole general systems theory.

Since the individual approaches to general systems theory are far from being well elaborated and no efforts have been made to compare them in all details, it is impossible to predict whether they will merge into a single theory (the union of all of them) or will remain separate because of essential differences. In any case, both the deductive and the inductive methods should participate in the development of the theory.

The Mesarovic theory is the oldest of the three approaches described and, as such, is also more highly developed than the other two. Mesarovic and Eckman initiated the theory at the beginning of the 1960's [12]. The principal educational and research center for this approach is in Cleveland, Ohio (Systems Research Center, Case Western Reserve University).

With no previous related publications, Wymore presented the entire conceptual framework of his theory in his book [41], published in 1967. The development of the theory has been motivated principally by the needs of systems engineering in the broadest sense (including the engineering of social systems). Unfortunately, there has not been enough time to elaborate the methodology associated with the approach. In particular, very little has been done in regard to system synthesis, which is of primary importance in engineering. Most of the work on the theory is done in Tucson, Arizona (Department of Systems Engineering, University of Arizona).

Some ideas proposed by a group using the pseudonym K. Vasspeg [39], in which I had participated for several years, some concepts introduced by Svoboda [36, 37], and my earlier work in cybernetics [24] contributed in developing the conceptual framework of my approach. As the conceptual framework is so new, there has not been enough time to develop a well-organized methodology. Except for some minor studies [27, 28], most of the effort has been spent in applying the conceptual framework to systems with two-valued variables (logic or switching circuits) [26, 29]. At present, the educational and research center for this approach is in Binghamton, New York (School of Advanced Technology, State University of New York at Binghamton).

Although the Mesarovic theory is more highly developed than the other two, none of them is satisfactorily developed from the methodological point of view. For instance, little has been done in general systems synthesis, and even less in the modification of individual conceptual frameworks to fuzzy sets [17, 33, 34]. Although methods solving some portions of various problems concerning general systems have been programmed for computers elsewhere, no serious effort has been made to integrate them. Neither Mesarovic nor Wymore has incorporated the investigation of probabilistic (stochastic) systems in his theory.

All these present deficiencies in individual approaches to general systems theory suggest the following trends:

1. A comparison of individual general systems theories and a unification wherever possible. Even if we manage in the future to unify the formal theories of general systems, it seems reasonable to preserve both the deductive and the inductive ways of presenting them.

2. An elaboration of a well-organized methodology of general systems, either based on a unified theory, or developed within individual conceptual frameworks if the unification fails. The methodology should cover probabilistic and fuzzy systems as well as deterministic ones. It should incorporate not only classical system-type problems in the natural sciences or classical engineering (electrical, mechanical, etc.), but also new problems which arise in the social and life sciences or in social engineering, as described by John H. Milsum in Chapter 5 and Walter Buckley in Chapter 6.

3. A development of a large-scale interactive and adaptive aggregate of computer hardware and software for work oriented to general systems in the sense described by Orchard in Chapter 7.

An investigation of the general properties of the various system theories is recognized as an important trend. This investigation, which includes different aspects, is directed essentially to the creation of a metatheory applicable to individual general systems theories. As such, it may have a great impact on the unification of existing theories of general systems.

In Chapter 11, Lars Löfgren addresses himself to various *metatheoretical aspects* of formal general systems theories. He demonstrates the importance of formal theories in general, and of formal system theories in particular. Pursuing this argument, he uses mathematical logic and the theory of computability. Questions concerning the explicatory and predictive power of a theory, its communicability, its syntactic information, and the problem of reducing one theory to another are among those which Löfgren expounds in considerable depth and exactitude in Chapter 11. He also investigates various problems associated with a formalization of some highly sophisticated types of systems, such as learning, evolutionary, or reproductive systems.

Löfgren is a strong advocate of formalization in general systems theory. Following his argument, one can hardly disagree with his thesis: "Everything that can be effectively explained can be formalized." Then his belief that "the problem of agreement on a suitable logical basis for a group of scientists is far less of a problem for them than to interact at all without formalizing their ideas" seems to be fully justified.

The work by Joseph V. Cornacchio, as summarized in Chapter 10, has also a metatheoretical flavor, though limited to a particular aspect—*topological*

structures of abstract models of general systems. He has been motivated by two considerations:

1. The need for topological structures in the abstract formulation of general systems models.
2. The exploration of relationships between the extended topological concepts introduced by Hammer [19] and mathematical models in general systems theory.

Cornacchio demonstrates that a topological structure is a fundamental trait of a broad class of specific systems, including those in such diverse disciplines as engineering, computer science, and the social, behavioral, or natural sciences. In the examples presented from these areas, the topological structure involved is that of the classical topological space, which is based on the fundamental concept of neighborhood. An example of the introduction of such a structure in the model of continuous functional systems is discussed in detail. Cornacchio then demonstrates, by rigorous methods, a fundamental relationship between the closure spaces introduced by Hammer and the set-theoretic structure characterizing the Wymore model of a general system. However, as he points out, more work is required before the usefulness of such generalized structures in formally representing the intuitive notions of approximation and continuity—notions which originally motivated the need for a topological structure—can be assessed. Another question is raised concerning the role of generalized topological structure in arbitrary general systems models. Apparently, the strongest case for the introduction of such generalized structures is that of the finite systems in mathematics. Hammer [19] (and other references in Chapter 12) gives a detailed argument for the latter case.

Cornacchio's work has evolved from graduate seminars on the mathematical aspects of general systems theory, held at the School of Advanced Technology, State University of New York at Binghamton. It is included, with other topics, in a set of class notes entitled *General Systems Theory: Mathematics, Models, and Methods*.

The developments of various general systems theories, as well as the metatheory, seem likely to converge to a *general systems profession*. The latter will be involved in developing sophisticated methods, supported by powerful computing facilities, to solve system-type problems irrespective of the disciplines involved. The general systems profession will provide services to other professions. As such, it will have to be adaptive to the needs arising from individual areas of human activities.

It seems reasonable to expect that the general systems profession will develop into several specific areas, such as systems engineering, systems science, systems art, systems philosophy, systems methodology, and systems

education. Both research and education will contribute to its development. It is hard to assert which of them will contribute more. As far as research is concerned, I have already mentioned some important trends. Without any doubt, a proper treatment of system complexity as a parameter is of primary importance. We will probably search for new and unorthodox forms for representing systems. For instance, the power of natural languages to express complex relationships in simple forms, which are fully satisfactory in many cases, seems to be promising. Unfortunately, little has been done in the data processing of natural languages.

As far as the impact of education on the development of the general systems profession is concerned, two aspects should be mentioned:

1. A need for training a sufficient number of system specialists to extend and accelerate basic research in the methodology of general systems. The isolated courses in general systems theory which have been offered so far are no longer sufficient. They should be extended into organized curricula, based on a conceptional framework (hopefully unified), and including courses on sophisticated mathematical tools, computer programming, principles of modeling, simulation techniques, operations research, principles of measurements, automata theory, theory of languages, and other pertinent aspects.

2. A need for familiarizing specialists in various disciplines with fundamental concepts and simple principles of general systems, to make them able to communicate with systems specialists and with people from other disciplines. Scientific and engineering disciplines, as well as humanities and the arts, should be included. It is an excellent experience to have in the same class students educated in various disciplines. They are first introduced to some concepts and principles of general systems. Each of them then presents a demonstration of the interpretations of the general systems concepts and principles to his special discipline. It is not expected that students taking such a course, or a sequence of such courses, will become system specialists. On the other hand, they should be able to distinguish problems which can be solved within general systems methodology from those which must be solved within a particular discipline. They should also be able to formulate properly their problems for system specialists, and to interpret correctly the results obtained from the latter. In addition, through the seminar talks, they learn about the peculiarities of other disciplines. Generally, after such a course (or sequence of courses) students are better prepared for interdisciplinary teamwork, even though each remains essentially specialized in his original discipline.

People involved in general systems theory are usually referred to as *generalists*, while those working in a classical discipline are called *specialists*. However, if one works solely in general systems theory, he then becomes a general systems specialist. He specializes in generalizations. Let us call him a *system theorist* or a *specialized generalist*.

For the last several years, an increasing amount of cooperation has been demanded of general systems theory from such areas as biology, psychology, health care, economics, management, and political science. On account of this, a danger arises: the system theorist may try to solve any problem presented to him by the specialist. It is very possible that the system theorist may find a solution that will not lend itself (prove helpful) to the problem proper. In such a case the theorist has not only hindered the specialist in his search for a solution but has, in addition, done a disservice to the entire general systems profession.

My assertion is that the fixed idea of powerful systems theorists, who can solve almost all problems for almost all disciplines, should be recognized as a myth and treated accordingly. A system theorist specializes in a study of the general principles of systems, and a few hours, days, or even weeks of concentrating on another discipline can give him only a very naive understanding of its peculiarities, needs, and problems. He cannot spend several years of study on each discipline in regard to which his advice is sought. If he claims that he is able to solve problems in various disciplines because of his knowledge of general principles, then he is either naive or dishonest.

A system theorist cannot master the various disciplines in which he will work sufficiently to enable him to solve all the specialized problems that may arise. But a specialist in, say, health care can easily grasp the foundations of general systems theory in a relatively short time. He would then be called a *generalized specialist*.

It is my belief that the generalized specialist is the person who will increasingly be in demand. We can characterize him as follows. He is essentially specialized in a discipline; at the same time, he is familiar, to a reasonable depth, with basic concepts, principles, and methods of general systems. In addition, he is aware of the capabilities and limitations of contemporary computers, is able to use these machines, and has some proficiency in computer programming. Even though he is not expected to solve complex system-type problems, he is able to communicate them properly to system theorists.

A peculiar facet of general systems theory is its *terminology*. Although it aspires to be the language for interdisciplinary communication, system terminology is a loose collection of languages used by various individuals or groups. It is unfortunate that, while different names are frequently used for identical concepts, different concepts are often denoted by an identical name. Such ambiguity is a source of considerable confusion and misunderstanding. This mishmash of terms casts doubts on the general systems theory as a whole.

Clearly, an attempt to unify the terminology should be of the highest priority. The task is not easy. It would require that a list of basic concepts of general systems theory be prepared and that each concept in this list be

identified with any associate names used by various system theorists. Then, an attempt should be made to select one term for each concept. This selection ought to be based on general agreement among the people involved. Under no circumstances should one set of terms be considered as superior to other sets.

The comparison of individual conceptual frameworks used in individual approaches to general systems theory appears to be very difficult. A meta-theory must be used to decide whether one concept is identical to, is different from, or is a proper subset of a concept drawn from another theory. When two theories are built axiomatically, the comparison may be considered as a sophisticated formal exercise in the metatheory. However, when inductively built theories are considered, additional difficulties, primarily of a semantic nature, may arise. In such cases personal contact between persons representing different approaches is almost a necessity if any progress in unifying the terminology is to be made. A series of well-organized workshops, devoted to comparison of conceptual frameworks and unification of the terminology of general systems theory, may prove helpful.

This book does not escape the terminological "jungle" characterizing general systems theory. As a collection of contributions representing different views on general systems theory, it reflects the existing terminological differences. In the following paragraphs, I attempt to point out some of these differences. I use my own concepts and terms described in Chapter 7 and in [25] as comparison references. The comparison should be considered only as a first approximation intended to provide the reader with some guidance.

Although I distinguish the concept of *object* (a part of reality subject to investigation) from the concept of *system* (some precisely defined properties of the object), the term *system* is used by some (von Bertalanffy, Weinberg, Milsum) for both of these concepts. Von Bertalanffy uses the terms *real system* and *conceptual system* for my concepts of object and system, respectively. Weinberg frequently uses the term *model of a system* rather than *system* when he wants to distinguish an object from a system. In my terminology, I use the term *model* not as an approximation of an object but rather as a relation of similarity between two systems.

Ashby's *real machine* [3] has the same meaning as my concept of *object*. In addition, his concept of a *variable* (or a *variable quantity*) is the same as mine. He is in complete agreement with my view when he says that "every real 'machine' embodies no less than an infinite number of variables, all but a few of which must of necessity be ignored" [3]. He then defines a system by a set of variables selected "from those available on the real 'machine'." This is consistent with one of my basic definitions of the system (the definition by a *collection of variables and a space-time resolution level*).

Zadeh uses the term *object* for "a set of variables together with a set of relations between them" [42, 45], which, in my terminology, is quite close to

(but not identical with) the concept of *behavior*. However, he occasionally uses the term *physical object* in the same sense in which I use *object*.

In formal theories of general systems (Mesarovic, Wymore), the concept of object has no meaning. However, the term *object* is used in a different sense. For instance, Mesarovic employs it as a synonym for an abstract set participating in a relation or, in a special case, a set of values of a variable (a *resolution level* in my terminology).

My concept of *system activity* is used under different names in different approaches to general systems theory. For instance, Ashby, Mesarovic, Weinberg, and Zadeh use, respectively, the terms *a line of behavior*, *a general time system*, *a chronological graph*, and *a class of time functions*.

The concept of *behavior* in the sense in which I use it (a time-invariant relation between certain kinds of variables) was suggested by Svoboda for discrete systems [36, 37]. This concept, which plays an important role in my conceptual framework, is not directly employed in any of the other approaches. Although it is included in the Mesarovic concept of a general system (a relation defined on a collection of abstract sets), it has not been developed within the theory itself.

The concept to which I refer as the *state-transition structure* can be found, with certain modification, in almost every approach to general systems theory. One modification is the *state-determined system* introduced by Ashby [3, 4]. It is used also by Weinberg (Chapter 4) and Zadeh [40, 45], as well as in the theory of finite automata (or finite-state machines) [7, 14,15] and other system theories [23, 45]. Another modification is represented by the Wymore definition of a system.

The concept of *program* (an initial state and a set of time instances imposed on the state-transition structure) corresponds to the Mesarovic *abstract dynamical system*. Clearly, it is meaningful in the case of the state-determined system, the Wymore system, and other forms in which the state-transition structure is involved.

Let us now address ourselves to the question: *What is new in general systems theory*? With the variety of views on general systems theory, we should expect a variety of answers. General systems theory is considered as a *formal theory* (Mesarovic, Wymore), a *methodology* (Ashby, Klir), a *way of thinking* (Bertalanffy, Churchman [11]), a *way of looking at the world* (Weinberg), a *search for an optimal simplification* (Ashby, Weinberg), an *educational tool* (Boulding [8], Klir, Weinberg), a *metalanguage* (Löfgren), or, prospectively, a *profession* (Klir). Each of these views and possibly others not mentioned here contain points that are new. This fact makes the answer to our question rather complex. To summarize, we may say that general systems theory in the broadest sense has generated innovations in the following manner:

1. *A new way of looking at the world* has evolved in which individual phenomena are viewed as interrelated rather than isolated, and *complexity has become a subject of interest.*

2. *Certain concepts, principles, and methods have been shown not to depend on the specific nature of the phenomena involved.* These can be applied, without any modification, in quite diverse areas of science, engineering, humanities, and the arts, thus introducing links between classical disciplines and allowing the concepts, ideas, principles, models, and methods developed in different disciplines to be shared.

3. *New possibilities* (principles, paradigms, methods) *for special disciplines have been discovered by making investigations on the general level.*

I have tried to outline the basic trends in general systems theory as I see them at this time. The reader is now invited to turn to the individual authors to enjoy the many facets of this theory and its trends.

REFERENCES

1. Arbib, M. A. (Editor), *Algebraic Theory of Machines, Languages, and Semigroups.* New York: Academic Press, 1968.

2. Arbib, M. A., *Theories of Abstract Automata.* Englewood Cliffs, N.J.: Prentice-Hall, 1969.

3. Ashby, W. R., *Design for a Brain.* New York: Wiley, 1952.

4. Ashby, W. R., *An Introduction to Cybernetics.* New York: Wiley, 1956.

5. Ashby, W. R., "Constraint Analysis of Many-Dimensional Relations." *General Systems Yearbook*, Vol. IX, pp. 99–105, 1964.

6. Blackwell, W. A., *Mathematical Modeling of Physical Networks.* New York: Macmillan, 1968.

7. Booth, T. L., *Sequential Machines and Automata Theory.* New York: Wiley, 1967.

8. Boulding, K. E., "General Systems as a Point of View." In *Views on General Systems Theory*, M. D. Mesarovic (Editor), pp. 25–38. New York: Wiley, 1964.

9. Bremermann, H. J., "Optimization Through Evolution and Recombination." In *Self-Organizing Systems*, M. C. Yovits et al. (Editors), pp. 93–106. Washington, D.C.: Spartan Books, 1962.

10. Brown, J. G., "A Note on Fuzzy Sets." *Information and Control*, Vol. 18, No. 1, pp. 32–39, 1971.

11. Churchman, C. W., *The Systems Approach.* New York: Dell Publishing Co., 1968.

12. Eckman, D. P., and M. D. Mesarovic, "On Some Basic Concepts of General Systems Theory." In *Proceedings of the Third International Conference on Cybernetics*, pp. 104–118. Namur, Belgium: 1961.

13. Friedman, G. J., and C. T. Leondes, "Constraint Theory." *IEEE Transactions on Systems Science and Cybernetics*, Vol. SSC-5, Nos. 1, 2, 3, pp. 48–56, 132–140, 191–199, January, April, July 1969.

14. Gill, A., *Introduction to the Theory of Finite-State Machines.* New York: McGraw-Hill, 1962.

15. **Ginsburg, S.,** *An Introduction to Mathematical Machine Theory.* Reading, Mass.: Addison-Wesley, 1962.

16. **Ginsburg, S.,** *The Mathematical Theory of Context Free Languages.* New York: McGraw-Hill, 1966.

17. **Goguen, J. A.,** "The Logic of Inexact Concepts." *Synthese,* Vol. 19, pp. 325–373, 1968–69.

18. **Gukhman, A. A.,** *Introduction to the Theory of Similarity.* New York: Academic Press, 1965.

19. **Hammer, P. C.,** "Extended Topology and Systems." *Mathematical Systems Theory,* Vol. 1, No. 1, pp. 135–142, 1967.

20. **Hartmanis, J.,** and **R. E. Stearns,** *Algebraic Structure Theory of Sequential Machines.* Englewood Cliffs, N.J.: Prentice-Hall, 1966.

21. **Hopcroft, J. E.,** and **J. D. Ullman,** *Formal Languages and Their Relation to Automata.* Reading, Mass.: Addison-Wesley, 1969.

22. **Hunt, B. R.,** "Polynomial Representation of Finite-State Machines." *IEEE Transactions on Systems Science and Cybernetics,* Vol. SSC-5, No. 1, pp. 94–96, January 1969.

23. **Kalman, R. E.,** et al. (Editors), *Topics in Mathematical System Theory.* New York: McGraw-Hill, 1969.

24. **Klir, G. J.,** and **M. Valach,** *Cybernetic Modelling.* London: ILIFFE Books, 1967; Princeton, N.J.: Van Nostrand, 1967.

25. **Klir, G. J.,** *An Approach to General Systems Theory.* New York: Van Nostrand Reinhold, 1969.

26. **Klir, G. J.,** and **M. A. Marin.** "A Multimodel and Computer Oriented Methodology for Synthesis of Sequential Discrete Systems." *IEEE Transactions on Systems Science and Cybernetics,* Vol. SSC-6, No. 1, pp. 40–48, January 1970.

27. **Klir, G. J.,** "On the Relation Between Cybernetics and General Systems Theory." In *Progress in Cybernetics,* J. Rose (Editor), pp. 155–165. New York: Gordon and Breach, 1970.

28. **Klir, G. J.,** "A Study of Organizations of Self-Organizing Systems." *Proceedings of the Sixth International Conference on Cybernetics,* pp. 81–98. Namur, Belgium: 1971.

29. **Klir, G. J.,** *An Introduction to the Methodology of Switching Circuits.* New York: Van Nostrand Reinhold, 1972.

30. **Koenig, H.,** et al., *Analysis of Discrete Physical Systems.* New York: McGraw-Hill, 1967.

31. **Mesarovic, M. D.,** "On Some Metamathematical Results as Properties of General Systems." *Mathematical Systems Theory,* Vol. 2, No. 2, pp. 357–361, 1968.

32. **Mesarovic, M. D.,** et al., *Theory of Hierarchical, Multilevel Systems.* New York: Academic Press, 1970.

33. **Paz, A.,** *Introduction to Probabilistic Automata.* New York: Academic Press, 1971.

34. **Shannon, C. E.,** and **J. McCarthy** (Editors), *Automata Studies.* Princeton, N.J.: Princeton University Press, 1956.

35. **Skoglund, V.,** *Similitude: Theory and Applications.* Scranton, Pa., International Textbook Co., 1967.

36. **Svoboda, A.,** "Synthesis of Logical Systems of Given Activity." *IEEE Transactions on Computers,* Vol. EC–12, No. 6, pp. 904–910, December 1963.

37. **Svoboda, A.,** "Behavior Classification of Digital Systems." In *Information Processing Machines,* Vol. 10, pp. 25–42. Prague: Academia, 1964.

38. **Thorn, D. C.,** *An Introduction to Generalized Circuits.* Belmont, Calif.: Wadsworth Publishing Co., 1963.

39. **Vasspeg, K.,** "On Organizing of Systems." In *Information Processing Machines,* Vol. 11, pp. 167–176. Prague: Academia, 1965.

40. **Whyte, L. L.,** et al. (Editors), *Hierarchical Structures.* New York, American Elsevier, 1969.

41. **Wymore, A. W.,** *A Mathematical Theory of Systems Engineering: The Elements.* New York: Wiley, 1967.

42. **Zadeh, L. A.,** and **C. A. Desoer,** *Linear System Theory.* New York: McGraw-Hill, 1963.

43. **Zadeh, L. A.,** "Fuzzy Sets." *Information and Control,* Vol. 8, No. 3, pp. 338–353, June 1965.

44. **Zadeh, L. A.,** "Fuzzy Sets and Systems." In *System Theory,* J. Fox, (Editor), pp. 29–37. New York: Polytechnic Press, 1965.

45. **Zadeh, L. A.,** and **E. Polak** (Editors), *System Theory.* New York: McGraw-Hill, 1969.

Part I

Historical and General Aspects

1. The History and Status of General Systems Theory

LUDWIG VON BERTALANFFY

Center for Theoretical Biology, State University of New York at Buffalo

1.1. HISTORICAL PRELUDE

In order to evaluate the modern "systems approach," it is advisable to look at the systems idea not as an ephemeral fashion or recent technique, but in the context of the history of ideas. (For an introduction and a survey of the field see [15], with an extensive bibliography and Suggestions for Further Reading in the various topics of general systems theory.)

In a certain sense it can be said that the notion of system is as old as European philosophy. If we try to define the central motif in the birth of philosophical-scientific thinking with the Ionian pre-Socratics of the sixth century B.C., one way to spell it out would be as follows. Man in early culture, and even primitives of today, experience themselves as being "thrown" into a hostile world, governed by chaotic and incomprehensible demonic forces which, at best, may be propitiated or influenced by way of magical practices. Philosophy and its descendant, science, was born when the early Greeks learned to consider or find, in the experienced world, an order or *kosmos* which was intelligible and hence controllable by thought and rational action.

One formulation of this cosmic order was the Aristotelian world view with its holistic and teleological notions. Aristotle's statement, "The whole is more than the sum of its parts," is a definition of the basic system problem which is

still valid. Aristotelian teleology was eliminated in the later development of Western science, but the problems contained in it, such as the order and goal-directedness of living systems, were negated and by-passed rather than solved. Hence the basic system problem is still not obsolete.

A more detailed investigation would enumerate a long array of thinkers who, in one way or another, contributed notions to what nowadays we call systems theory. If we speak of hierarchic order, we use a term introduced by the Christian mystic, Dionysius the Aeropagite, although he was speculating about the choirs of angels and the organism of the Church. Nicholas of Cusa [5], that profound thinker of the fifteenth century, linking Medieval mysticism with the first beginnings of modern science, introduced the notion of the *coincidentia oppositorum*, the opposition or indeed fight among the parts within a whole which nevertheless forms a unity of higher order. Leibniz's hierarchy of monads looks quite like that of modern systems; his *mathesis universalis* presages an expanded mathematics which is not limited to quantitative or numerical expressions and is able to formalize all conceptual thinking. Hegel and Marx emphasized the dialectic structure of thought and of the universe it produces: the deep insight that no proposition can exhaust reality but only approaches its coincidence of opposites by the dialectic process of thesis, antithesis, and synthesis. Gustav Fechner, known as the author of the psychophysical law, elaborated, in the way of the nature philosophers of the nineteenth century, supraindividual organizations of higher order than the usual objects of observation—for example, life communities and the entire earth, thus romantically anticipating the ecosystems of modern parlance. Incidentally, the present writer wrote a doctoral thesis on this topic in 1925.

Even such a rapid and superficial survey as the preceding one tends to show that the problems with which we are nowadays concerned under the term "system" were not "born yesterday" out of current questions of mathematics, science, and technology. Rather, they are a contemporary expression of perennial problems which have been recognized for centuries and discussed in the language available at the time.

One way to circumscribe the Scientific Revolution of the sixteenth-seventeenth centuries is to say that it replaced the descriptive-metaphysical conception of the universe epitomized in Aristotle's doctrine by the mathematical-positivistic or Galilean conception. That is, the vision of the world as a teleological cosmos was replaced by the description of events in causal, mathematical laws.

We say "replaced," not "eliminated," for the Aristotelian dictum of the whole that is more than its parts still remained. We must strongly emphasize that order or organization of a whole or system, transcending its parts when these are considered in isolation, is nothing metaphysical, not an anthropo-

morphic supersitition or a philosophical speculation; it is a fact of observation encountered whenever we look at a living organism, a social group, or even an atom.

Science, however, was not well prepared to deal with this problem. The second maxim of Descartes' *Discours de la Méthode* was "to break down every problem into as many separate simple elements as might be possible." This, similarly formulated by Galileo as the "resolutive" method, was the conceptual "paradigm" [35] of science from its foundation to modern laboratory work: that is, to resolve and reduce complex phenomena into elementary parts and processes.

This method worked admirably well insofar as observed events were apt to be split into isolable causal chains, that is, relations between two or a few variables. It was at the root of the enormous success of physics and the consequent technology. But questions of many-variable problems always remained. This was the case even in the three-body problem of mechanics; the situation was aggravated when the organization of the living organism or even of the atom, beyond the simplest proton-electron system of hydrogen, was concerned.

Two principal ideas were advanced in order to deal with the problem of order or organization. One was the comparison with man-made machines; the other was to conceive of order as a product of chance. The first was epitomized by Descartes' *bête machine*, later expanded to the *homme machine* of Lamettrie. The other is expressed by the Darwinian idea of natural selection. Again, both ideas were highly successful. The theory of the living organism as a machine in its various disguises—from a mechanical machine or clockwork in the early explanations of the iatrophysicists of the seventeenth century, to later conceptions of the organism as a caloric, chemodynamic, cellular, and cybernetic machine [13]—provided explanations of biological phenomena from the gross level of the physiology of organs down to the submicroscopic structures and enzymatic processes in the cell. Similarly, organismic order as a product of random events embraced an enormous number of facts under the title of "synthetic theory of evolution" including molecular genetics and biology.

Notwithstanding the singular success achieved in the explanation of ever more and finer life processes, basic questions remained unanswered. Descartes' "animal machine" was a fair enough principle to explain the admirable order of processes found in the living organism. But then, according to Descartes, the "machine" had God for its creator. The evolution of machines by events at random rather appears to be self-contradictory. Wristwatches or nylon stockings are not as a rule found in nature as products of chance processes, and certainly the mitochondrial "machines" of enzymatic organization in even the simplest cell or nucleoprotein molecules are incomparably

more complex than a watch or the simple polymers which form synthetic fibers. "Survival of the fittest" (or "differential reproduction" in modern terminology) seems to lead to a circuitous argument. Self-maintaining systems must exist before they can enter into competition which leaves systems with higher selective value or differential reproduction predominant. That self-maintenance, however, is the explicandum; it is not provided by the ordinary laws of physics. Rather, the second law of thermodynamics prescribes that ordered systems in which irreversible processes take place tend toward most probable states and hence toward destruction of existing order and ultimate decay [16].

Thus neovitalistic currents, represented by Driesch, Bergson, and others, reappeared around the turn of the present century, advancing quite legitimate arguments which were based essentially on the limits of possible regulations in a "machine," of evolution by random events, and on the goal-directedness of action. They were able, however, to refer only to the old Aristotelian "entelechy" under new names and descriptions, that is, a supernatural, organizing principle or "factor."

Thus the "fight on the concept of organism in the first decades of the twentieth century," as Woodger [56] nicely put it, indicated increasing doubts regarding the "paradigm" of classical science, that is, the explanation of complex phenomena in terms of isolable elements. This was expressed in the question of "organization" found in every living system; in the question whether "random mutations *cum* natural selection provide all the answers to the phenomena of evolution" [32] and thus of the organization of living things; and in the question of goal-directedness, which may be denied but in some way or other still raises its ugly head.

These problems were in no way limited to biology. Psychology, in gestalt theory, similarly and even earlier posed the question that psychological wholes (e.g., perceived *gestalten*) are not resolvable into elementary units such as punctual sensations and excitations in the retina. At the same time sociology [49, 50] came to the conclusion that physicalistic theories, modeled according to the Newtonian paradigm or the like, were unsatisfactory. Even the atom appeared as a minute "organism" to Whitehead.

1.2. FOUNDATIONS OF GENERAL SYSTEMS THEORY

In the late 1920's von Bertalanffy wrote:

Since the fundamental character of the living thing is its organization, the custom-ary investigation of the single parts and processes cannot provide a complete explanation of the vital phenomena. This investigation gives us no information

about the coordination of parts and processes. Thus the chief task of biology must be to discover the laws of biological systems (at all levels of organization). We believe that the attempts to find a foundation for theoretical biology point at a fundamental change in the world picture. This view, considered as a method of investigation. we shall call "*organismic biology*" and, as an attempt at an explanation, "*the system theory of the organism*" [7, pp. 64 ff., 190, 46, condensed].

Recognized "as something new in biological literature" [43], the organismic program became widely accepted. This was the germ of what later became known as general systems theory. If the term "organism" in the above statements is replaced by other "organized entities," such as social groups, personality, or techonological devices, this is the program of systems theory.

The Aristotelian dictum of the whole being more than its parts, which was neglected by the mechanistic conception, on the one hand, and which led to a vitalistic demonology, on the other, has a simple and even trivial answer— trivial, that is, in principle, but posing innumerable problems in its elaboration:

The properties and modes of action of higher levels are not explicable by the summation of the properties and modes of action of their components *taken in isolation*. If, however, we know the *ensemble* of the components and the *relations existing between them*, then the higher levels are derivable from the components [10, p. 148].

Many (including recent) discussions of the Aristotelian paradox and of reductionism have added nothing to these statements: in order to understand an organized whole we must know both the parts and the relations between them.

This, however, defines the trouble. For "normal" science in Thomas Kuhn's sense, that is, science as conventionally practiced, was little adapted to deal with "relations" in systems. As Weaver [51] said in a well-known statement, classical science was concerned with one-way causality or relations between two variables, such as the attraction of the sun and a planet, but even the three-body problem of mechanics (and the corresponding problems in atomic physics) permits no closed solution by analytical methods of classical mechanics. Also, there were descriptions of "unorganized complexity" in terms of statistics whose paradigm is the second law of thermodynamics. However, increasing with the progress of observation and experiment, there loomed the problem of "organized complexity," that is, of interrelations between many but not infinitely many components.

Here is the reason why, even though the problems of "system" were ancient and had been known for many centuries, they remained "philosophical" and did not become a "science." This was so because mathematical techniques were lacking and the problems required a new epistemology; the whole force of "classical" science and its success over the centuries militated

against any change in the fundamental paradigm of one-way causality and resolution into elementary units.

The quest for a new "gestalt mathematics" was repeatedly raised a considerable time ago, in which not the notion of quantity but rather that of relations, that is, of form and order, would be fundamental [10, p. 159f.]. However, this demand became realizable only with new developments.

The notion of general systems theory was first formulated by von Bertalanffy, orally in the 1930's and in various publications after World War II:

> There exist models, principles and laws that apply to generalized systems or their subclasses irrespective of their particular kind, the nature of the component elements, and the relations or "forces" between them. We postulate a new discipline called *General System Theory*. General System Theory is a logico-mathematical field whose task is the formulation and derivation of those general principles that are applicable to "systems" in general. In this way, exact formulations of terms such as wholeness and sum, differentiation, progressive mechanization, centralization, hierarchial order, finality and equifinality, etc., become possible, terms which occur in all sciences dealing with "systems" and imply their logical homology (von Bertalanffy, 1947, 1955; reprinted in [15, pp. 32, 253].

The proposal of general systems theory had precursors as well as independent simultaneous promoters. Köhler came near to generalizing gestalt theory into general systems theory [33]. Although Lotka did not use the term "general system theory," his discussion of systems of simultaneous differential equations [39] remained basic for subsequent "dynamical" system theory. Volterra's equations [21], originally developed for the competition of species, are applicable to generalized kinetics and dynamics. Ashby, in his early work [1], independently used the same system equations as von Bertalanffy employed, although deriving different consequences.

Von Bertalanffy outlined "dynamical" system theory (see Section 1.3(a)), and gave mathematical descriptions of system properties (such as wholeness, sum, growth, competition, allometry, mechanization, centralization, finality, and equifinality), derived from the system description by simultaneous differential equations. Being a practicing biologist, he was particularly interested in developing the theory of "open systems," that is, systems exchanging matter with environment as every "living" system does. Such theory did not then exist in physical chemistry. The theory of open systems stands in manifold relationships with chemical kinetics in its biological, theoretical, and technological aspects, and with the thermodynamics of irreversible processes, and provides explanations for many special problems in biochemistry, physiology, general biology, and related areas. It is correct to say that, apart from control theory and the application of feedback models, the theory of *Fliessgleichgewicht* and open systems [8, 12] is the part of

general systems theory most widely applied in physical chemistry, biophysics, simulation of biological processes, physiology, pharmacodynamics, and so forth [15]. The forecast also proved to be correct that the basic areas of physiology, that is, metabolism, excitation, and morphogenesis (more specifically, the theory of regulation, cell permeability, growth, sensory excitation, electrical stimulation, center function, etc.), would "fuse into an integrated theoretical field under the guidance of the concept of open system" [6, Vol. II, pp. 49 ff.; also 15, p. 137 f.].

The intuitive choice of the open system as a general system model was a correct one. Not only from the physical viewpoint is the "open system" the more general case (because closed systems can always be obtained from open ones by equating transport variables to zero); it also is the general case mathematically because the system of simultaneous differential equations (equations of motion) used for description in dynamical system theory is the general form from which the description of closed systems derives by the introduction of additional constraints (e.g., conservation of mass in a closed chemical system) (cf. [46], p. 80 f.).

At first the project was considered to be fantastic. A well-known ecologist, for example, was "hushed into awed silence" by the preposterous claim that general system theory constituted a new realm of science [24], not foreseeing that it would become a legitimate field and the subject of university instruction within some 15 years.

Many objections were raised against its feasibility and legitimacy [17]. It was not understood then that the exploration of the properties, models, and laws of "systems" is not a hunt for superficial analogies, but rather poses basic and difficult problems which are partly still unsolved [10, p. 200 f.].

According to the program, "system laws" manifest themselves as analogies or "logical homologies" of laws that are formally identical but pertain to quite different phenomena or even appear in different disciplines. This was shown by von Bertalanffy in examples which were chosen as intentionally simple illustrations, but the same principle applies to more sophisticated cases, such as the following:

It is a striking fact that biological systems as diverse as the central nervous system, and the biochemical regulatory network in cells should be strictly analogous . . . . It is all the more remarkable when it is realized that this particular analogy between different systems at different levels of biological organization is but one member of a large class of such analogies [45].

It appeared that a number of researchers, working independently and in different fields, had arrived at similar conclusions. For example, Boulding wrote to the present author:

I seem to have come to much the same conclusions as you have reached, though approaching it from the direction of economics and the social sciences rather than from biology—that there is a body of what I have been calling "general empirical theory," or "general system theory" in your excellent terminology, which is of wide applicability in many different disciplines [15, p. 14; cf. 18].

This spreading interest led to the foundation of the Society for General Systems Research (initially named the Society for the Advancement of General System Theory), an affiliate of the American Association for the Advancement of Science. The formation of numerous local groups, the task group on "General Systems Theory and Psychiatry" in the American Psychiatric Association, and many similar working groups, both in the United States and in Europe, followed, as well as various meetings and publications. The program of the Society formulated in 1954 may be quoted because it remains valid as a research program in general systems theory:

Major functions are to: (1) investigate the isomorphy of concepts, laws, and models in various fields, and to help in useful transfers from one field to another; (2) encourage the development of adequate theoretical models in the fields which lack them; (3) minimize the duplication of theoretical effort in different fields; (4) promote the unity of science through improving communication among specialists.

In the meantime a different development had taken place. Starting from the development of self-directing missiles, automation and computer technology, and inspired by Wiener's work, the cybernetic movement became ever more influential. Although the starting point (technology versus basic science, especially biology) and the basic model (feedback circuit versus dynamic system of interactions) were different, there was a communality of interest in problems of organization and teleological behavior. Cybernetics too challenged the "mechanistic" conception that the universe was based on the "operation of anonymous particles at random" and emphasized "the search for new approaches, for new and more comprehensive concepts, and for methods capable of dealing with the large wholes of organisms and personalities" [25]. Although it is incorrect to describe modern systems theory as "springing out of the last war effort" [19]—in fact, it had roots quite different from military hardware and related technological developments—cybernetics and related approaches were independent developments which showed many parallelisms with general system theory.

1.3. TRENDS IN GENERAL SYSTEMS THEORY

This brief historical survey cannot attempt to review the many recent developments in general systems theory and the systems approach. For a critical discussion of the various approaches see [30, pp. 97 ff.], and [27, Book II].

With the increasing expansion of systems thinking and studies, the definition of general systems theory came under renewed scrutiny. Some indication as to its meaning and scope may therefore be pertinent. The term "general system theory," was introduced by the present author, deliberately, in a catholic sense. One may, of course, limit it to its "technical" meaning in the sense of mathematical theory (as is frequently done), but this appears unadvisable because there are many "system" problems asking for "theory" which is not presently available in mathematical terms. So the name "general systems theory" may be used broadly, in a way similar to our speaking of the "theory of evolution," which comprises about everything ranging from fossil digging and anatomy to the mathematical theory of selection; or "behavior theory," which extends from bird watching to sophisticated neurophysiological theories. It is the introduction of a new paradigm that matters.

(a) Systems science; mathematical systems theory. Broadly speaking, three main aspects can be indicated which are not separable in content but are distinguishable in intention. The first may be circumscribed as *systems science*, that is, scientific exploration and theory of "systems" in the various sciences (e.g., physics, biology, psychology, social sciences), and general systems theory as the doctrine of principles applying to all (or defined subclasses of) systems.

Entities of an essentially new sort are entering the sphere of scientific thought. Classical science in its various disciplines, such as chemistry, biology, psychology, or the social sciences, tried to isolate the elements of the observed universes—chemical compounds and enzymes, cells, elementary sensations, freely competing individuals, or whatever else may be the case—in the expectation that by putting them together again, conceptually or experimentally, the whole or system—cell, mind, society—would result and would be intelligible. We have learned, however, that for an understanding not only the elements but their interrelations as well are required—say, the interplay of enzymes in a cell, the interactions of many conscious and unconscious processes in the personality, the structure and dynamics of social systems, and so forth. Such problems appear even in physics, for example, in the interaction of many generalized "forces" and "fluxes" (irreversible thermodynamics; cf. Onsager reciprocal relations), or in the development of nuclear physics, which "requires much experimental work, as well as the development of additional powerful methods for the handling of systems with many, but not infinitely many, particles" [23]. This requires, first, the exploration of the many systems in our observed universe in their own right and specificities. Second, it turns out that there are general aspects, correspondences, and isomorphisms common to "systems." This is the domain of *general systems theory*. Indeed, such parallelisms or isomorphisms appear (sometimes surprisingly) in otherwise totally different "systems."

General systems theory, then, consists of the scientific exploration of "wholes" and "wholeness" which, not so long ago, were considered to be metaphysical notions transcending the boundaries of science. Novel concepts, models, and mathematical fields have developed to deal with them. At the same time, the interdisciplinary nature of concepts, models, and principles applying to "systems" provides a possible approach toward the unification of science.

The goal obviously is to develop general systems theory in mathematical terms (a "logico-mathematical field," as this author wrote in the early statement cited in Section 1.2) because mathematics is the exact language permitting rigorous deductions and confirmation (or refusal) of theory. Mathematical systems theory has become an extensive and rapidly growing field. "System" being a new "paradigm" (in the sense of Thomas Kuhn), contrasting to the predominant, elementalistic approach and conceptions, it is not surprising that a variety of approaches have developed, differing in emphasis, focus of interest, mathematical techniques, and other respects. These elucidate different aspects, properties and principles of what is comprised under the term "system," and thus serve different purposes of theoretical or practical nature. The fact that "system theories" by various authors look rather different is, therefore, not an embarrassment or the result of confusion, but rather a healthy development in a new and growing field, and indicates presumably necessary and complementary aspects of the problem. The existence of different descriptions is nothing extraordinary and is often encountered in mathematics and science, from the geometrical or analytical description of a curve to the equivalence of classical thermodynamics and statistical mechanics to that of wave mechanics and particle physics. Different and partly opposing approaches should, however, tend toward further integration, in the sense that one is a special case within another, or that they can be shown to be equivalent or complementary. Such developments are, in fact, taking place.

System-theoretical approaches include general system theory (in the narrower sense), cybernetics, theory of automata, control theory, information theory, set, graph and network theory, relational mathematics, game and decision theory, computerization and simulation, and so forth. The somewhat loose term "approaches" is used deliberately because the list contains rather different things, for example, models (such as those of open system, feedback, logical automaton), mathematical techniques (e.g., theory of differential equations, computer methods, set, graph theory), and newly formed concepts or parameters (information, rational game, decision, etc.). These approaches concur, however, in that, in one way or the other, they relate to "system problems," that is, problems of interrelations within a superordinate "whole." Of course, these are not isolated but frequently overlap, and the same problem

can be treated mathematically in different ways. Certain typical ways of describing "systems" can be indicated; their elaboration is due, on the one hand, to theoretical problems of "systems" as such and in relation to other disciplines, and, on the other hand, to problems of the technology of control and communication.

No mathematical development or comprehensive review can be given here. The following remarks, however, may convey some intuitive understanding of the various approaches and the way in which they relate to each other.

It is generally agreed that "system" is a *model* of general nature, that is, a conceptual analog of certain rather universal traits of observed entities. The use of models or analog constructs is the general procedure of science (and even of everyday cognition), as it is also the principle of analog simulation by computer. The difference from conventional disciplines is not essential but lies rather in the degree of generality (or abstraction): "system" refers to very general characteristics partaken by a large class of entities conventionally treated in different disciplines. Hence the interdisciplinary nature of general systems theory; at the same time, its statements pertain to formal or structural commonalities abstracting from the "nature of elements and forces in the system" with which the special sciences (and explanations in these) are concerned. In other words, system-theoretical arguments pertain to, and have predictive value, inasmuch as such general structures are concerned. Such "explanation in principle" may have considerable predictive value; for specific explanation, introduction of the special system conditions is naturally required.

A system may be defined as a set of elements standing in interrelation among themselves and with the environment. This can be expressed mathematically in different ways. Several typical ways of system description can be indicated.

One approach or group of investigations may, somewhat loosely, be circumscribed as *axiomatic*, inasmuch as the focus of interest is a rigorous definition of system and the derivation, by modern methods of mathematics and logic, of its implications. Among other examples are the system descriptions by Mesarovic [41], Maccia and Maccia [40], Beier and Laue [4] (set theory), Ashby [2] (state-determined systems), and Klir [30] (UC = set of all couplings between the elements and the elements and environment; ST = set of all states and all transitions between states).

Dynamical system theory is concerned with the changes of systems in time. There are two principal ways of description: internal and external [47].

Internal description or "classical" system theory (foundations in [9], [11], and [15, pp. 54 ff.]; comprehensive presentation in [46]; an excellent introduction into dynamical system theory and the theory of open systems,

following the line of the present author, in [3]) defines a system by a set of n measures, called state variables. Analytically, their change in time is typically expressed by a set of n simultaneous, first-order differential equations:

$$\frac{dQ_n}{dt} = f_i(Q_1, Q_2, \ldots, Q_n). \tag{1.1}$$

These are called dynamical equations or equations of motion. The set of differential equations permits a formal expression of system properties, such as wholeness and sum, stability, mechanization, growth, competition, final and equifinal behavior and others [9, 11, 15]. The behavior of the system is described by the theory of differential equations (ordinary, first-order, if the definition of the system by Eq. 1.1 is accepted), which is a well-known and highly developed field of mathematics. However, as was mentioned previously, system considerations pose quite definite problems. For example, the theory of stability has developed only recently in conjunction with problems of control (and system): the Liapunov († 1918) functions date from 1892 (in Russian; 1907 in French), but their significance was recognized only recently, especially through the work of mathematicians of the U.S.S.R.

Geometrically, the change of the system is expressed by the trajectories that the state variables traverse in the state space, that is, the n-dimensional space of possible location of these variables. Three types of behavior may be distinguished and defined as follows:

1. A trajectory is called *asymptotically stable* if all trajectories sufficiently close to it at $t = t_0$ approach it asymptotically when $t \to \infty$.

2. A trajectory is called *neutrally stable* if all trajectories sufficiently close to it at $t = 0$ remain close to it for all later time but do not necessarily approach it asymptotically.

3. A trajectory is called *unstable* if the trajectories close to it at $t = 0$ do not remain close to it as $t \to \infty$.

These correspond to solutions approaching a time-independent state (equilibrium, steady state), periodic solutions, and divergent solutions, respectively.

A time-independent state,

$$f_i(Q_1, Q_2, \ldots, Q_n) = 0, \tag{1.2}$$

can be considered as a trajectory degenerated into a single point. Then, readily visualizable in two-dimensional projection, the trajectories may converge toward a stable node represented by the equilibrium point, may approach it as a stable focus in damped oscillations, or may cycle around it in undamped oscillations (stable solutions). Or else, they may diverge from an

unstable node, wander away from an unstable focus in oscillations, or from a saddle point (unstable solutions).

A central notion of dynamical theory is that of *stability*, that is, the response of a system to perturbation. The concept of stability originates in mechanics (a rigid body is in stable equilibrium if it returns to its original position after sufficiently small displacement; a motion is stable if insensitive to small perturbations), and is generalized to the "motions" of state variables of a system. This question is related to that of the existence of equilibrium states. Stability can be analyzed, therefore, by explicit solution of the differential equations describing the system (so-called indirect method, based essentially on discussion of the eigenwerte λ_i of Eq. 1.1). In the case of nonlinear systems, these equations have to be linearized by development into Taylor series and retention of the first term. Linearization, however, pertains only to stability in the vicinity of equilibrium. But stability arguments without actual solution of the differential equations (direct method) and for nonlinear systems are possible by introduction of so-called *Liapunov functions*; these are essentially generalized energy functions, the sign of which indicates whether or not an equilibrium is asymptotically stable [28, 36].

Here the relation of dynamical system theory to control theory becomes apparent; control means essentially that a system which is not asymptotically stable is made so by incorporating a controller, counteracting the motion of the system away from the stable state. For this reason the theory of stability in internal description or dynamical system theory converges with the theory of (linear) control or feedback systems in external description (see below; cf. [48]).

Description by ordinary differential equations (Eq. 1.1) abstracts from variations of the state variables in space which would be expressed by partial differential equations. Such field equations are, however, more difficult to handle. Ways of overcoming this difficulty are to assume complete "stirring," so that distribution is homogeneous within the volume considered; or to assume the existence of compartments to which homogeneous distribution applies, and which are connected by suitable interactions (compartment theory) [44].

In *external description*, the system is considered as a "black box"; its relations to the environment and other systems are presented graphically in block and flow diagrams. The system description is given in terms of inputs and outputs (*Klemmenverhalten* in German terminology); its general form are transfer functions relating input and output. Typically, these are assumed to be linear and are represented by discrete sets of values (cf. yes-no decisions in information theory, Turing machine). This is the language of control technology; external description, typically, is given in terms of communication (exchange of information between system and environment and within the

system) and control of the system's function with respect to environment (feedback), to use Wiener's definition of cybernetics.

As mentioned, internal and external descriptions largely coincide with descriptions by continuous or discrete functions. These are two "languages" adapted to their respective purposes. Empirically, there is an obvious contrast between regulations due to the free interplay of forces within a dynamical system, and regulations due to constraints imposed by structural feedback mechanisms [15], for example, the "dynamic" regulations in a chemical system or in the network of reactions in a cell on the one hand, and control by mechanisms such as a thermostat or homeostatic nervous circuit on the other. Formally, however, the two "languages" are related and in certain cases demonstrably translatable. For example, an input-output function can (under certain conditions) be developed as a linear nth-order differential equation, and the terms of the latter can be considered as (formal) "state variables"; while their physical meaning remains indefinite, formal "translation" from one language into the other is possible.

In certain cases—for example, the two-factor theory of nerve excitation (in terms of "excitatory and inhibitory factors" or "substances") and network theory (McCulloch nets of "neurons")—description in dynamical system theory by continuous functions and description in automata theory by digital analogs can be shown to be equivalent [45]. Similarly predator-prey systems, usually described dynamically by Volterra equations, can also be expressed in terms of cybernetic feedback circuits [55]. These are two-variable systems. Whether a similar "translation" can be effectuated in many-variables systems remains (in the present writer's opinion) to be seen.

Internal description is essentially "structural," that is, it tries to describe the systems' behavior in terms of state variables and their interdependence. External description is "functional"; the system's behavior is described in terms of its interaction with the environment.

As this sketchy survey shows, considerable progress has been made in mathematical systems theory since the program was enunciated and inaugurated some 25 years ago. A variety of approaches, which, however, are connected with each other, have been developed.

Today mathematical system theory is a rapidly growing field, but it is natural that basic problems, such as those of hierarchical order [53], are approached only slowly and presumably will need novel ideas and theories. "Verbal" descriptions and models (e.g., [20], [31], [42], [52]) are not expendable. Problems must be intuitively "seen" and recognized before they can be formalized mathematically. Otherwise, mathematical formalism may impede rather than expedite the exploration of very "real" problems.

A strong system-theoretical movement has developed in psychiatry, largely through the efforts of Gray [26]. The same is true of the behavioral

sciences [20] and also of certain areas in which such a development was quite unexpected, at least by the present writer—for example, theoretical geography [29]. Sociology was stated as being essentially "a science of social systems" [14]; not foreseen was, for instance, the close parallelism of general system theory with French structuralism (e.g., Piaget, Levy-Strauss; cf. [37]) and the influence exerted on American functionalism in sociology ([22]: see especially pp. 2, 96, 141).

(*b*) *Systems technology.* The second realm of general systems theory is *systems technology*, that is, the problems arising in modern technology and society, including both "hardware" (control technology, automation, computerization, etc.) and "software" (application of system concepts and theory in social, ecological, economical, etc., problems). We can only allude to the vast realm of techniques, models, mathematical approaches, and so forth, summarized as systems engineering or under similar denominations, in order to place it into the perspective of the present study.

Modern technology and society have become so complex that the traditional branches of technology are no longer sufficient; approaches of a holistic or systems, and generalist and interdisciplinary, nature became necessary. This is true in many ways. Modern engineering includes fields such as circuit theory, cybernetics as the study of "communication and control" (Wiener [54]), and computer techniques for handling "systems" of a complexity unamenable to classical methods of mathematics. Systems of many levels ask for scientific control: ecosystems, the disturbance of which results in pressing problems like pollution; formal organizations like bureaucracies, educational institutions, or armies; socioeconomic systems, with their grave problems of international relations, politics, and deterrence. Irrespective of the questions of how far scientific understanding (contrasted to the admission of irrationality of cultural and historical events) is possible, and to what extent scientific control is feasible or even desirable, there can be no dispute that these are essentially "system" problems, that is, problems involving interrelations of a great number of "variables." The same applies to narrower objectives in industry, commerce, and armament.

The technological demands have led to novel conceptions and disciplines, some displaying great originality and introducing new basic notions such as control and information theory, game, decision theory, the theory of circuits, of queuing and others. Again it transpired that concepts and models (such as feedback, information, control, stability, circuits) which originated in certain specified fields of technology have a much broader significance, are of an interdisciplinary nature, and are independent of their special realizations, as exemplified by isomorphic feedback models in mechanical, hydrodynamic, electrical, biological and other systems. Similarly, developments originating in

pure and in applied science converge, as in dynamical system theory and control theory. Again, there is a spectrum ranging from highly sophisticated mathematical theory to computer simulation to more or less informal discussion of system problems.

(*c*) *Systems philosophy*. Third, there is the realm of *systems philosophy* [38], that is, the reorientation of thought and world view following the introduction of "system" as a new scientific paradigm (in contrast to the analytic, mechanistic, linear-causal paradigm of classical science). Like every scientific theory of broader scope, general systems theory has its "meta-scientific" or philosophical aspects. The concept of "system" constitutes a new "paradigm," in Thomas Kuhn's phrase, or a new "philosophy of nature," in the present writer's [14] words, contrasting the "blind laws of nature" of the mechanistic world view and the world process as a Shakespearean tale told by an idiot, with an organismic outlook of the "world as a great organization."

First, we must find out the "nature of the beast": what is meant by "system," and how systems are realized at the various levels of the world of observation. This is *systems ontology*.

What is to be defined and described as system is not a question with an obvious or trivial answer. It will be readily agreed that a galaxy, a dog, a cell, and an atom are "systems." But in what sense and what respects can we speak of an animal or a human society, personality, language, mathematics, and so forth as "systems"?

We may first distinguish *real systems*, that is, entities perceived in or inferred from observation and existing independently of an observer. On the other hand, there are *conceptual systems*, such as logic or mathematics, which essentially are symbolic constructs (but also including, e.g., music); with *abstracted systems* (science) [42] as a subclass, that is, conceptual systems corresponding with reality. However, the distinction is by no means as sharp as it would appear.

Apart from philosophical interpretation (which would take us into the question of metaphysical realism, idealism, phenomenalism, etc.) we would consider as "objects" (which partly are "real systems") entities given by perception because they are discrete in space and time. We do not doubt that a pebble, a table, an automobile, an animal, or a star (and in a somewhat different sense an atom, a molecule, and a planetary system) are "real" and existent independently of observation. Perception, however, is not a reliable guide. Following it, we "see" the sun revolving around the earth, and certainly do not see that a solid piece of matter like a stone "really" is mostly empty space with minute centers of energy dispersed in astronomical distances. The spatial boundaries of even what appears to be an obvious object or "thing"

actually are indistinct. From a crystal consisting of molecules, valences stick out, as it were, into the surrounding space; the spatial boundaries of a cell or an organism are equally vague because it maintains itself in a flow of molecules entering and leaving, and it is difficult to tell just what belongs to the "living system" and what does not. Ultimately all boundaries are dynamic rather than spatial.

Hence an object (and in particular a system) is definable only by its cohesion in a broad sense, that is, the interactions of the component elements. In this sense an ecosystem or social system is just as "real" as an individual plant, animal, or human being, and indeed problems like pollution as a disturbance of the ecosystem, or social problems strikingly demonstrate their "reality." Interactions (or, more generally, interrelations), however, are never directly seen or perceived; they are conceptual constructs. The same is true of the objects of our everyday world, which by no means are simply "given" as sense data or simple perceptions but also are constructs based on innate or learned categories, the concordance of different senses, previous experience, learning processes, naming (i.e. symbolic processes), etc. all of which largely determine what we actually "see" or perceive [cf. 34]. Thus the distinction between "real" objects and systems as given in observation and "conceptual" constructs and systems cannot be drawn in any common-sense way.

These are profound problems which can only be indicated in this context. The question for general systems theory is what statements can be made regarding material systems, informational systems, conceptual systems, and other types—questions which are far from being satisfactorily answered at the present time.

This leads to *systems epistemology*. As is apparent from the preceding, this is profoundly different from the epistemology of logical positivism or empiricism, even though it shares the same scientific attitude. The epistemology (and metaphysics) of logical positivism was determined by the ideas of physicalism, atomism, and the "camera theory" of knowledge. These, in view of present-day knowledge, are obsolete. As against physicalism and reductionism, the problems and modes of thought occurring in the biological, behavioral and social sciences require equal consideration, and simple "reduction" to the elementary particles and conventional laws of physics does not appear feasible. Compared to the analytical procedure of classical science, with resolution into component elements and one-way or linear causality as the basic category, the investigation of organized wholes of many variables requires new categories of interaction, transaction, organization, teleology, and so forth, with many problems arising for epistemology, mathematical models and techniques. Furthermore, perception is not a reflection of "real things" (whatever their metaphysical status), and knowledge not a simple approximation to "truth" or "reality." It is an interaction between

knower and known, and thus dependent on a multiplicity of factors of a biological, psychological, cultural, and linguistic nature. Physics itself teaches that there are no ultimate entities like corpuscles or waves existing independently of the observer. This leads to a "perspective" philosophy in which physics, although its achievements in its own and related fields are fully acknowledged, is not a monopolitistic way of knowledge. As opposed to reductionism and theories declaring that reality is "nothing but" (a heap of physical particles, genes, reflexes, drives, or whatever the case may be), we see science as one of the "perspectives" that man, with his biological, cultural, and linguistic endowment and bondage, has created to deal with the universe into which he is "thrown," or rather to which he is adapted owing to evolution and history.

The third part of systems philosophy is concerned with the relations of man and his world, or what is termed *values* in philosophical parlance. If reality is a hierarchy of organized wholes, the image of man will be different from what it is in a world of physical particles governed by chance events as the ultimate and only "true" reality. Rather, the world of symbols, values, social entities and cultures is something very "real"; and its embeddedness in a cosmic order of hierarchies tends to bridge the gulf between C. P. Snow's "two cultures" of science and the humanities, technology and history, natural and social sciences, or in whatever way the antithesis is formulated.

This humanistic concern of general systems theory, as this writer understands it, marks a difference to mechanistically oriented system theorists speaking solely in terms of mathematics, feedback, and technology and so giving rise to the fear that systems theory is indeed the ultimate step toward the mechanization and devaluation of man and toward technocratic society. While understanding and emphasizing the role of mathematics and of pure and applied science, this writer does not see that the humanistic aspects can be evaded unless general systems theory is limited to a restricted and fractional vision.

Thus there is indeed a great and perhaps puzzling multiplicity of approaches and trends in general systems theory. This is understandably uncomfortable to him who wants a neat formalism, to the textbook writer and the dogmatist. It is, however, quite natural in the history of ideas and of science, and particularly in the beginning of a new development. Different models and theories may be apt to render different aspects and so are complementary. On the other hand, future developments will undoubtedly lead to further unification.

General systems theory is, as emphasized, a model of certain general aspects of reality. But it is also a way of seeing things which were previously overlooked or bypassed, and in this sense is a methodological maxim. And like every scientific theory of broader compass, it is connected with, and tries to give its answer to perennial problems of philosophy.

PROBLEMS

1.1. What is internal and external system description?

1.2. Enumerate the main steps in the history of the systems idea.

1.3. What do you understand by "classical" system theory?

1.4. Define "system."

1.5. Define the concept and function of "model" in science.

1.6. Define the concept of "emergence" in terms of systems theory.

1.7. Discuss mechanism and vitalism in biology, and indicate similar conceptions in other sciences.

1.8. Compare gestalt theory, structuralism, and functionalism.

1.9. What is the relation between cybernetics and systems theory?

1.10. Define real and conceptual systems, and discuss their relationship.

1.11. Indicate the relations between the systems idea and dialectical materialism.

1.12. What is the importance of "open system" in biology and in general systems theory?

REFERENCES

1. **Ashby, W. R.,** "Effect of Controls on Stability." *Nature* (London), Vol. 155, No. 3933, pp. 242–243, February 1945.

2. **Ashby, W. R.,** *An Introduction to Cybernetics*, 3rd ed. New York: Wiley, 1958.

3. **Beier, W.,** *Biophysik*, 3rd ed. Jena: Fischer, 1968. English translation in preparation.

4. **Beier, W.,** and **W. Laue,** "On the Mathematical Formulation of Open Systems and Their Steady States." In *Unity Through Diversity. Festschrift in Honor of Ludwig von Bertalanffy*, W. Gray and N. Rizzo (Editors), Book II: *General and Open Systems.* New York: Gordon and Breach, 1971.

5. **Bertalanffy, L. von,** *Nikolaus von Kues.* München: G. Müller, 1928.

6. **Bertalanffy, L. von,** *Theoretische Biologie*, Vols. I, II. Berlin: Borntraeger, 1932, 1942; 2nd ed., Bern: Francke, 1951.

7. **Bertalanffy, L. von,** *Modern Theories of Development.* Translated by J. H. Woodger. Oxford: Oxford University Press, 1934; New York: Harper Torchbooks, 1962. German original: *Kritische Theorie der Formbildung.* Berlin: Borntraeger, 1928.

8. **Bertalanffy, L. von,** "Der Organimus als physikalisches System betrachtet." *Die Naturwissenschaften*, Vol. 28, pp. 521–531, 1940. Reprinted in *General System Theory. Foundations, Development, Applications.* New York: George Braziller, 1968.

9. **Bertalanffy, L. von,** "Zu einer allgemeinen Systemlehre." *Blätter für Deutsche Philosophie* Vol. 18, 1945; extract in *Biologia Generalis*, Vol. 19, pp. 114–129, 1949. Reprinted in *General System Theory. Foundations, Development, Applications.* New York: George Braziller, 1968.

10. Bertalanffy, L. von, *Problems of Life. An Evaluation of Modern Biological Thought.* London: Watts & Co.; New York: Wiley, 1952; Harper Torchbooks, 1960. German original: *Das biologische Weltbild.* Bern: Francke, 1949.

11. Bertalanffy, L. von, "An Outline of General System Theory." *British Journal of the Philosophy of Science*, Vol. 1, pp. 134–164, 1950. Reprinted in *General System Theory. Foundations, Development, Applications.* New York: George Braziller, 1968.

12. Bertalanffy, L. von, *Biophysik des Fliessgleichgewichts.* Translated by W. H. Westphal. Braunschweig: Vieweg, 1953. Revised edition with W. Beier and R. Laue in preparation.

13. Bertalanffy, L. von, "Zur Geschichte theoretischer Modelle in der Biologie." *Studium Generale*, Vol. 8, pp. 290–298, 1965.

14. Bertalanffy, L. von, *Robots, Men and Minds.* New York: George Braziller, 1967.

15. Bertalanffy, L. von, *General System Theory. Foundations, Development, Applications.* New York: George Braziller, 1968.

16. Bertalanffy, L. von, "Chance or Law." In *Beyond Reductionism*, A. Koestler and J. R. Smythies (Editors). New York: Hutchinson, 1969.

17. Bertalanffy, L. von, C. G. Hempel, R. E. Bass, and H. Jonas, "General System Theory: A New Approach to Unity of Science." *Human Biology*, Vol. 23, pp. 302–361, 1951.

18. Boulding, K. E., *The Image.* Ann Arbor: University of Michigan Press, 1965.

19. Buckley, W., *Sociology and Modern Systems Theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1967.

20. Buckley, W. (Editor), *Modern Systems Research for the Behavioral Scientist. A Sourcebook.* Chicago: Aldine Publishing Co., 1968.

21. D'Ancona, V., *Der Kampf ums Dasein.* In *Abhandlungen zur exakten Biologie*, L. von Bertalanffy (Editor). Berlin: Borntraeger, 1939. English translation: *The Struggle for Existence.* Leiden: E. J. Brill, 1954.

22. Demerath, N. J., III, and R. A. Peterson (Editors), *System, Change and Conflict. A Reader on Contemporary Sociological Theory and the Debate over Functionalism.* New York: Free Press, 1967.

23. De-Shalit, A., "Remarks on Nuclear Structure," *Science*, Vol. 153, pp. 1063–1067, 1966.

24. Egler, F. E., "Bertalanffian Organismicism," *Ecology*, Vol. 34, pp. 443–446, 1953.

25. Frank, L. K., G. E. Hutchinson, W. K. Livingstone, W. S. McCulloch, and N. Wiener, "Teleological Mechanisms." *Annals of the New York Academy of Sciences*, Vol. 50, Art. 4, pp. 187–278, October 1948.

26. Gray, W., F. D. Duhl, and N. D. Rizzo (Editors), *General Systems Theory and Psychiatry.* Boston: Little Brown, 1968.

27. Gray, W., and N. Rizzo (Editors), *Unity Through Diversity. Festschrift in Honor of Ludwig von Bertalanffy.* Espec. Book II: *General and Open Systems*, and Book IV: *General Systems in the Behavioral Sciences.* New York: Gordon and Breach, 1971.

28. Hahn, W., *Theory and Application of Liapunov's Direct Method.* Englewood Cliffs, N.J.: Prentice-Hall, 1963.

29. Harvey, D., *Explanation in Geography.* London: Arnold, 1969; New York: St. Martin's Press, 1969.

30. Klir, G. J., *An Approach to General Systems Theory.* New York: Van Nostrand Reinhold, 1969.

31. Koestler, A., "The Tree and the Candle." In *Unity Through Diversity*, W. Gray and N. Rizzo (Editors), Book II: *General and Open Systems.* New York: Gordon and Breach, 1971.

32. **Koestler, A.,** and **J. R. Smythies** (Editors), *Beyond Reductionism.* New York: Hutchinson, 1969.

33. **Köhler, W.,** "Zum Problem der Regulation." *Roux's Archiv.*, Vol. 112, 1927.

34. **Kraft, V.,** *Die Grundlagen der Erkenntnis und der Moral.* Berlin: Duncker and Humblot, 1968.

35. **Kuhn, T.,** *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1962.

36. **La-Salle, J.,** and **S. Lefschetz,** *Stability by Liapunov's Direct Method.* New York: Academic Press, 1961.

37. **Laszlo, E.,** "Systems and Structures—Toward Bio-social Anthropology." In *Unity Through Diversity,* W. Gray and N. Rizzo (Editors), Book IV: *General Systems in the Behavioral Sciences.* New York: Gordon and Breach, 1971.

38. **Laszlo, E.,** *Introduction to Systems Philosophy.* New York: Gordon and Breach, 1971.

39. **Lotka, A. J.,** *Elements of Physical Biology* (1925). New York: Dover, 1956.

40. **Maccia, E. S.,** and **G. S. Maccia,** *Development of Educational Theory Derived from Three Educational Theory Models,* Project 5–0638. Columbus, Ohio: The Ohio State Research Foundation, 1966.

41. **Mesarovic, M. D.,** "Foundations for a General Systems Theory." In *Views of General Systems Theory,* M. D. Mesarović (Editor), pp. 1–24. New York: Wiley, 1964.

42. **Miller, J. G.,** "Living Systems: Basic Concepts." In *General Systems Theory and Psychiatry,* W. Gray, F. D. Duhl, and N. D. Rizzo (Editors). Boston: Little, Brown, 1968.

43. **Needham, J.,** "Review of 'Theoretische Biologie,' Vol. I, by L. von Bertalanffy." *Nature* (London), Vol. 132, 1933.

44. **Rescigno, A.,** and **G. Segre,** *Drug and Tracer Kinetics.* Waltham, Mass.: Blaisdell, 1966.

45. **Rosen, R.,** "Two-Factor Models, Neural Nets and Biochemical Automata." *Journal of Theoretical Biology,* Vol. 15, pp. 282–297, 1967.

46. **Rosen, R.,** *Dynamical System Theory in Biology,* Vol. I: *Stability Theory and Its Applications.* New York: Wiley, 1970.

47. **Rosen, R.,** "A Survey of Dynamical Descriptions of System Activity." In *Unity Through Diversity,* W. Gray and N. Rizzo (Editors), Book II: *General and Open Systems.* New York: Gordon and Breach, 1971.

48. **Schwarz, H.,** *Einführung in die moderne Systemtheorie.* Braunschweig: Vieweg, 1969.

49. **Sorokin, P. A.,** *Contemporary Sociological Theories* (1928). New York: Harper Torchbooks, 1964.

50. **Sorokin, P. A.,** *Sociological Theories of Today.* New York: Harper & Row, 1966.

51. **Weaver, W.,** "Science and Complexity." *American Scientist,* Vol. 36, pp. 536–544, 1948.

52. **Weiss, P. A.,** "Life, Order and Understanding." *The Graduate Journal,* Vol. III, Supplement, University of Texas, 1970.

53. **Whyte, L. L., A. G. Wilson,** and **D. Wilson** (Editors), *Hierarchical Structures.* New York: American Elsevier, 1969.

54. **Wiener, N.,** *Cybernetics,* New York: Wiley, 1948.

55. **Wilbert, H.,** "Feind-Beute-Systeme in kybernetischer Sicht." *Oecologia* (Berlin), Vol. 5, pp. 347–373, 1970.

56. **Woodger, J. H.,** *Biological Principles.* London: Routledge and Kegan Paul, 1929; New York: Humanities Press, 1967.

2. The Uses of Mathematical Isomorphism in General System Theory

ANATOL RAPOPORT

Department of Psychology, University of Toronto

EDITOR'S COMMENTS

Although Chapter 2 is not heavily dependent on mathematics, the reader is expected to be familiar at least with the basic concepts and principles of differential equations, graph theory, and the theory of games. The following books are recommended as prerequisite readings in these areas:

Ziebur, A. D., *Topics in Differential Equations*. Belmont, Calif.: Dickenson Publishing Co., 1970.
Ore, O., *Graphs and Their Uses*. New York: Random House, 1963.
Berge, S., *The Theory of Graphs and Its Applications*. New York: Wiley, 1962.
Owen, G., *Finite Mathematics*. Philadelphia: W. B. Saunders, 1970. (Chapter VI, "The Theory of Games," pp. 267–328.)

2.1. INTRODUCTION

The dichotomies analytic-synthetic, atomistic-holistic, local-global, and differential-integral all have related meanings. The first term of each dichotomy refers to detail, analysis, investigation of parts or of local conditions; the second term, to the perception of wholes, *gestalten*, and configurations. In the context of mathematics, the nature of the dichotomy is clearest. Thus the fundamental concept of the differential calculus is the *derivative*, defined at a point; the integral calculus introduces the concept of the *integral*, essentially a summation. The differential calculus is analytic; the integral calculus, synthetic.

Classical science branched off from philosophy by adopting primarily analytic methods of investigation. Galilean mechanics put *instantaneous* velocity at the center of attention, and Newtonian mechanics began with associating an instantaneous force with the derivative of instantaneous velocity. This formulation deliberately turned away from earlier holistic or teleological conceptions of motion, according to which motion was thought to be determined by the "nature" of the moving body or by the tendency of bodies to achieve a "natural" final position. For instance, ancient astronomy assumed circular motion of planets as a reflection of their perfect or divine nature. Medieval physics explained the falling of stones by their "striving" to get as close as possible to their "natural" position at the center of the earth, and the rise of smokes by a similar striving to reach their proper abode in the realm of fire beyond the stars. In contrast, Galilean and Newtonian mechanics explained motion by analyzing forces acting at particular points at particular times.

The analytical point of view spread from the physical to the biological sciences. In physiology, explanations came to be formulated in terms of "efficient" rather than "final" (teleological) causes, to use the traditional Aristotelian distinction. That is to say, the question "Why does this organ act as it does?" came to mean a demand for an answer in terms of chains of physicochemical events rather than in terms of the "purpose" served by the organ.

In short, scientific investigation became sharper, narrower, more "local," concentrated more on specific events than on global, holistic, all-encompassing images of "meaning" and "purpose." With increased concentration came the fragmentation of science into ever-multiplying disciplines, each with its own inbred vocabulary, methods, and foci of interest. A reaction against this trend was already discernible in the warning uttered by Alfred North Whitehead: Science, he wrote, is living off the capital accumulated in the seventeenth century, and the capital is nearing extinction.

It was, however, easier to point to the danger than to indicate a way out of the impasse. A return to the language and concepts of prescientific natural philosophy would be futile. It seems to be the fate of philosophy to become obsolete as soon as science steps in to formulate speculations as concrete hypotheses subject to rigorous methods of investigation. If synthetic, holistic ideas were to be reintroduced into science, they would have to be couched in a language as disciplined as that of analytic science. Certain formulations of general systems theory are attempts to achieve this goal. The present chapter is concerned with these trends.

2.2. THEORY IN THE STRICT AND IN THE BROAD SENSE

Strictly speaking, general systems theory is not a "theory" in the sense that most scientific theories are theories. A *scientific theory* embodies a certain content area, a class of events, say, the motions of heavenly bodies or the phenomena associated with the propagation of light or sound, or those manifested in the working of a nervous system or the fluctuations of market prices. A scientific theory seeks to establish connections among events, ideally to discover necessary and sufficient conditions for the occurrence of a class of events, expressed in the paradigm of a scientific assertion: "if so, . . . then so." In the most mature sciences, this paradigm is mathematicized. The connections are stated in mathematical equations, relating variable quantities to each other. In an equation, the "if" part is the assignment of values to independent variables; the "then" part is the resulting assignment of corresponding values to dependent variables. Each assignment is a particular "if so, . . . then so" assertion. Hence, the equation represents potentially an infinity of such assertions.

A theory can be understood, however, in a sense broader than that described above. Before a theory is in a position to make "if so, . . . then so" assertions, the variables about which the assertions are to be made must be sought out and defined. The search for the variables and for appropriate definitions can also be regarded as a theoretical investigation. In this broader sense, a theory may be no more than a scheme of classifications and definitions. (A definition is, of course, prerequisite to classification). In this broader sense, general systems theory can, perhaps, be considered a theory, since it does undertake the task, first of all, of defining "a system" and then of classifying systems according to certain criteria.

It stands to reason that all definitions and classifications are, a priori, arbitrary. That is to say, a definition is not a proposition to which a truth value can be assigned. A definition is no more and no less than an agreement to use a word or a phrase in a certain way. The only criterion of adequacy

applicable to a definition concerns the possibility of deciding whether something does or does not belong to the class defined. The same applies to a classification or a taxonomy which is a consequence of definitions.

Nevertheless, if we insist that a taxonomy is not the goal of a theory but is only a prerequisite for pursuing a theoretical goal, that is, if we keep the narrower meaning of a theory in mind (that of generating " if so, . . . then so " statements), then we cannot view definitions and classifications as arbitrary. Good classifications are those that are likely to produce concepts from which a far-reaching theory (in the strict sense of the word) can be constructed. A poor classification is one that has no such far-reaching implications.

For example, there is no a priori reason why bodies should not be classified by their shapes or ordered in magnitude by their volumes (extension in space), as was done by Descartes in developing his theory of motion. But from the point of view of a theory of motion, Descartes' classification scheme proved to be sterile, whereas a magnitude ordering of bodies according to their masses proved immensely fruitful.

Again, there is no a priori reason against classifying animals into large and small, dangerous and harmless, or edible and inedible. Indeed, such classifications have served specific practical purposes. None of them, however, has led to a theory of evolution, whereas the taxonomies undertaken by biologists, based on more fundamental features (which they somehow perceived to be relevant), had this outcome. That the theory of evolution is a theory in the strict sense, involving predictive assertions, should not be obscured by the fact that it concerns predominantly past events. It none the less involves predictions, for instance, to the effect that fossils of animals of a certain type will be found in rocks of a certain age. And, of course, genetics, which is tightly interlaced with the theory of evolution, contains assertions involving (statistical) predictions.

2.3. THE TASK OF GENERAL SYSTEMS THEORY

The task of general systems theory can be formulated as follows: to prepare definitions and hence classifications of systems that are likely to generate fruitful theories in the narrow sense. There is a connection between this task and the one mentioned earlier—that of counteracting the increasing fractionation of science. The connection will be seen in the definition of a system on which most system theorists appear to agree.

The idea underlying all definitions of a system is that of a collection of entities and sets of relations among them. Superimposed on this idea are additional criteria, stated with varying degrees of precision (or vagueness), singled out by people with varying interests. At times, these criteria are

suggested not so much by logical considerations as by the kind of thing that has already been taken as a prototype of a system. A living organism serves understandably as a prototype of this sort for the biologist, a machine for the engineer, an automaton or a computer for a cybernetician. The concept of system becomes somewhat vague when exemplified by a "philosophical system" or a "system of beliefs." Indeed, the question is sometimes raised whether entities that have no material existence can be properly called systems. Some of these entities, however, have such clearly discernible structures (well-defined relations among the elements that compose them) that it seems most natural to extend the concept of system to them. Languages and mathematical systems are obvious examples.

2.4. MATHEMATICAL SYSTEMS AND ISOMORPHISM

A *mathematical system* consists of a set of elements (e.g., numbers, points, vectors, matrices, etc.) and the precisely specified relations among them. In fact, the elements of a mathematical system are defined exclusively in terms of the specified relations among them. It is important to recognize that mathematical operations can also be defined in terms of relations, so that the concept of "operation" is logically redundant. It is introduced in the definition of a mathematical system in the interests of more vivid conceptualization or as a consequence of terminological inertia. For instance, the operation of addition can be represented as a ternary relation: an ordered triple of elements can be said to be in the given relation if and only if the sum of the first two equals the third. In symbols:

$$R(a, b, c) \leftrightarrow a + b = c. \tag{2.1}$$

A mathematical system is contentless. For instance, the set of positive integers closed under the operation of addition applies equally well to camels, to oranges, and to years.

Two mathematical systems are said to be *isomorphic* to each other if a one-to-one correspondence can be established between the elements of one and those of the other and if all the relations defined on the elements of one hold also among the corresponding elements of the other. Isomorphism between two mathematical systems induces a conceptual isomorphism between the concrete systems they represent. In other words, two concrete systems can be said to be conceptually isomorphic to each other if both can be represented by the same mathematical model.

The concept of isomorphism leads to a classification of all systems that can be represented by mathematical models. The classification of such systems becomes simply an image of a classification of mathematical models

representing them. The logical advantages of such a classification are at once apparent. For instance, one mathematical system can be seen immediately as a generalization of another, that is, as including the latter as a special case. The induced classification of corresponding concrete systems immediately displays one class as including the other. If the systems are represented by isomorphic mathematical systems (or by the same model), all the theorems of the mathematical system are applicable to all consequences derived from the definition of the concrete systems. It follows that the concept of mathematical isomorphism is a powerful tool for integrating theories of concrete systems. In this way, the integration goal of general systems theory is served.

A frequently cited example of isomorphism between concrete systems is that between a harmonic oscillator and an electrical circuit containing an inductance, a resistance, and a capacitance in series. The mechanical and electrical systems are illustrated in Figure 2.1.
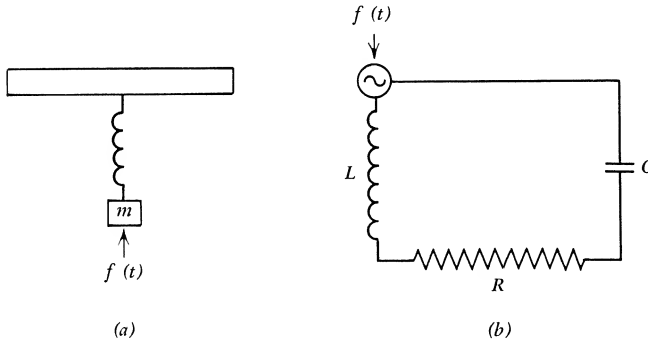
Figure 2.1. (*a*) A mechanical harmonic oscillator. (*b*) An electric circuit with inductance, resistance, and capacitance in series.

The position of the mass suspended by a spring is given as a function $x(t)$ of time by the general solution of the second-order differential equation:

$$m\ddot{x} + r\dot{x} + cx = f(t), \tag{2.2}$$

where m is the mass of the body, r is a constant denoting the coefficient of friction in the system, c is the elasticity of the spring, and $f(t)$ represents an externally impressed, vertically acting force upon the body as a function of time.

The quantity of charge at any point in the circuit is given as a function of t by $q(t)$, which is a solution of the differential equation

$$L\ddot{q} + R\dot{q} + C^{-1}q = f(t), \tag{2.3}$$

where L is the inductance of the circuit, R the resistance, C the capacitance, and $f(t)$ an externally imposed electromotive force. Clearly, the mathematical models are isomorphic under the correspondence

$$x \leftrightarrow q, \qquad m \leftrightarrow L, \qquad r \leftrightarrow R, \qquad c \leftrightarrow C^{-1}. \tag{2.4}$$

The correspondence suggests analogies between position and charge, between mass and inductance, between friction and electrical resistance, and between elasticity and capacitance. Although the analogy between friction and resistance is intuitively apparent, the other analogies may not be. However, their role as corresponding elements in a mathematical isomorphism *suggests* the analogies and along with them corresponding conceptualizations. Indeed, it is easy for an electrical engineer to think of the inverse of capacitance as an elastic spring that offers a counterforce in proportion to its compression, interpreted as the accumulation of charge, hence of a potential across the capacitor.

Actually, all theoretical thinking is analogical. What distinguishes a possibly justified analogy from a fortuitous one is the relatively far-reaching or superficial nature of the relations that suggest the analogy. Metaphors reflect analogies: what the king is to his realm, a head is to its body—hence the king is "head of state." This analogy is superficial, and pressing its implications is not likely to produce a fruitful political theory. But the analogy "Inductance is to charge what mass is to position," although not nearly so immediately comprehensible as the analogies expressed in every-day language, is incomparably more to the point.

The language of science is replete with metaphorical extensions of meanings suggested by mathematical isomorphisms. One speaks of the flux of heat and of magnetic flux, although "flowing" is far-fetched in the first instance and even more so in the second. These metaphors, however, reflect a basic *mathematical* connection among the various meanings of "flux."

A remarkable metaphor was coined by Mandelbrot [1], namely, the "temperature" of a language corpus. It has, of course, nothing to do with the emotional content of the corpus, or with any physical aspect. Mandelbrot drew on the analogy between statistical and physical entropy, and so was led to another analogy between physical temperature and an analogous mathematical quantity describing a language corpus.

The connection between information and entropy was noted earlier in statistical mechanics by Norbert Wiener. Whether the connection is merely a by-product of mathematical formalism or is a reflection of "physical reality" has been the subject of some controversy. From the point of view of general systems theory, the controversy is vacuous. If we assume that the subject of discourse is never "physical reality" per se but rather always our *knowledge* or conception of physical reality, we must conclude that any structural des-

cription of physical reality is the only content of knowledge; hence that any structural description that reaches " all the way " to elements and relations not further analyzable may be identified with "reality" itself. This conception actually implies a definition of physical reality.

One could argue that inductance is not physically identical with mass despite their structurally analogous roles in electrical and mechanical systems. However, the argument can be supported only by *further* analysis of inductance and of mass, whereby the distinction between them can presumably be revealed. That is to say, "inductance" and "mass" are not identical *if* further analysis reveals *structural* differences between them. The philosophical view which declares knowledge to be knowledge of structure underlies the trend in general systems theory that puts mathematical isomorphism at the foundation.

2.5. STATIC AND DYNAMIC THEORIES

The mathematical definition of a system is actually a definition of a mathematical model assumed to be isomorphic to a system abstracted from the one described. A system consists of elements and relations among them that are singled out for attention. In particular, the elements may be variables in terms of which the behavior of the elements is to be described. The structure of the system is described as a set of relations among the elements (or the variables).

If the elements are variables, one defines the state of the system as the totality of values assumed by the variables at some moment of time. A *static theory* assumes that the state remains constant. Then the equations defining the relations among the variables permit us to deduce the values of some of the variables when the values of others have been given. A *dynamic theory* considers primarily a succession of states. Furthermore, the system is *deterministic* if knowledge of the values of the relevant variables at a moment of time permits us to deduce the state of the system at any future or any preceding moment. The system is *probabilistic* or *stochastic* if knowledge of the values of the variables at a given moment permits us to predict only probability distributions of these variables at some future moment (or to postdict them for some past moment). Mathematically speaking, probabilistic systems can be redefined as deterministic ones if distribution functions instead of numerical values of variables are taken to be the elements, as, for example, the wave functions of quantum mechanics.

As a simplest example of a static system, consider a gas enclosed in a volume in a state of equilibrium. The variables of interest are pressure (constant throughout the volume, since the gas is in equilibrium), temperature (constant

for the same reason), and volume. The three variables are connected by the equation of state. Thus knowledge of the values of any two of the variables permits us to deduce the value of the third.

A planetary system is an example of a dynamic system. Here the variables of interest are the instantaneous positions and velocities of the planets. The relative positions determine the gravitational forces, and these, given the masses as parameters, determine the accelerations. Forces, masses, and accelerations are related through the differential equations governing the behavior of the system. If the initial state of the system defined in this manner is known, successive states can, in principle, be calculated.

The same conception of system applies to the automaton. An automaton is a system whose behavior at any moment is defined by the totality of values assumed by a set of variables, some of which are associated with the internal state of the automaton and others with inputs presented to it. At specified moments the automaton receives an input from some source. The input, like the internal state, is specified by the values of a set of variables. The input, together with the internal state of the automaton, determines an output and a new state. It should be noted that the " state " of the automaton may involve in its description all previous states, as is the case in an automaton with a " memory." A digital computer is a well-known example of an automaton. The rules governing the changes of state as they depend on both the state and the input constitute a *program*. A program, together with a set of inputs, determines a set of outputs.

2.6. SYSTEMS REPRESENTED BY DIFFERENTIAL EQUATIONS

As has been said, the mathematical models describing various systems suggest a *classification* of the systems. We shall first discuss some classes of dynamic systems.

Consider the class of systems isomorphic to systems of ordinary differential equations. The latter can be classified by orders, degrees, numbers of dependent variables, and so forth. The simplest such systems are linear ones with constant coefficients. The prototype is

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{ij} x_j + b_i \qquad (i = 1, 2, \ldots, n; a_{ij}, b_i \text{ constants}). \qquad (2.5)$$

A system of this sort can be equivalently represented by a single nth-order differential equation. The general solution of all such systems is known. It has certain properties that make systems represented by linear differential equations with constant coefficients especially accessible to intuitive understanding. The deterministic character of such systems is also evident. Given

a set of initial conditions, that is, initial values of the dependent variables, $x_1, x_2, \ldots, x_n$, the "world line" of the vector $(x_1, x_2, \ldots, x_n)$ is thereby determined, indeed unambiguously, since no two world lines intersect. Next, the world lines of the components of this vector can be only of certain simple types. As a function of time, such a world line is represented by a sum of exponential functions, where the exponents are, in general, complex. These complex numbers are roots (called eigenvalues) of a certain polynomial determined by the constants of the system. The real parts of these roots constitute the exponential components of the solutions: they indicate either an unlimited growth or a steady decline in the magnitude of the variables. The complex parts are the periodic components: they indicate oscillations of the magnitudes. Oscillations with constant amplitudes can occur only in very special cases, namely, when all the roots of the polynomial are pure imaginary numbers. Note that all of these findings are results of mathematical deduction *only*, not empirical generalizations.

With these facts in mind, we turn to the classification of systems induced by the classification of sets of differential equations. We note first that no planetary system or, in general, any mechanical system in which forces are nonlinear functions of distances (such as gravitational forces) can be represented by a linear system of ordinary differential equations. What kind of systems, then, can be represented by sets of linear ordinary differential equations? Obviously, systems in which forces or their analogs are linear functions of positions. The harmonic oscillator described above is a system of this sort. It is characterized by the fact that the restoring force of a spring is proportional to the displacement of the mass from its equilibrium position. Without an impressed external force, the equation of such a system becomes

$$\ddot{x} + r\dot{x} + cx = 0, \tag{2.6}$$

where, since the unit of mass is arbitrary, we have set $m = 1$. By the nature of "friction" the parameter r must be nonnegative, and by the nature of "elasticity" the parameter c is positive. The formal solution of this equation is

$$x = Ae^{at} + Be^{bt}, \tag{2.7}$$

where A and B are fixed by the initial conditions, while a and b are the roots of $y^2 + ry + c = 0$. From Eq. 2.6 we see that the types of motion of a particle representing a harmonic oscillator are rather strictly limited. The oscillation is essentially a damped one except in the very special case where $r = 0$. Since the derivative of both the exponential function and the sinusoidal function essentially reproduces each of these functions, it follows that the directed velocity of the particle (or, in the case of the electric circuit, the current) follows the same type of world line as the position (or the charge).

Next, we note that the second-order differential equation, Eq. 2.6, is equivalent to a system of two first-order differential equations, namely,

$$\frac{dx}{dt} = y, \qquad \frac{dy}{dt} = -ry - cx. \tag{2.8}$$

We conclude that the solutions of this system, namely, the functions $x(t)$, $y(t)$ that satisfy it, must also be damped oscillations (so long as $r \neq 0$), or undamped periodic functions in the special case where $r = 0$.

Now, the system of the general type represented by Eq. 2.5 can be pictured as a model of a situation quite different physically from either a harmonic oscillator or an electrical circuit. To see this, let us generalize the second-order system slightly to obtain a homogeneous system of two first-order linear differential equations;

$$\begin{aligned} \dot{x} &= ax + by, \\ \dot{y} &= cx + dy, \end{aligned} \tag{2.9}$$

with $ad - bc \neq 0$.

In words, the rate of change of each variable is influenced in an additive way by two functions, each proportional, respectively, to the magnitude of each variable. Suppose that x and y are concentrations of a substance in a system of chemical reactions. The rate of change of concentration can be either enhanced or inhibited by the concentrations themselves. For instance, if in the system of Eq. 2.9 $a > 0$, x is an autocatalytic agent, while $a < 0$ reflects the assumption that x disintegrates as a rate proportional to its own concentration. Similarly, if $b > 0$, y acts as a catalytic agent for the formation of x; if $y < 0$, y inhibits the formation of x. Similar considerations apply to c and d.

The general solution of the system in Eq. 2.9 is

$$\begin{aligned} x &= Ae^{r_1 t} + Be^{r_2 t}, \\ y &= \left(\frac{r_2}{b} - a\right)Ae^{r_1 t} + \left(\frac{r_2}{b} - a\right)Be^{r_2 t}, \end{aligned} \tag{2.10}$$

where A and B are determined by the initial conditions, and r_1, r_2 are the roots of the characteristic polynomial.

From this solution we can read off directly the behavior of the hypothetical chemical system represented by the model postulated. In particular, we can conclude that the concentrations of x and y will *not* undergo undamped oscillations except in the very special case when $a = -d$ and $bc < a^2$.

Moreover, if this special case does not obtain, each of the concentrations must approach either zero or infinity. The latter conclusion is physically

absurd. It is a consequence of our having failed to introduce a constraint in the form of the conservation of mass, namely, $x + y = m$, a constant. Furthermore, because mass cannot be negative, we must have $x \geq 0$, $y \geq 0$. Then the system will be represented by a single equation:

$$\dot{x} + ax + b(m - x) = (a - b)x + g, \qquad (2.11)$$

where $g = bm$, $x \geq 0$.

The derivative vanishes when $x = g/(b - a)$. If $0 < g/(b - a) < m$, $ab < 0$, the system approaches an equilibrium where $x/y = -b/a$. Otherwise one or the other of the substances must eventually disappear. There will be no oscillations.

It is important to note that in this case the sum of the concentrations is restrained by the conservation of mass to remain constant. The situation would be different if our system of differential equations were nonhomogeneous. Consider the system

$$\frac{dx}{dt} = ax + by + h, \qquad \frac{dy}{dt} = cx + dy + k. \qquad (2.12)$$

The equilibrium state of such a system, if it exists, is obtained by solving simultaneously the two linear equations obtained by setting dx/dt and dy/dt equal to zero. If $ad - bc$, the determinant of the system, does not vanish, a unique equilibrium is determined, in which the concentrations of the substances do *not* depend on the initial concentrations. This sort of equilibrium, called a steady state, is established with the *environment* of the system. In this case, the system is open. It has sources and/or sinks represented by the constant terms.

This purely mathematical result has a bearing on the observation made many years ago by von Bertalanffy [2] that so-called open systems, in contrast to closed systems, exhibit a *principle of equifinality*, that is, a tendency to achieve a final state relatively independently of initial conditions. In other words, open systems may tend to "resist" perturbations that take them away from some steady state, that is, may tend to return to the steady state. They can exhibit *homeostasis*.

As illustrations, von Bertalanffy cited the growth behavior of certain animals. When their growth is interrupted, say, by diet deficiencies, they can often "catch up" when the deficiencies are remedied, as if they strived to achieve their predestined size regardless of outside intervention. The connection between these phenomena and the characteristics of "open systems" (as they are called in physics) is admittedly far-fetched, and von Bertalanffy's conclusions have a distinct teleological flavor. On the other hand, the mathematical example we have given removes all ambiguity (admittedly in a very

special case). Here an "open system" is identified with a system of differential equations that is nonhomogeneous; a closed system, with one that is homogeneous with a constraint representing the conservation of mass. Equifinality is a mathematical consequence in the former but not in the latter.

We note in passing that, if the conservation of mass equation is not imposed on a homogeneous system with a nonvanishing determinant $(bc - ad)$, the only equilibrium is the trivial solution $x = 0$; $y = 0$. On the other hand, if the determinant does vanish, there is an infinity of "equilibria." The model is immediately generalizable to any number of variables.

From a still more general point of view, one can examine sets of differential equations for mathematical properties of their solutions and interpret these properties in terms of concrete systems that are reasonable realizations of the mathematical systems. In particular, since the general solution of any set of linear differential equations is known, we can immediately deduce certain interesting properties of any system that can be so represented. One such property is that a system represented by a set of linear differential equations having constant coefficients with a nonvanishing determinant has at most one equilibrium state. This is not, in general, true for nonlinear systems. Furthermore, the stability or instability of such equilibria is immediately deducible from relations among the coefficients. Stability is the tendency of a system to return to its equilibrium; instability is a tendency to move away from equilibrium if slightly disturbed.

As already noted, widely different interpretations can be proposed for a given set of differential equations. An interesting application of the linear system model was made by Richardson [3] in his theory of arms races. Assuming that the level of armaments of a rival power provides the stimulus for increasing one's own level, and that one's own level of armaments serves as an inhibitory factor on further increases, Richardson represented the situation by a pair of differential equations isomorphic to the systems of Eq. 2.12. Here x and y can take on both positive and negative values. If positive, they represent the respective degrees of "hostility" of two powers vis-à-vis each other, reflected, say, in the armament budgets or some similar index; if negative, they represent the level of good will or cooperation, reflected, say, in trade volumes.

It turns out that the system is stable, that is, will tend toward an equilibrium, if $ad < bc$; otherwise the system is unstable. In the latter case, depending on the initial conditions (say, armament budgets and trade volumes), the system will reflect either a runaway armament race or, on the contrary, disarmament and ever-increasing cooperation.

The next class of sets of differential equations deserving attention comprises differential equations of the second degree. Here the variables on the right

appear either as squared or multiplied pairwise. Several situations might be represented by mathematical systems of this sort; the following are examples.

1. A system of bimolecular chemical reactions, where the rate of change of each substance is enhanced or inhibited by the frequency of collisions of pairs of molecules.

2. A system of social interactions, where the variables represent proportions of individuals, say, holding particular views or exhibiting particular forms of behavior. The views of an individual may change as he encounters other individuals holding similar or different views. Therefore the rates of change of the "concentrations" of the several views will be proportional to the rates of such encounters, hence to pairwise products of the "concentrations" of the various types of individuals.

3. Ecological systems containing predators and prey. As a simple example of such a system, consider the following pair of differential equations of the second degree:

$$\frac{dx}{dt} = -ax^2 + bxy,$$

$$\frac{dy}{dt} = -bxy - cy^2 + my \qquad (a, b, c, m > 0). \tag{2.13}$$

The biomass of the predators is represented by x; that of their prey, by y. The predators can increase only through encounters with prey, as represented by bxy. They must, of course, also meet with members of the opposite sex of the same species in order to reproduce. However, the net effect of encounters between members of the same species (among the predators) is assumed to be negative. This effect is attributed to the "crowdedness factor."

The biomass of the prey, on the other hand, increases in the absence of predators at a rate proportional to itself as represented by my. Moreover, encounters with predators result in a decrease of the biomass of the prey (since they are eaten). Hence the coefficient of the xy term, representing the frequency of predator-prey encounters, is negative in the second equation. The crowdedness factor is represented by $-cy^2$.

Such a system has equilibria at $x = y = 0$ and at $x = mb/(b^2 + ac)$, $y = ma/(b^2 + ac)$. The first equilibrium is trivial; the second reveals the dependence of the equilibrium on the birth rates, the frequency of encounters, and the crowdedness factor.

The very simplest nonlinear system of interest is the elementary contagion equation in one variable:

$$\frac{dx}{dt} = ax(b - x). \tag{2.14}$$

Here x is the fraction of individuals infected in an epidemic, b is the fraction of susceptible individuals, and a is the constant of proportionality. The rate of new infections is assumed to be proportional to the frequency of encounters between infected and uninfected individuals, hence to the product of their concentrations, $x(b - x)$.

The solution of this equation is the well-known logistic curve, which represents the growth of the fraction of infected individuals in the population under the assumptions stated.

2.7. GENERAL SYSTEMS THEORY AND MATHEMATICAL MODELS

At this point it may appear to the reader that the particular trend of general systems theory described here is simply that of constructing mathematical models for a variety of phenomena. There is a difference, however, between mathematical model building and the general systems approach. The model builder attempts to construct an adequate mathematical description of a given phenomenon. To develop a mathematical theory of contagion, the model builder might begin with the simplest model, the logistic equation derived from the simplest postulates of random contacts between the infected and the uninfected. Since few actually observed contagion processes will be adequately described or predicted by a model so simple, the theoretician might try to modify it so as to take more factors into account. For instance, he might make the parameter representing the probability of contagion upon encounters an explicit function of time, presumably to take account of the waxing or waning "virulence" of the epidemic. Or he might make this parameter a function of the time elapsed since infection. Furthermore, he might introduce a rate of removal of individuals from the population, due either to death or to recovery with immunity. The whole thrust of the effort would be to improve the accuracy of the description of the phenomenon under study.

In contrast, the mathematical trend in general systems theory makes the mathematical system, rather than a concrete system, the point of departure. Thus the theoretical investigation is an investigation of the mathematical properties of an abstract system of relations. These properties being common to all mathematical systems isomorphic to the given one, the conclusions drawn are expected to apply to all concrete systems of which the mathematical systems are adequate representations. This program naturally guides the

system theorist along the lines of least mathematical resistance, and the simpler mathematical systems are investigated first. Attempts are made to fit concrete systems into the mold of the now-familiar mathematical system. Inevitably, the perceived success of the attempt is colored by the all-too-human desire to see one's enterprise succeed. Nevertheless, the failures are often too conspicuous to be ignored. Here the advantages of the mathematical approach become evident. The reasons for the failure can sometimes be inferred from explicit inadequacies of the postulates that characterize the type of mathematical system, and directions for generalization are indicated. For instance, the failure of linear systems of differential equations to describe adequately certain systems of interactions can be traced to the absence of multiple equilibria in linear systems, suggesting the investigation of the next level of complexity, systems of the second degree. The refusal of the general second-degree system to yield to general methods of solution suggests the investigation of special cases, bringing to attention systems of interactions where such special cases may be relevant. In particular, the system of second-degree differential equations becomes manageable in the special case in which the dependent variables in the derivative is a factor of all the second-degree terms on the right:

$$\frac{dx_i}{dt} = x_i \sum_{j=1}^{n} a_j x_j + b_i \qquad (i = 1, 2, \ldots, n). \tag{2.15}$$

The assumption is reasonable in an ecological setting, for it asserts that the rate of increase of a particular population depends (aside from a source or sink) only on the frequencies of encounter of *that* population with members of other populations.

At this point the difference between the general systems approach and mathematical model building can be pointed out. A particular mathematical assumption suggests to the general systems theorist contexts far removed from the particular one examined. For example, the assumption just mentioned (that the rate of change in a population is independent of encounters of pairs of species that do not include it) suggests a connection with an analogous assumption often made in stochastic learning theory, namely, that the change in the probability of a particular response alternative is independent of the relative frequencies of *other* response alternatives. And this, in turn, suggests possible connections with an assumption often made in decision theory, namely, that the preference order among alternative courses of action is not affected by the introduction or the deletion of outcomes not preferred to any of the outcomes associated with the alternatives in question. In short, the logical structure of the assumptions, not their particular interpretation, is of central interest to the general system theorist.

2.8. AN EXAMPLE FROM THERMODYNAMICS

At this point a particular concrete context can serve as an "anchorage" for generalizing concepts.

The well-known second law of thermodynamics states that in an isolated system the entropy will tend toward a maximum. For certain types of open systems, this law can be generalized, namely, the *rate of entropy* production tends to a minimum. It can be seen that the isolated system is a special case, in which the rate of entropy production tends to zero.

Let us exemplify an open system by the electric circuit shown in Figure 2.2. At the moment the switch is closed, the dissipated power is given by

$$P_0 = \frac{E_1{}^2}{R_1}. \tag{2.16}$$

As the capacitor becomes charged, the dissipated power tends to

$$P_\infty = \frac{E_1{}^2}{R_1 + R_2}. \tag{2.17}$$

The power dissipated is given as a function of E_2 by

$$P = \frac{(E_1 - E_2)^2}{R_1{}^2} + \frac{E_1{}^2}{R_2}. \tag{2.18}$$

Setting $\partial P/\partial E_2 = 0$, we obtain

$$E_2 = E_1 \frac{R_1}{R_1 + R_2}, \tag{2.19}$$

as expected, so that the final state is characterized by the minimum power production. We have thus exemplified the generalization of the second law
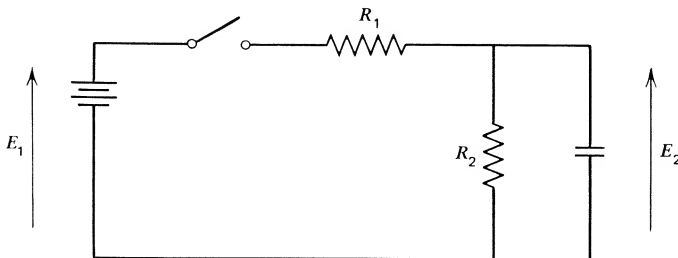
Figure 2.2. Electric circuit, open system.

for an open system. (The battery on the left of Figure 2.2, being an input to the system, is assumed to be replenished.)

Next, suppose that R_2 is replaced by a generator with an internal resistance r_p and a voltage output μE_2, where E_2 is now the "command signal." At $t = 0$, $E_2 = 0$ and $P_0 = E_1{}^2/R_1$ as before. After a sufficient lapse of time, the capacitor may be neglected, so that

$$E_1 - E_2 = iR_1,$$

$$E_2 - \mu E_2 = ir_p,$$

$$E_2 = E_1 \frac{r_p}{r_p + R_1(1 - \mu)}, \qquad (2.20)$$

$$i = E_1 \frac{1 - \mu}{R_1 + r_p/(1 - \mu)},$$

$$P_\infty = i^2 R = \frac{E_1{}^2(R_1 + r_p)}{R_1 + r_p/(1 - \mu)}.$$

It turns out that, if μ lies outside the interval $(-\sqrt{1 + r_p/R_1}, \sqrt{1 + r_p/R_1})$, $P_\infty > P_0$ and the principle of minimizing the rate of entropy production is not satisfied.

The reason for the discrepancy is clear. We have identified entropy with P, the number of joules per second that are turned into heat. But the feedback generator is supplying power to the system. There is thus a negative term in the dissipation of power, namely,

$$P_2 = -\mu E_2 i_2 = -\mu E_2 \frac{E_2 - \mu E_2}{r_p} = -\mu E_2{}^2 \frac{1 - \mu}{r_p}, \qquad (2.21)$$

so that

$$P - P_2 = \frac{(E - E_2)^2}{R_1} + E_2{}^2 \frac{(1 - \mu)^2}{r_p} + \mu E_2{}^2 \frac{1 - \mu}{r_p}. \qquad (2.22)$$

Differentiating partially with respect to E_2, we have

$$\frac{\partial}{\partial E_2}(P - P_2) = 2 \left[\frac{E_2 - E_1}{R_1} + E_2 \frac{(1 - \mu)^2}{r_p} + \mu E_2 \frac{1 - \mu}{r_p} \right]. \qquad (2.23)$$

Thus, if $(P - P_2)$ is to be minimized, we must set

$$\frac{\partial}{\partial E_2}(P - P_2) = 0, \qquad (2.24)$$

obtaining

$$E_2 = \frac{E_1 r_p}{r_p + R_1(1 - \mu)}, \tag{2.25}$$

the value eventually approached.

Now let P_1 be the power supplied to the system from the environment (not from the feedback generator), and let P_2 be the rate at which energy is being stored in the capacitor. Then

$$P_1 = E_1 i_1, \quad P_c = (i_1 + i_2)E_2, \quad P = P_1 + P_2 - P_c, \quad P - P_2 = P_1 - P_c. \tag{2.26}$$

Thus, what is being minimized is the difference between the power supplied to the system from the outside and the power being stored. Collapsing the part of the system including the feedback loop as a "negative resistance," we can say that the principle of minimizing entropy production is satisfied.

Let us now proceed to a further generalization with generators in parallel, as shown in Figure 2.3. We introduce feedback loops so that $E_3 - \mu E_2$, $E_5 = \nu E_4$. As before, we calculate the dissipated power:

$$P = \frac{(E_1 - E_2)^2}{R_1} + \frac{(E_2 - E_3)^2}{R_2} + \frac{(E_2 - E_4)^2}{R_3} + \frac{(E_4 - E_5)^2}{R_4}, \tag{2.27}$$

and the power supplied by the generators:

$$\Delta = E_3 i_3 + E_5 i_5 = E_3 \frac{E_3 - E_2}{R_2} + E_5 \frac{E_5 - E_4}{R_4}. \tag{2.28}$$

Consider the function $\psi = P - \Delta$. If ψ is to be minimized, we must have

$$\frac{\partial \psi}{\partial E_2} = \frac{\partial \psi}{\partial E_4} = 0, \tag{2.29}$$
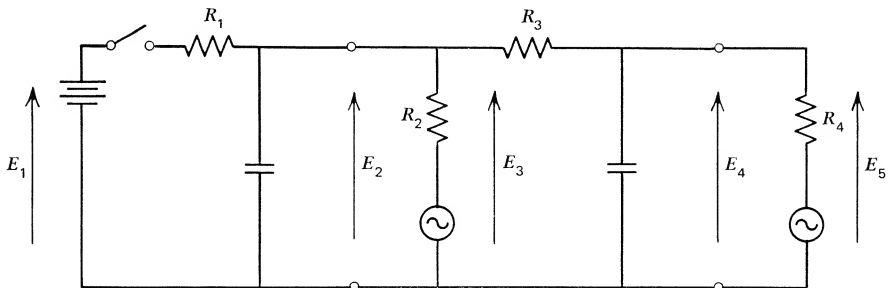
Figure 2.3. Generators in parallel.

which leads to

$$i_2 = i_1 + i_3 = i_5, \tag{2.30}$$

the steady-state conditions. Thus ψ, which plays the role of the rate of entropy production, is minimized in the steady state.

However, if we *cross* the feedback loops so that $E_3 = \mu E_4$ and $E_5 = \nu E_2$, and if we set $\partial \psi / \partial E_4 = 0$, we obtain

$$-2i_2 - 2i_5 - \mu \frac{E_2}{R_2} + \nu \frac{E_5}{R_4} = 0. \tag{2.31}$$

But $i_2 + i_5 = 0$ must hold in the steady state. Therefore we must have

$$\frac{-\mu E_2}{R_2} + \frac{\nu E_2}{R_4} = 0, \tag{2.32}$$

which is satisfied only if $E_2 = 0$, or if $\mu / \nu = R_2 / R_4$, contrary to the assumption that the parameters are arbitrary. Clearly, ψ is not the function to be minimized in the steady-state condition. The question arises, Does a state function minimized in the steady-state condition exist in this case, and, if so, how is it to be interpreted in terms of minimizing the rate of entropy production so as to represent a generalization of the second law [4]?

We see that the generalization, if such can be found, depends on the topology of the circuit. This circumstance raises a general question which, to my knowledge, is still unanswered in general systems theory. How is the second law of thermodynamics to be generalized for open systems *with feedback loops* in an arbitrary topological arrangement? The question is tantalizing inasmuch as feedback loops suggest that *information* is somehow fed into the system, or that the system obtains "information about itself" and so leads to a relation between information and entropy, already noted from the mathematical isomorphism connecting the two concepts.

2.9. STATIC THEORIES

The comparatively recent extensions of mathematical analysis to areas beyond its traditional scope, specifically to the analysis of biological and social systems, have provided ample illustrations of the potential of the mathematical general systems approach. In this connection it is important to point

out that static theories play a more important part in biological and social sciences than in the physical sciences. There are two reasons for this, one fairly obvious, the other more subtle. The obvious reason is that dynamic theories (those which explain or predict the changes that systems undergo in time) are, on the whole, more difficult to construct than static theories concerned primarily with the structure rather than with the behavior of systems. For this reason it could be said that the social sciences have not reached the maturity associated with the construction of dynamic system models. In physics, too, statics antedated dynamics both in mechanics and in theories of electricity and magnetism. Curiously, classical "thermodynamics" is actually a misnomer, the discipline having been concerned primarily with equilibrium states. Nonequilibrium, that is, truly dynamic thermodynamics, is still a quite young branch of physics, as is indicated by its many still unsolved problems.

In biology, too, "statics" (e.g., taxonomy and anatomy) antedate their dynamic counterparts, the theory of evolution and physiology. In the social sciences, as has been said, dynamic theories are still in a nascent state.

The other reason why static theories are of greater importance in biology and especially in the social sciences than in the physical sciences is that the static theories in the former are "richer" than those in the latter sciences. Classification and taxonomy play relatively insignificant parts in physics, a science concerned largely with laws governing the behavior of *all* matter. Classification plays a more important role in chemistry. However, since the discovery of a finite number of elements as the building blocks of all substances, chemical taxonomy presents few difficulties. Methods of classification become of greater importance in biology, where the fundamental concepts, like organism, adaptation, and natural selection, require complex acts of *recognition*. Finally, these methods become of primary importance in social science, where a set of fundamental concepts to serve as the building blocks of a unified theory is very far from a reality. The distillation of such concepts is the goal of primarily static theories.

The primary question of a static theory, "What shall be singled out for attention?", is preliminary to questions of how the concepts representing what has been singled out can be used to describe processes and to derive "if so, ... then so" propositions. Perhaps it is chiefly for this reason that so much of what goes under the name of "social science theories" does not transcend an elaboration of definitions.

We turn now to "*static*" *general systems theory*. Two branches of mathematics are particularly suited to its modern development: the *theory of graphs* and the *theory of games*. In our example of entropy production in an electric circuit, we have already run into a problem of relating the thermodynamics of such circuits to the properties of linear graphs.

2.10. SYSTEMS REPRESENTED BY LINEAR GRAPHS

The *theory of graphs* is a theory of structure par excellence. A linear graph is a set of points and an unspecified binary relation, *R*, which either holds or does not hold between every pair of points in the set. If the relation is symmetric, that is, if *xRy* if and only if *yRx*, the graph is an ordinary (*undirected*) *graph*. Otherwise the system is defined by a *directed graph* (or *digraph*). Any situation consists, in the last analysis, of a set of elements and a set of relations among them. A graph having only a single binary relation is consequently the simplest description of a structure.

A graph can be diagrammatically represented by lines connecting some of its points. If the graph is directed, the lines are arrows; otherwise, not. A picture of a graph immediately suggests a communication network or an organizational chart. Indeed, the basic features of such networks or charts are frequently represented by diagrams of graphs.

A graph is a *topological* rather than a geometric structure inasmuch as the distances between connected points are not part of its definition. Only the *fact* that the two points are or are not connected is indicated. These topological features are at times the most essential ones in the descriptions of many systems. For instance, it may be of importance to know whether it is or is not possible to reach points of the graph from other points via intervening connections, regardless of the physical distance that separates the entities represented by the points. To take another example, a biological species is sometimes defined as a collection of organisms that can interbreed. The possibility of interbreeding does not necessarily depend on the possibility of mating between members of any heterosexual pair. A Chihuahua may not be able to mate with a Saint Bernard, at least not without artificial insemination. Yet the genes of Chihuahuas and Saint Bernards can mingle via mating between members of intervening breeds. Thus the graph showing possible paths of genes of dog species is probably a *connected graph*. An interbreeding population (species) can, perhaps, be defined in terms of a connected graph, where the links represent possible matings.

Strict mathematical isomorphism can be established between two individual graphs in the usual way. If to every point of one graph there corresponds exactly one point of the other (and vice versa), and if the relations between the corresponding points are preserved, the two graphs are said to be isomorphic. All graphs isomorphic to each other naturally define a type or a class of graphs. Thus a systematic taxonomy immediately suggests itself, and with it a taxonomy of systems represented by the graphs. Attention is then directed to those properties of systems that reflect the properties of the types of graphs that represent them. For example, a type of organization can be defined by the

type of graph that represents the flow of information in it or the relations of subordination or superordination. The graph representing the structure of the system may reveal certain critical features, for instance, a critical communication link, which, if cut, would disconnect the graph, or an exceptional " node " through which all paths from one part of the graph to another must pass. Questions about the efficacy of organization or the vulnerability of communication networks often revolve around the existence of such elements.

A serious limitation in the development of a general theory of graphs is the size of the graph. As has been said, the classification of graphs that occurs immediately is that into classes determined by isomorphisms. Roughly speaking, two graphs are isomorphic to each other if one can be turned into the other by relabeling the points. Since n points can be labeled in $n!$ different ways, the number of graphs defined by their equivalence classes is immensely reduced. Thus, if $n = 10$, then 3,628,800 graphs become essentially a single graph. Still, even this drastic reduction leaves a superastronomical number of equivalence classes even when n is only moderately large. For instance, when $n = 10$, the number of labeled pairs is 55, and so the number of distinct labeled graphs is 2^{55}, which is incomparably larger than 3,628,800. Clearly, a systematic study of graphs as mathematical systems must resort to other classifications. The lines of least mathematical resistance suggest singling out some easily describable classes of graphs, for instance, regular graphs, characterized by an equal number of links converging at each point. The usefulness of this approach in general systems theory depends, of course, on the existence of interesting systems with content to which the model of, say, the regular graph applies. If such systems are difficult to find, the applicability of the mathematical theory is questionable.

More promising is the generalization of graph theory in the probabilisitic direction. Instead of considering a specific graph with specific properties (of which, as we have seen, there is an embarrassing multitude), the probabilistic theory puts the " randomly constructed " graph at the center of interest. It examines the probability distributions and the statistical expectations of the properties of such randomly constructed graphs.

As an example, consider a directed graph constructed in the following manner. There are n points. From each point a given number a of links " grow." Each link connects " at random " to one of the points of the graph. More precisely, the probability that a given link connects on a given point is $1/n$. Given this definition of a randomly constructed graph, we can inquire into several of its properties, probabilistically defined. For example:

1. What is the probability distribution of the " in-degree " of the points of the graph ? (The in-degree of a point of a directed graph is the number of links converging on it.)

2. What is the probability that, from an arbitrarily selected point of the directed graph constructed in this manner, a path exists to every other point?

3. If X is the random variable representing the number of points to which a path exists from an arbitrarily selected point, what is the distribution of X? What is the expectation of X?

4. What is the probability that a directed random graph constructed in this manner is connected? What is the distribution of the random variable Y, representing the number of connected components of the graph?

The same questions can be asked about the ordinary (undirected) graph formed by "erasing the arrows" on the links of the directed graph.

As an example, we shall examine the expectation of the number (or fraction) of points to which a directed path exists from an arbitrarily selected point. Assume that the construction of the graph starts with a single point. Let P_1 be the fraction of points contacted by the a links issuing from this point. Actually P_1 is a random variable, but we shall be considering only its expectation. Similarly, let P_2 be the fraction of points *newly* contacted by the links issuing from the points of the first "remove," and so on. The recursion formula for the expectations turns out to be [5]

$$P_{t+1} = \left(1 - \sum_{j=1}^{t} P_j\right)\left[1 - \left(1 - \frac{1}{N}\right)^{aP_t N}\right] \tag{2.33}$$

with the initial condition $P_0 = 1/N$, where N is the total number of points. For large N, the iteration formula can be approximated by

$$P_{t+1} = (1 - X_t)(1 - e^{-aP_t}), \tag{2.34}$$

where $X_t = \sum_{j=1}^{t} P_j$.

We seek $X_\infty = \gamma$, the expected total fraction of points so contacted. The expression for γ is easily obtained. Rewriting P_t as $X_t - X_{t-1}$, we have

$$\frac{X_{t+1} - X_t}{1 - X_t} = 1 - e^{-a(X_t - X_{t-1})}, \tag{2.35}$$

$$\frac{1 - X_{t+1}}{1 - X_t} = e^{-a(X_t - X_{t-1})}, \tag{2.36}$$

$$e^{aX_t}(1 - X_{t+1}) = e^{aX_{t-1}}(1 - X_t) \qquad (t = 0, 1, 2, \ldots). \tag{2.37}$$

Hence

$$e^{aX_t}(1 - X_{t+1}) = K, \text{ a constant.} \tag{2.38}$$

For large N, $X_0 = 1/N \cong 0$ and $X_1 = a/N \cong 0$; hence, $K \cong 1$.

Letting t go to infinity, we have

$$e^{a\gamma}(1 - \gamma) = 1,$$
$$\gamma = 1 - e^{-a\gamma}, \tag{2.39}$$

which is the sought expression for γ.

We note that the graph of γ as a function of a consists of two branches, one of which is $\gamma = 0$. Since only nonnegative values of γ have physical significance, the graph of γ versus a is the line $\gamma = 0$ for $0 \leqq a \leqq 1$, and thereafter the upper branch of the curve approaching $\gamma = 1$ asymptotically.

The process described can be viewed as a simple contagion model. Beginning with a single " infected " individual, the number of infected individuals grows as the *newly* infected come in contact with the non-yet-infected. Clearly, the model differs from the logistic model in that the period of infectuousness of each infected individual is limited to the time during which he comes in contact with a individuals. The model is not substantially modified if the number of individuals with which each infected individual comes in contact is a random variable, whose expectation is a. Equation 2.39 gives the expected fraction of individuals *ever* infected (implicitly) as a function of a. If $a \leqq 1$, this expected fraction is zero, strictly speaking—infinitesimal compared with the total population, assumed to be "infinite." For $a > 1$, the fraction is finite. Specifically, for $a = 2$, the expected fraction is about 0.8; for $a = 3$, it is about 0.95; and so on.

Concrete systems representable by graphs of this sort (modified by biases to be described below) are common. Consider a sociogram. Each individual in a population selects a individuals as his friends or in some other specified manner. If we represent the naming of each individual by an arrow from the chooser to the chosen, we have a directed graph. Clearly, real sociograms are not "randomly constructed graphs," since friends are not randomly chosen. However, having derived certain probabilistic features of a randomly constructed graph, we can examine the corresponding features of sociograms to note the ways in which the latter differ from randomly constructed graphs. The differences may suggest concepts for characterizing biased graphs by the values of their bias parameters.

A bias in the construction of a graph involves a departure from equiprobability or from statistical independence. For example, if, in the construction of the graph described above, the links issuing from each point terminated on the various points with variable probabilities, this would reflect a " popularity " bias: the individuals in the population have unequal probabilities of being chosen as friends. The bias would reflect itself in the departure of the distribution of in-degree from the Poisson distribution, to which it reduces under random choices. In fact, it can be shown that, *whatever* the popularity bias is, the resulting in-degree distribution must be "flatter" than the Poisson dis-

tribution. On the other hand, if we introduce a "reciprocity" bias, the distribution of in-degree should be "sharper" than the Poisson distribution. (A reciprocity bias is the tendency of individuals chosen as friends to reciprocate the choice, hence a departure from statistical independence of choices.)

Another sort of bias that can be introduced into the model is a "transitivity bias," according to which friends of friends are likely to be friends. The effect of both a reciprocity bias and a transitivity bias should be to depress the value of γ for a given value of a. This can be seen easily in the extreme case. If all the choices are reciprocated, the population splits into mutually exclusive cliques, so that the expected number of individuals that can be "reached" by a path of choices from a given individual is no larger than the size of the clique.

Evidence of all three biases has been found in actual sociograms. It is interesting to note that even a considerable reciprocity bias is not sufficient to overcome the popularity bias in the sense of making the distribution of in-degree sharper instead of flatter than the Poisson distribution with the same number of choices per individual.

Another interesting finding is that the order in which friends are named ("best friend," "second best friend," etc.) is reflected in the way in which the number of contacted individuals grows with the removes. As expected (cf. Figure 2.4), it grows more slowly when the graph is constructed by tracing through the first two friends; more rapidly when tracing through the second and third friends; and so on.

Let us now consider the probability that a randomly constructed graph is connected. This problem was attacked by Erdös and Renyi [6], who defined a random graph in a somewhat different manner. Theirs is an undirected graph. Associated with n points are $\frac{1}{2}n(n-1)$ pairs. Of these, N pairs are selected (without replacement) at random and joined with links. The probability that the resulting graph is connected is defined as the ratio of the total number of connected graphs (with labeled points) to the total number of (labeled) graphs with n points and N links. Now, if N is kept constant and n tends to infinity, clearly the probability of being connected tends to zero. However, if N increases along with n, this is not necessarily the case. Indeed, if N increases sufficiently rapidly with respect to n, the probability of connectedness may tend to 1. Erdös and Renyi determined the so-called *threshold function* for $N(n)$, that is, if the function has the form

$$N(n) = \frac{n}{2} \log_e n + an \qquad (a = \text{constant}), \qquad (2.40)$$

then the probability of connectedness tends to a finite limit, namely, $e^{-e^{-2a}}$.

The question naturally arises of how the probability is affected if the graph is biased. For instance, if the points are immersed in a metric space, and if the probability of joining two points is some function of the distance between
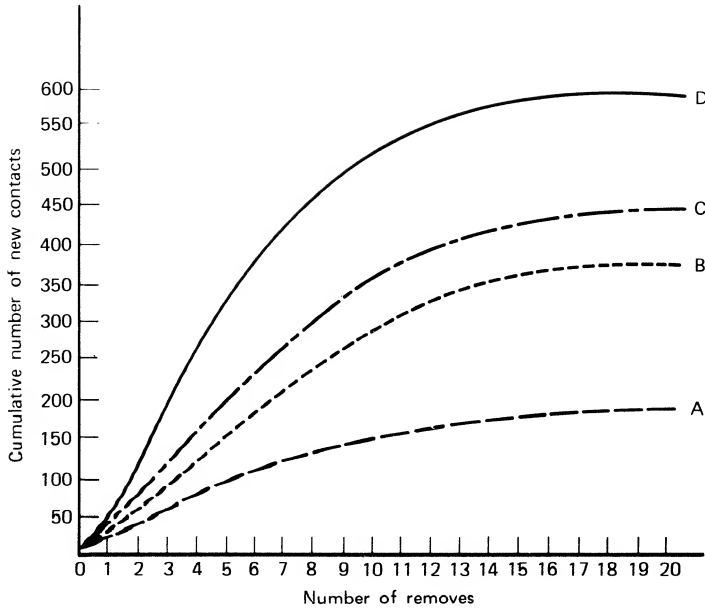
Figure 2.4. A: tracing through first and second friends; B: tracing through third and fourth friends; C: tracing through fifth and sixth friends; D tracing through two randomly selected friends (theoretical.) All observations are averages of 30 tracings beginning with 10 arbitrarily selected individuals.

them, will this bias increase or decrease the probability of connectedness? Intuition does not provide an immediate answer. We have seen from the previous example that a kind of distance bias (the tendency to choose friends of friends and to reciprocate choices) depresses the probability of connectedness of a sociogram, since the expected number of individuals to whom a path exists from a randomly chosen individual becomes smaller thereby. On the other hand, consider the points of a rectangular grid. Assume that four links issuing from every point connect with certainty to the four immediately neighboring points. This constitutes a strong distance bias; yet it ensures that the resulting graph of any size is connected.

The problem of the connectivity probability of a biased graph was attacked in a very special case by Perkins [7]. His graph has n points and $n - 1$ links. Such a graph is connected if and only if it is a *tree*, that is, has no cycles. Perkins introduces the simplest sort of distance bias, namely, the n points are divided into two equal classes. Points of the same class are "near" each other; those belonging to different classes are "far" from each other. Of the $n - 1$ links a certain fraction, θ, are used to connect only pairs of points

within the same class. These are *endogamous* connections. The remaining fraction $(1 - \theta)$ of the links connect only pairs of points belonging to different classes. These are *exogamous* connections. Clearly, if $\theta = 1$, the graph cannot be connected. However, if $\theta = 0$, that is, if all the connections are exogamous, the graph can be connected. For a fixed n, the probability of connectedness may be considered as a function of θ. Perkins has shown that, as θ increases from 0 to 1, the probability of connectedness at first decreases, then passes through a minimum, next increases to a maximum of about $\theta = 0.9$, and finally rapidly decreases until it vanishes (as it must) at $\theta = 1$.

An analogous result for a graph with an arbitrary number of points and links and an arbitrary number of classes might shed some light on a long-standing biological question. Consider a species with certain established mating patterns. If mating is completely random, the graph, in which points represent individuals and links represent matings, is a randomly constructed graph, except, of course, that is is *bichromatic*: links can be established only between members of the opposite sex. This restriction should not be confused with a distance bias imposed on the mating pattern, that is, a preference for either endogamous or exogamous matings. If the population is divided into a number of subpopulations, and if only endogamous matings can occur, clearly the resulting graph cannot be connected. The usual evolutionary consequence of this is eventual separation of the species into several species that can no longer interbreed. However, even if mating is random, the resulting graph, being a random one, need not be connected.

We can now ask the following question: Is there an "optimal" distribution of endogamous and exogamous matings which maximizes the probability of connectedness of the graph representing the flow of genes? In other words, do species subdivided into subspecies with only occasional interbreeding maximize the probability that a mutation "selected for" in the process of natural selection will spread throughout the species? The question concerning the rate of spread is, of course, a related one. These questions are subsumed under the presently unsolved problems of general systems theory.

2.11. THEORY OF GAMES—A GENERAL SYSTEMS APPROACH TO A THEORY OF "RATIONAL" CONFLICT

The illuminating power of static (i.e., purely structural) mathematical models has been demonstrated most clearly in the *theory of games*. Game theory is often depicted as a theory of conflict between or among ideally "rational" actors. Its name derives from games of strategy (chess, bridge, etc.) cited as models of such conflicts. A mathematical model of a system always involves singling out some of its features regarded in some way as

essential and considering the properties or the behavior of such idealized systems, which are defined *only* in terms of the features singled out.

The typical game of strategy certainly involves dynamics, inasmuch as it generally proceeds from move to move and so from situation to situation. A principal achievement of game theory was the reduction of the dynamics of these processes to the logico-strategic structure of the game, that is, to a static description. The concept that effects this reduction of the dynamics of a game to statics is that of *strategy*, defined roughly as a plan specifying what a player will do in every possible situation that may arise in the course of the game. A game with a finite number of players and a finite number of moves has also a finite number of strategies available to each player. To be sure, this number is in general superastronomical. However, as long as it is finite, it allows the game theorist to conceptualize a game as a single, independent choice of strategy by each of the players. Therefore, if the game has just two players (a two-person game), it can be represented as a matrix whose rows and columns represent, respectively, the strategies available to each of the players and whose entries are the payoffs associated with the resulting outcome of the game. If there are more than two players, the representation becomes an *n*-dimensional matrix. A matrix is a mathematical object. Certain properties of game matrices reflect the logico-strategic structures of the games they represent and so induce a classification of games in accordance with criteria especially relevant to questions considered in the theory of "rational" conflict.

The number of dimensions of the matrix (the number of players) provides one obvious classification. Here the very important qualitative differences that distinguish two-person games from games with more than two players (*n*-person games) are already revealed in the properties of *n*-dimensional payoff matrices ($n > 2$) in a way that does not occur on immediately intuitive grounds. These differences are revealed in the different roles played by so-called equilibrium outcomes in constant-sum two-person games and constant-sum *n*-person games. (In a constant-sum game, the sum of the payoffs of the players is independent of strategy choices.)

For the moment, let us consider the payoff matrix of a two-person constant-sum game. Such a matrix may or may not have a saddle point, that is, a payoff entry to the player choosing rows that is at the same time minimal with respect to all the entries of the same row and maximal with respect to all the entries of the same column. It is easily seen that, if there are several such entries, they must all be equal; moreover, they must be so placed in the matrix that any pair of strategy choices that contain saddle points must intersect at a saddle point. It is shown in game theory that all games of perfect information (i.e., those in which each player's moves are known to the other players immediately after they are made) have saddle points.

From the property of saddle points the game theoreticians draw certain consequences. First, a "rational player" playing a game with a saddle point cannot do better than choose a strategy containing a saddle point. Second, it does not matter which strategy he chooses, as long as it contains a saddle point. Finally, in a conflict represented by a game of this sort, secrecy (concealing one's strategy choice from the opponent) confers no advantage.

Properties of games without saddle points are quite different. In these games no "best" strategy can be singled out from among the available ones. The rational player must resort to a *mixed strategy*, essentially to letting a random device choose a strategy for him. Secrecy in the actual choice of strategy is of essence here, but not in the matter of choosing the "optimal mixture."

In this way, a classification of two-party conflicts is suggested that may not have occurred on intuitive grounds. Indeed, games like chess, nim, and tic-tac-toe are all put into one class, as games with saddle points; matching pennies and poker, being games without saddle points, are put into another class. The "complexity" of the game, as conventionally understood, in terms of the number of possible situations that can occur has nothing to do with this classification. As a consequence of game-theoretic analysis, another "dimension" of complexity appears, namely, the criterion that determines whether or not a game *has* a best strategy among the available ones.

When the logico-strategic structure of non-constant-sum games and of games with more than two players is examined, the concept of "rationality" must be still further refined and generalized. It turns out that the prescripts of individual rationality may differ essentially from those of collective rationality, a circumstance that leads to a radical re-examination of the meaning of the phrase "rational decision" and consequently to the raising of important, sometimes perplexing, psychological issues. These issues come to the foreground in the wake of paradoxes that arise when the concepts of rationality that are adequate on one level of conflict (e.g., two-person constant-sum games) are applied on another level (e.g., non-constant-sum and *n*-person games). These paradoxes are revealed in the process of pure logico-strategic analysis of conflict situations, that is, entirely in the spirit of general systems theory.

Again we see the fundamental difference between the general systems approach and the method of mathematical model building. An attempt to construct a mathematical model of some conflict abstracted from a real-life situation would lead to attempts to bring the model into agreement with the observed characteristics of the conflict. These attempts may or may not be successful. If they are, the problem of generalizing the results still looms. The abstract mathematical approach singles out an aspect of conflict that is, to begin with, conducive to mathematical description. The resulting analysis

does not, in general, yield results that are applicable to descriptions of real-life situations, because game-theoretic results are much too abstract. Nevertheless, the theoretical yield of this approach is great. It leads to new concepts which do not ordinarily emerge from generalizations of observations. Thus, in presenting the method and content of game theory, it is often necessary to dispel a prevalent and understandable misconception that game theory is concerned with the strategic analysis of specific conflict situations, in particular, of specific games of strategy. In fact, however, hardly any finding of game theory is of use in acquiring skill in any game of strategy worth playing, because any game of sufficient complexity to be of interest is far too complex to yield to an algorithm designed to find the best strategies. What game theory does is establish the *meaning* of "best strategy" when such exists. It raises and clarifies the question of whether a "best strategy" exists at all in the various senses in which it can be defined. It says a great deal about the nature of the conceptual problems involved in logico-strategic analysis.

This sort of clarification characterizes all mathematical research directed at the investigation of general systems rather than at the solution of specific problems. For instance, the general theory of differential equations is directed not so much at finding solutions of differential equations as they present themselves, as at establishing existence and uniqueness theorems and at investigating the general nature of solutions of different classes of differential equations and of sets of differential equations. The results of this sort of research are an enlargement of the conceptual repertoire, a building up of "intellectual capital," so to say, that may yield rich dividends in later specific investigations. At this point it is important to note that later specific investigations to which the theory may direct our attention may be quite different from those that we might have undertaken, had our conceptual repertoire not been enlarged by the general theory.

A case in point is the investigation of so-called *cooperative* games inspired by the limits imposed on "classical" game theory. By "classical" game theory we may choose to understand the theory of the two-person constant-sum game that culminated in von Neumann's minimax theorem [8]. The theorem establishes the existence of a "best" strategy available to each of the two players of a constant-sum game, this strategy being either "pure" (if the game has a saddle point) or "mixed" otherwise. This best strategy turns out to be a minimax strategy, one that guarantees a certain minimal payoff to the player. It also turns out, however, that the application of this principle in non-constant-sum games does not in general result in the best outcomes that players can achieve collectively. Introducing the concept of collective rationality is tantamount to introducing the assumption that whatever the entire set of n players can achieve by cooperating, regardless of the conflict of interest that divides them, they will achieve. However, the

question of how this optimal joint payoff will be apportioned among the individual players still remains open. In this way, the thrust of game theory is directed away from the search for optimal strategies for each individual player and toward the search for principles by which to *adjudicate* the conflict of interest among the players who have cooperated to achieve the jointly optimal goal and are now faced with the problem of dividing the prize.

The shift of focus raises problems that cannot be formulated without certain assumptions of "equity" or analogous concepts that are irrelevant in the context of "classical" game theory. The necessity for introducing such concepts without abandoning the standards of rigor demanded in mathematical analysis compels the game theorist to face problems inherent in some areas of philosophy. A case in point is the so-called utilitarian calculus, which poses the problem of choosing a course of action that leads to "the greatest good for the greatest number." The operational definition of this comfortable phrase is far from obvious. "The greatest good for the greatest number" can be easily defined if one assumes a *welfare function* that assigns a "social utility" to each outcome of a decision. However, this social utility is meaningful only if certain assumptions are made about the nature of "utility"; for instance, that utility is not only a measurable quantity, but also a transferable and conservative one. This assumption clearly is not warranted in many specific situations, and the theoretician is forced to redefine "collectively optimal decision" independently of a welfare function. Thereby the problem of apportionment must also be redefined. For instance, the solution of the two-person non-constant-sum game proposed by Nash [9] is independent of a welfare function. It is derived from a set of axioms in which the possibility of interpersonal comparison of utilities is specifically denied. This denial is reflected in the requirement that the solution must remain invariant with respect to *independent* positive linear transformations of each player's payoffs, thus rendering the "sum of the payoffs" mathematically meaningless. Nash's solution was later extended by Harsanyi [10] to the *n*-person game. In contrast, the solution of the *n*-person game proposed by Shapley [11] (the so-called *Shapley value*) rests on the assumption that utility is a transferable, conservative commodity, hence that interpersonal comparison of utilities is possible.

All these analyses are made in the context of taking into account only the logico-strategic structure of a conflict of interest, totally abstracted from empirical content. As such, the analysis reflects the activity of the pure mathematician. However, as soon as the conclusions are applied to the classification of "conflicting systems" on the basis of the logico-strategic structure of the conflicts, with a view to drawing structural analogies between conflicting systems of similar structure but widely differing empirical content, one passes into the realm of general systems theory.

As usual, the principal "payoff" of such analysis is conceptual rather than pragmatic. For instance, it could be as difficult to apply the results of n-person game theory to "conflict resolution" as to apply the results of classical game theory to techniques of winning games. What the analysis does is clarify the essential features of the problems involved in a way that is seldom possible in concrete situations obscured by the complexities and ambiguities of their particular aspects.

As a simplest example, we may cite the result, proved in n-person game theory, that all three-person constant-sum games fall into just two classes. In one, including so-called *inessential* games, no advantage accrues to any pair of players if they join in a coalition against the third. In the other, comprising so-called *essential* games, such advantage does accrue. Moreover, all the essential games are, in a sense, isomorphic to a single game, which can be formulated as follows: A unit of utility is to be apportioned among the three, and any two can appropriate the entire unit. In this way, the notion of isomorphism reduces what seems to be a large class of problems to a single problem, effecting a drastic conceptual simplification. Essential games with more than two players can no longer be subsumed under a single game, but again the notion of isomorphism suggests equivalence classes, each of which includes a large number of games with an identical logico-strategic structure and induces a corresponding taxonomy of "rational" conflicts.

2.12. CONCLUSION

The rewards of general system analysis come typically in the form of new problems rather than in the form of solutions to old ones. The rewards are none the less real, since the formulation of new problems usually involves a sharpening of newly found concepts and a redirection of intellectual energy to new, sometimes virginal, domains.

Specifically, the trend in general systems theory centered around exploiting the notion of mathematical isomorphism in the interest of creating a purely logico-structural, content-free taxonomy of systems has yielded rich intellectual dividends. This approach blends, on the one side, with that of the pure mathematician and, on the other, with the builder of mathematical models abstracted from specific content. The central goal of this approach, however, is that of investigating the consequences of the classification of systems *induced* by mathematical isomorphisms. In a way, the method provides the antidote to the fractionation of science generated by increasing specialization, and to the "exhaustion of intellectual capital" against which Alfred North Whitehead warned almost half a century ago.

PROBLEMS

2.1. Consider the pair of differential equations

$$\dot{x} = ax + by + h, \qquad \dot{y} = cx + dy + k.$$

Assign all possible combinations of signs to the parameters, and examine the resulting system for stability.

Under what conditions does an equilibrium exist?

Under what conditions is the equilibrium in the first quadrant?

2.2. Given $\gamma = 1 - e^{-a\gamma}$, plot γ against a.

2.3. A graph with 5 points and 6 links is constructed as follows. Six of the 10 pairs of points are selected at random and each pair is joined by 1 link. What is the probability that the resulting graph is connected?

2.4. A graph with 5 points is constructed as follows. A point is selected at random. Two links issue from it, each of which terminates on any of the 5 points with equal probability. From each of the points so contacted, 2 links issue again, terminating on any of the 5 points with equal probability. The process continues until the number of links equals or exceeds 6. What is the probability that the resulting graph is connected?

2.5. A directed graph represents a dominance relation if for every pair of points A, B, either $A > B$ (A dominates B) or $B > A$ but not both. How many nonisometric graphs of 5 points are there?

2.6. Show that, if a matrix representing a two-person zero-sum game has several saddle points,

(a) the payoffs at all saddle points are equal;

(b) if each player chooses a strategy containing a saddle point, then the outcome is a saddle point.

2.7. Show that, if the game is not zero-sum, neither (a) nor (b) in Problem 2.6 is necessarily true.

2.8. Partition the directed graphs in Figure 2.5 on the basis of the isomorphic relation. Determine the one-to-one correspondence between the two point sets for each pair of isomorphic graphs.

2.9. A finite-state (a Mealy) machine is defined as a quintuple $M = (X, Y, Z, f, g)$, where X is a finite nonempty set of stimuli, Y is a finite nonempty set of responses, Z is a finite nonempty set of internal states, f is an output function: $y = f(x, z)$, and g is a transition (or next-state) function: $z' = g(x, z)$. Symbols $x \in X$, $y \in Y$, $z \in Z$, and $z' \in Z$ represent, respectively, the present stimulus, response, internal state, and next internal state. Specify the conditions under which machine $M_1 = (X_1, Y_1, Z_1, f_1, g_1)$ is isomorphic to machine $M_2 = (X_2, Y_2, Z_2, f_2, g_2)$.
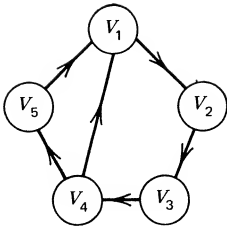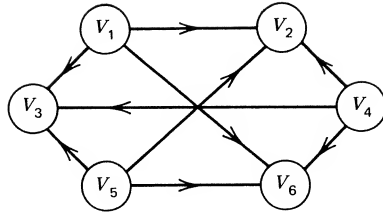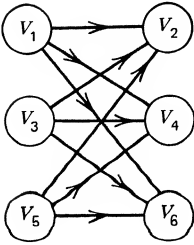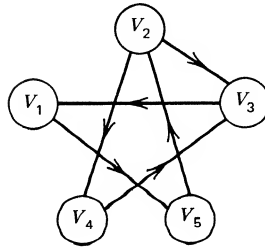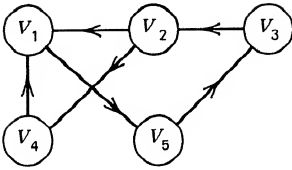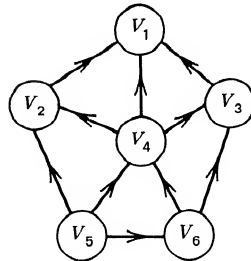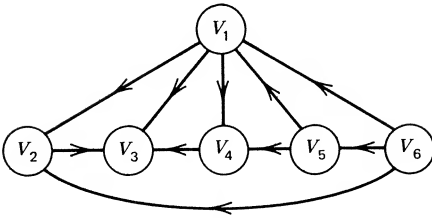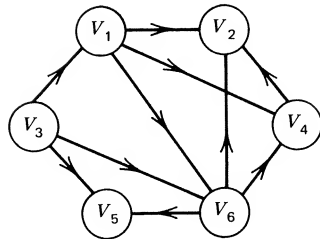
Figure 2.5. Examples of directed graphs (Problem 2.8).

REFERENCES

1. **Mandelbrot, B.,** ''An Informational Theory of Statistical Structure of Language.'' In *Communication Theory*, W. Jackson (Editor). London: Butterworth, 1953; New York: Academic Press, 1953.

2. **Bertalanffy, L. von,** ''General System Theory.'' In *Main Currents in Modern Thought*, Vol. 71, 1955. Reprinted in *General Systems Yearbook*, Vol. 1, pp. 1–10, 1956.

3. **Richardson, L. F.,** *Arms and Insecurity*. Pittsburgh: The Boxwood Press, 1960; Chicago: Quadrangle Books, 1970.

4. **Foster, C., A. Rapoport,** and **E. Trucco,** ''Some Unsolved Problems in the Theory of Non-Isolated Systems.'' *General Systems Yearbook*, Vol. 2, pp. 9–29, 1957.

5. **Solomonoff, R.,** and **A. Rapoport,** ''Connectivity of Random Nets.'' *Bulletin of Mathematical Biophysics*, Vol. 13, pp. 107–117, 1951.

6. **Erdos, P.,** and **A. Renyi,** ''On the Evolution of Random Graphs.'' *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, Vol. 5, pp. 17–61, 1960.

7. **Perkins, L. S.,** *A Study of Random Graphs*. Doctoral Dissertation. University of Michigan, Ann Arbor, 1970.

8. **Neumann, J. von,** ''Zur Theorie der Gesellschaftspiele.'' *Mathematische Annalen*, Vol. 100, pp. 295–320, 1928.

9. **Nash, J. F.,** ''Two-person Cooperative Games.'' *Econometrica*, Vol. 21, No. 1, pp. 128–140, 1953.

10. **Harsanyi, J. C.,** ''A Bargaining Model for the *n*-Person Game.'' In *Contributions to the Theory of Games*, IV, R. D. Luce and A. W. Tucker (Editors). Annals of Mathematics Studies, Vol. 40. Princeton, N.J.: Princeton University Press, 1959.

11. **Shapley, L. S.,** ''A Value for *n*-Person Games.'' In *Contributions to the Theory of Games*, II, H. W. Kuhn and A. W. Tucker (Editors). Annals of Mathematics Studies, Vol. 28. Princeton, N.J.: Princeton University Press, 1953.

3. Systems and Their Informational Measures

W. ROSS ASHBY

University College, University of Wales, Cardiff

3.1. "MODERN" SCIENCE

At about the time of World War II, science developed a vast new extension, of which we see today only the beginning. Until that time the method of science was essentially to take everything to pieces and to examine each piece in isolation. A living organism was dissected to organs, these were studied as collections of cells, and each cell was examined as a collection of molecules. The endocrine *system* was studied through the reactions of each gland, isolated as far as experimental technique allowed. The philosophers of science tended to exalt this analytic approach as the uniquely "scientific" method. With the method went such rules as "Vary only one factor at a time."

For two hundred years (after Newton) this method yielded such an abundance of discoveries and advances that most workers felt little inclination to complain; the biologists, however, were uneasy, feeling that something

important was missing. Unfortunately, they had no *rigorous* language in which to say what was missing; hence author after author tried to say, " The whole has something not to be found in a collection of its parts," but succeeded only in convincing the already convinced. Then, in the 1930's, general systems theory arose, mostly through the work of Ludwig von Bertalanffy, who saw not only that the study of parts (in " classic " science) must be supplemented by the study of wholes, but also that there *exists* a science of wholes, with its own laws, methods, logic, and mathematics.

Also working in the 1930's, R. A. Fisher had appreciated how fundamentally limited was the approach through analysis, taking one variable at a time. Being compelled (professionally) to work with agricultural soils, he was forced to insist, first, that the method of studying all the variables, one at a time, would exhaust many lifetimes, and, second, that the method of examining the parts individually was fundamentally incapable of giving information about the interactions between the variables.

Thus arose one of the basic contributions to the modern epistemology of " the system." It was based, it should be noted, not on the *results* of field experiments but on getting clear, *before an experiment was started*, what could give information about what. Fisher insisted, for instance, that a series of plots all given nitrogenous fertilizer, together with a series all given phosphates, could not provide information about what would happen if the two were given together. (In soil, nonadditivity is so common and gross that one must expect and prepare for it.) He defined " interaction " so clearly, and showed so well how to measure it, that by the time of World War II it had become a routine ingredient in the analysis of variance.

The third major component of the modern epistemology of systems was also developing at that time. The biologists had long been aware of the importance of two-way causation, when A affects what happens at B, and B affects what happens at A. The basic relation of organism to environment shows just this circularity. But the biologists had insufficient mathematical or logical techniques for handling such a conceptually difficult process, and attempts to treat the matter in plain language usually ended in confusion. Meanwhile, however, the radio engineers, having discovered the tremendous potentialities in " reaction " (feeding some of the output back to the input), were also having great difficulties, for they too found such circular causal chains most troublesome. Nyquist's theorem [1], however, provided the basis for a breakthrough; by World War II powerful techniques were available for understanding not only how such circular systems behave, but also how a variable can interact with its own past values, and for handling the problems involved.

After World War II these separate understandings, and the technical methods they had engendered, became generally known. They were at once

seen to have applications far wider than to soil and radio; the modern science of systems had arrived. The development of the large, general-purpose computer, coming at the same time, reinforced the change, partly by making systems study practical, and partly by encouraging everyone to think about systems as changing by small discrete steps, a method that makes many of these matters much easier conceptually.

Today, science is developing along lines not included in the "classic" form. The classic method, as stated previously, dealt essentially with parts alone: the difficulties of "interaction" were evaded. Gases were studied in the "perfect" form (notice the adjective!), in which the molecules were so far apart that the chance of two colliding was infinitesimal. A ray of light was regarded as unaffected by the simultaneous transmission of another ray of light. The force which one electric charge exerted on another was unaffected by the presence of a third.

Even when classic science seemed to be dealing with a complex interacting system, its successes were achieved almost entirely with the "linear" type. An example is the surface of a lake, disturbed by, say, the splashes from a few thrown pebbles. Now, although the *position* of each water atom on the surface depends, as the circles of ripples spread, on *all* the circles, each set of circles spreads in its own way, riding over the others as if they were not there. In other words, the patterns set up by the pebbles do not interact. This illustrates the essential meaning of the widely used *principle of superposition*.

"Modern" science, however, is characterized by an uninhibited advance into the nonlinear. It not merely studies systems with high internal interaction but also confidently tackles systems in which it is the interactions themselves that are of interest. To the biologist (such as I am, coming from medicine) the change is immense. Whereas in the past the only resource for dealing with biological systems was to try to minimize the interactions between the parts, thereby often losing the real focus of interest, today nothing but time and money prevent us from treating real biological systems in all their complexity and richness.

3.2. INFORMATION IN A SYSTEM

What, then, of the future trend of systems theory? The subject has greatly expanded since its inception a generation ago. Today no one can be expert in all its branches, and I have had to specialize. What has attracted me has been a growing conviction that one of the key aspects of the subject is the *quantities of information* that are involved in the questions and operations variously considered.

My interest in this aspect dates from about 1954, when, having made the Homeostat [2, 3], I tried to advance to an adapting network. Taking 100 double triodes, I arranged each so that it could accept three inputs and give two outputs, and so that they could be joined freely into a net, either of planned connections or at random. I spent the next two years in deepening frustration, as I totally failed to grasp either what the net was doing or what I, in principle, could get it to do. I ended, however, with a deep conviction that the major factor in my defeat was that the quantities of communication (from machine to me as I observed it, or from me to machine as I controlled it) were impossibly large. The question " How large ? " leads to a whole branch of science (in communication theory) that I am still developing. Since then, all my experience has confirmed my impression, so I will now outline this trend in systems theory, as I see it at present.

3.3. SETS AND SYSTEMS

The application of communication theory to these topics has been greatly facilitated by the gradual realization (perhaps due to our having the digital computer always in our thoughts) that "machines" and other dynamic systems may perfectly well be regarded as going in small discrete steps. Though such variables as position or color may be thought of as naturally continuous, experience has shown that the application of communication theory to continuous variables is severely hampered by technical mathematical difficulties—a stimulating challenge to those who like mathematics for its own sake, but a heavy drag on those who just want to use it as a tool. When the variables are discrete, however, communication becomes conceptually very simple, essentially just counting, a branch of combinatorics. Furthermore, in this discrete form the whole theory can be joined to the theory of sets, especially in the form originally developed by the Bourbaki [4] group, and so linked to the whole body of rigorous mathematics. As such sets need not be restricted to real numbers (but might be the meteorologist's five types of cloud, or the three types of earned income, or the several strains of the influenza virus), the biologist is at last free to use the variables that are meaningful to him [5].

The Bourbaki school has shown abundantly how every complex relationship may be regarded (rigorously) as corresponding to some *subset of a product* (Cartesian) *space*. As a *sub*set, the relationship represents a selection and thus joins precisely to information theory, which concerns itself largely with measuring the intensities of selections. In Shannon's [6] original formulation the product space had two dimensions: x the sender, and y the receiver;

and any point $\langle x, y \rangle$ in it would correspond to "Message x sent and message y received." But Shannon's methods generalize readily and naturally to n dimensions [7–12], and I shall use this form here.

3.4. n-VARIABLE INFORMATION

The basic idea of the method is that, with a system of n variables, X_1, X_2, ..., X_n, the entropy of each one, $H(X_i)$, can be measured directly on that variable without reference to the others. If we also know the conjoint behavior (of the whole, as a vector), we can estimate the entropy of the system, $H(X_1, X_2, \ldots, X_n)$. Then the total transmission between the n variables is defined by

$$T(X_1 : X_2 : \ldots : X_n) = H(X_1) + \cdots + H(X_n) - H(X_1, \ldots, X_n).$$

This total transmission, a measure of the departure of the whole from internal statistical (probabilistic) independence, can then be analyzed quantitatively in various ways so as to throw light on the various internal relations between the parts.

In biological systems, the *probabilities* on which entropy was originally defined are often difficult to define or even of doubtful existence. "The probability that *this* starling will emit a call within the next 10 seconds"—how, operationally, can one decide whether the value should be 0.13 or 0.28? I have found, however, that one can often advantageously use the rather different basis of asking to what the entropy (or transmission) *would* rise *if* the extreme should happen, and all the probabilities be equal. The arithmetic in the two cases is identical, but whereas the first rests on a highly questionable assumption the second asks a straight question having a straight answer. We compute, in other words, not *the* channel capacity but what may be called the *least safe capacity*.

The latter quantity, though seldom of interest in engineering and infrequently mentioned in that field, is likely to be of considerable and practical importance in biological cases. An example may make the point clear. Suppose that a channel is to be able to carry four messages, one of which is much commoner (94 %) than the other three (3, 2, and 1 %). If the messages occur independently in time, the entropy is 0.415 bit, and a channel of this capacity will suffice, provided averaging is allowed. But suppose that the situation is one in which averaging, with its consequent delay, is not tolerable. Suppose, for instance, that the sender is a lighthousekeeper, some miles out from port, who has to transmit, as each ship arrives, one of these four messages:

Pilot required	(probability 0.94)
Customs required	(,, 0.03)
Health inspector required	(,, 0.02)
Police required	(,, 0.01)

In such a case the inequalities of the probabilities, though lowering the entropy, do nothing to help the sender; at every event he must be (equally) ready for all four, and his least safe capacity is 2.00 bits. Thus, in many biological cases, this quantity may well be more realistic than the originally defined capacity.

3.5. INTERACTION AND INFORMATION

This approach enables us to come to grips with the fundamental question: When an experimenter is trying either to study a system or to control it, how much will the informational quantities be increased if the system is changed from one having no interaction between its parts to one having full interaction between them?

What *will* happen cannot be predicted, for this will depend on the particular system, but the least safe capacity can be found unambiguously. Suppose that the system has n variables (any of which may itself be a vector without disturbing the argument), and suppose also, for simplicity, that each variable can take k values. (Inequalities of the k's is easily allowed for, but would be merely distracting here.) Consider first the set of systems with *no* interactions between the variables. In this set, the property holding for each variable (e.g., temperature is between 35 and 37°C) is *not* conditional on the values of the other variables. Hence the subset, corresponding to the relation embodied by the system, is some *rectangular* subset. (It may be only a single point if the system is defined very strictly, or a large region if defined broadly.) The number of such rectangular subsets can easily be reckoned: each variable can provide 2^k subsets of *its* values, and n of such variables provide $(2^k)^n$, that is, 2^{kn}. For the selection, or identification, of one of them, the least safe capacity required is its logarithm to base 2, that is, kn bits.

When interaction between the parts is allowed without restriction, the system may embody a relation corresponding to *any* subset of the space (its size also depending on its strictness of definition). As the whole space has k^n points, its subsets number $2^{(k^n)}$, and the selection, or identification, of one of them now requires (as least safe capacity) k^n bits.

Since k must be 2 or larger, these two quantities, kn and k^n, become enormously different when n is large (as is the case in the really interesting *systems* theory!) As an example, let k be only 2 and let $n = 1000$ (small in comparison with a slum of 10,000 persons, or a computer of 10^6 components,

or a brain of 10^{10} neurons). Without interaction, the least safe capacity (between experimenter and system) is 2000 bits, an easily manageable quantity. With interaction between the parts, however, it is 2^{1000}, that is, 10^{300} bits, larger by 297 orders of magnitude!

Clearly we may say that allowing interaction causes an *enormous* increase in the magnitude of the least safe capacity. And we can see that, when full interaction is occurring, when the system is a "whole" in the fullest sense, no mere doubling or trebling of the resources, or even a multiplying by a millionfold, is likely to be of any use.

3.6. THE QUANTAL LIMIT

The number 10^{300} is likely to be dismissed as fantastic, but I suggest that we examine it seriously. A useful fixed point, by which we can keep some sense of proportion, has been given by Bremermann [13]. He has shown that, because of the fundamental quantal coarseness of matter, nothing made of matter, whether as computer or brain, can transmit (or process) information at a rate faster than 10^{47} bits per gram per second. Take tons of computer and centuries of time, and you add only a few units to the exponent; therefore about 10^{70} bits is an undoubted bound to the quantities that are physically achievable.

Again, 10^{70} bits is likely to seem, at first glance, enough to satisfy the most ambitious systemist. A few trials, however, soon show that the number is, in fact, seriously restrictive when we deal with systems that are of the size commonly considered today and in which the interaction is really high. Here is one of many possible examples. Recent work has shown that the retina does not just send a point-to-point copy of the light-picture to the brain. Active processes go on in the retinal nerve net, so that what is reported to the brain about one point is highly conditional on (is a function of) what is happening at other points. Physiologists working on this "integrative" aspect no longer ask, "How does this *point* react to light?" but rather inquire, "How does this *system* (the retina) react?" By how much have they increased the amount of information that must come to them if they are to answer the question? The crudest calculation is sufficient to make the situation clear. Suppose that the retina has a million cells, each of which (for simplicity) is either active or inactive. Suppose, again for simplicity, that the end event is simply whether or not some cell in the cortex fires. To be able to say, "I now know how this system behaves," is to be able to say which of the $2^{1,000,000}$ states of the retina lead to firing and which do not. The least safe capacity requires, without further special conditions, a 1-bit transmission for each of these states. Thus the physiologist's first attempt to deal with

the retina as a whole implies, *prima facie*, an obligation to transmit $10^{300,000}$ bits, a quantity far beyond the quantal limit.

This example, in no way atypical, shows how the informational demand can increase explosively when interaction is allowed to enter freely, that is, when we become a systemist wholeheartedly. Systems theory is essentially a demand that we treat systems as wholes, composed of related parts, between which interaction occurs to a major degree. No one supports this demand more willingly than I do, but the examples given above show that, having won our battle for the admission of interaction, we must now learn moderation.

But what does " moderation " mean here ? The huge numbers have occurred primarily because we took the cases in which the interaction is extreme: every variable depended on (or entered into a function with) *every* other variable. Although this situation is obviously central in any theory of systems, and we should know as much about it as possible, it is rare in our terrestrial systems. A cell in the retina is acted on by some of its neighbors, but not immediately by every other retinal cell. In a society, not everybody communicates with, or is affected directly by, everybody else. In an oil refinery, not every variable is a direct function of every other variable. Systems theory, having broken away successfully from the extreme "classic" attempt to treat the whole as consisting of isolated parts, cannot go to the other extreme, in which the interactions are total, without exceeding the quantal limit. The future of systems theory, therefore, seems to lie in the study of systems that are sufficiently connected to be real systems yet by no means totally connected.

If this thesis is granted, it follows that, as the difficulty (with complex systems) is essentially quantitative, the further studies must also be essentially quantitative. In other words, we must get to know more about these "quantities of information" in complex systems. To me, this is the outstanding need of general systems theory for the next decade.

What more is there to know? I will now give some examples, mostly of recent development, to illustrate some of the various lines of research and development possible.

3.7. INFORMATIONAL REQUIREMENTS FOR COORDINATION

It has been shown [14] that any well-defined coordination (e.g., among the limbs of a tightrope walker, among the aircraft around an airport, among the fingers of a pianist) necessarily defines a least safe capacity between the parts involved. The basic idea is simply that the "coordination" of parts implies that their behaviors must deviate from statistical (probabilistic) independence: the deviation then provides the datum from which entropies, transmissions, and interactions can be computed.

The numerical quantities obtained may be of interest in many ways, often not so much for the numbers themselves as for the light they shed on some more general aspect. Thus one result, due to a suggestion by David Walter, is a demonstration that the amount of communication required for coordinating a number of parts does not necessarily increase if the number of parts increases. More parts may, in fact, require less communication. Here is an example to make the matter clear.

Let us suppose that the parts are vehicles on Jupiter, and that the communicational aspect is the dominating technical difficulty. Three sites on Jupiter are involved, such that one of them will have to be visited by a vehicle (but which of the three sites is not yet known). When a vehicle goes to the specified site, any other vehicles available must go to the other sites, but not two to the same site. (Which vehicle goes to the specified site does not matter.) The conditions can be met by one, two, or three vehicles: which number is least demanding in the total quantity of communication?

When there is only one vehicle, only the site, selected as 1 from 3, has to be transmitted, and $\log_2 3$ $(= 1.585$ bits) is the least safe capacity. Perhaps it is clearer (especially in the subsequent cases) to compute the transmission between S, the site selected, and X, the site visited by the (single) vehicle. Label the three sites a, b, and c. Only three combinations satisfy the conditions:

$$\langle s, x \rangle = \langle a, a \rangle, \qquad \langle b, b \rangle, \qquad \langle c, c \rangle.$$

Then $H(S) = H(X) = H(S, X) = \log_2 3 = 1.585$ bits and

$$T(S : X) = 1.585 \text{ bits.}$$

When there are three vehicles, the transmission is concerned entirely with securing the 1 : 1 distribution (for that ensures that the site will be visited). The combinations that satisfy the conditions are the 18 for each value of S with a permutation of a, b, c. Call the vehicles X_1, X_2, X_3. Each, as a variable, is able to take any of the values a, b, c. Then $H(S) = H(X_1) = H(X_2) = H(X_3) = H(6, 6, 6) = \log_2 3 = 1.585$ bits. $H(S, X_1, X_2, X_3) = \log_2 18 = 1 + 2 \log_2 3$. Hence

$$T(S : X_1 : X_2 : X_3) = 2.170 \text{ bits.}$$

The two extra vehicles have raised the communicational needs by 0.585 bit.

When there are two vehicles, the combinations permitted are these 12:

$$\langle s, x_1, x_2 \rangle = \begin{array}{lll} \langle a, a, b \rangle & \langle b, b, a \rangle & \langle c, c, a \rangle \\ \langle a, a, c \rangle & \langle b, b, c \rangle & \langle c, c, b \rangle \\ \langle a, b, a \rangle & \langle b, a, b \rangle & \langle c, a, c \rangle \\ \langle a, c, a \rangle & \langle b, c, b \rangle & \langle c, b, c \rangle \end{array}$$

$H(S) = H(X_1) = H(X_2) = \log_2 3 = 1.585$ bits; $H(S, X_1, X_2) = \log_2 12 = 2 + \log_2 3$. Hence

$$T(S : X_1 : X_2) = 1.170 \text{ bits},$$

less than either of the others. Thus the plausible "Send the fewest vehicles since this *must* imply the lowest communicational demand" is wrong.

3.8. THE NEED FOR SYMBOLISM

The translation of these abstract ideas into actual practical details (e.g., verbal instructions to the drivers of the vehicles, radio signals to their electronic controls) is part of the general topic of "coding." In this regard, another study of coordination has given an interesting sidelight on the need for symbolic, rather than actual, manipulation of variables.

Consider four variables—X_1, X_2, X_3, X_4—each of which may, if unrestricted, take any of the four values a, b, c, d, but which are constrained so that no two variables take the same value. As $H(X_1) = 2$ bits, and $H(X_1, X_2, X_3, X_4) = \log_2 24$, therefore

$$T(X_1 : X_2 : X_3 : X_4) = 3.415 \text{ bits}.$$

Such a total may be partitioned in many ways, and each one will correspond to some method of achieving the coordination. Thus the identity [12]

$$T(X_1 : X_2 : X_3 : X_4) = T(X_1 : X_2) + T(X_1 X_2 : X_3) + T(X_1 X_2 X_3 : X_4)$$

corresponds to the method by which X_1 and X_2 select their values in relation to one another while ignoring those of X_3 and X_4; then X_3, knowing the values taken by X_1 and X_2, selects its value appropriately; and finally X_4, knowing the values of the other three, selects its own.

Another way, the one that interests us here, operates through the identity [12]

$$T(X_1 : X_2 : X_3 : X_4) = T(X_1 : X_2) + T(X_3 : X_4) + T(X_1 X_2 : X_3 X_4).$$

This would correspond to an arrangement in which X_1 and X_2 mutually adjust their values (ignoring X_3 and X_4), while *independently* (perhaps simultaneously) X_3 and X_4 make their adjustments (ignoring X_1 and X_2). The third term requires also an adjustment of the *pair* (vector) $X_1 X_2$ to the *pair* $X_3 X_4$, while doing nothing to affect the adjustments within each pair. The actual quantities can readily be found to be (respectively)

$$3.415 = 0.415 + 0.415 + 2.585,$$

stating that the transmission between, say, X_1 and X_2 need not exceed 0.415 bit if the right code is used. At the first attempts to find a suitable code, all demanded more. Thus, if X_3 and X_4 have chosen b and c, then X_1 and X_2 apparently have to use the channel

$$
\text{Value of } X_2 \quad
\begin{array}{c|cc}
d & 1 & 0 \\
a & 0 & 1 \\
\hline
  & a & d
\end{array}
$$

Value of X_1

whose use demands 1 bit (per event).

Eventually a coding (i.e., a practical method of acting) is found that requires only the 0.415 bit, but it has a peculiar requirement: the values of X_1 and X_2 in the channel must come from a different set from those used in $T(X_3 : X_4)$. In other words, X_1 and X_2 must be coded into some arbitrary set (e.g., p, q, r, s) that will only later be translated into the actual a, b, c, d. Then X_1 and X_2 can use the channel

$$
\text{Value of } X_2 \quad
\begin{array}{c|cccc}
s & 1 & 1 & 1 & 0 \\
r & 1 & 1 & 0 & 1 \\
q & 1 & 0 & 1 & 1 \\
p & 0 & 1 & 1 & 1 \\
\hline
  & p & q & r & s
\end{array}
$$

Value of X_1

whose capacity is just the 0.415 bit (per event) required. (One is reminded here of the detailed planning that occurred before the landings in Normandy, when the provisional plans used such symbolic dates as D-day, D-plus 3, ..., with the real value of " D " left conditional on other plans being adjusted elsewhere.) The partition asserts that the two channels, each of capacity 0.415 bit, must act independently if the whole coordination is to be achieved with full efficiency. If the two channels try to work with the real, final events, they will have events in common and cannot be independent. Thus the use of *symbols* is a necessity.

3.9. INFORMATION AND REGULATORS

Information theory may also come to play a central role in the theory of control and regulation. This is the hard core of system theory, for here occur the problems that demand real skill, not mere verbal plausibility. In regard

to regulation, Conant [15] has proved the fundamental theorem (superseding the "law of requisite variety") that the capacity of any device or system as a regulator cannot exceed its capacity as a transmitter of information. The relation between regulation and transmission was proved by him, in fact, to be not just possible but fundamental. Here again, information theory is used not just to make measurements, but also to give a deeper insight into the nature of the process. Thus Conant has shown that the error-controlled regulator, often mistakenly thought to be the only type, is fundamentally distinct from the cause-controlled regulator and basically inferior to it (though often technically easier to make).

3.10. INFORMATION REQUIREMENTS OF DESIGN

Even within the design of a single system (when it becomes really complex) the informational point of view may well provide valuable strategic guidance. Consider, for instance, the very general case in which a designer is asked to design a system so that its output is some assigned function of its input. Usually the request comes in one form (e.g., construct a radar system of such-and-such sensitivity, costing not more than x dollars), and the designer translates this form into one on a drawing board, strewn with transistors, wires, and other components.

The whole situation may then be represented by a diagram like that in Figure 3.1. It looks familiar enough, but here I wish to focus attention on channel C, from designer to system, rather than on the more usual channel from X to Y. Certainly C *is* a channel (in any sufficiently well-defined case.) "Being a good designer" implies that he embodies a correspondence between
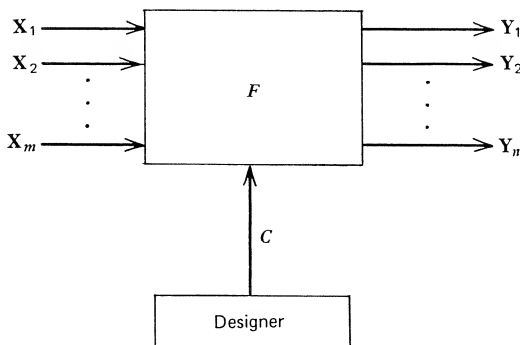
Figure 3.1. The design of a system.

the set of systems demanded of him and the set of drawings on his board. Briefly, he receives demands and emits drawings. His set of " messages sent " is the set of demands that may (professionally) be made on him. His set of " messages received " is the set of circuits that may appear on his board. In this process of translation we can identify, conceptually at least, all the usual features of transmission: entropy, noise, equivocation, error-correcting methods, and so on. Of special interest will be the channel's capacity, for this quantity reflects, for any act of design, how much information processing by the designer is implied by the translation. (Numerical values for the probabilities may well be unavailable, but here again the least safe capacity provides an interesting, perhaps a surprising, insight.)

Such an insight is given if we ask how the capacity required in channel C is related to the capacity in the channel from X to Y. We may start by noticing that the diagram in Figure 3.1 is highly misleading in one important aspect, for it suggests that input and output, X and Y, will enter into the matter symmetrically. This suggestion is grossly wrong. Channel C has as its set of " messages received " the set of *functions* from which is selected the final " message " F, when Y equals the required $F(X)$. If the set of " messages sent " (the set of demands that might be made of the designer) is as large, then the least safe capacity of C is the logarithm of their number. Now the number of functions, from a domain of d elements to a range of r elements, is r^d; and we see at once that the dependences on d and r are very different. The main fact here is that, when both d and r are greater than small integers, d is a far more powerful increaser (of the number of functions, and therefore of the work to be done) than r.

To be definite, suppose that the designer wishes to set up an air-traffic control. Information may come in from, say, 1000 variables; suppose that the instructions go out to as many. Take the simplest case, in which each variable has only two values. Then d and r are both 2^{1000}, and the number of functions from which the channel C must choose is this number raised to its own power; then the logarithm gives the required least safe capacity, which proves to be approximately 10^{300} bits. (Again the quantity has gone far beyond the quantal limits.)

Next suppose, as an interesting comparison, that the output Y, instead of having 2^{1000} possible values, is simplified to the extreme of having only 2: by how much does this lessen the designer's work? We now require the logarithm of $2^{(2^{1000})}$, which is (approximately) 10^{300}, making practically no difference. Thus, when a designer attempts to design (i. e., give a *particular* form to the function F) a system in which all the parts interact fully, complexity at the outputs can often be ignored: it is complexity at the inputs of the system that is to be feared.

3.11. SUMMARY

These examples may help to justify the thesis of this chapter: that when the systemist tackles really complex systems, he may find that he is unwittingly trying to transmit, either through his computer or through his own brain, quantities of information that are vastly beyond anything physically realizable. At the time of writing, our knowledge of these quantities is cripplingly insufficient. Hence one trend of major importance in systems theory over the next decade must be the growing enrichment of our knowledge in this direction.

PROBLEMS

3.1. A neuron is firing at every 100 msec, almost always, but on 10% of the occasions it does not fire. If there are no constraints holding from event to event,

(a) how much information is carried per pulse?

(b) how much is carried per gap?

(c) why is each gap carrying more information than each pulse?

3.2. Of three variables, A has two possible values, B has four, and C has eight. What are the maximal numerical values, in bits, of:

(a) $H(B)$; (b) $H(B, C)$; (c) $H(A, B, C)$; (d) $H_C(A)$; (e) $H_B(A, C)$; (f) $H_{AB}(C)$; (g) $T(B : C)$; (h) $T(A, B : C)$; (i) $T_A(B : C)$; (j) $T(A : B : C)$; (k) (the absolute magnitude of) $Q(A : B : C)$?

3.3. An array of 1000 lamps presents a "picture" by each lamp being either lit or unlit.

(a) How many "pictures" are possible?

(b) If a "pattern" is some subset of the set of all possible pictures, how many "patterns" are possible? (Express the number as 10^x).

If one pattern is to be selected or specified (from the set of all patterns), what is the least safe capacity required (in bits per selection of *pattern*)?

3.4. Over the n variables, $A, B, \ldots, N$, it has been found that

$$T(A : B : \ldots : N) = 0.$$

What are the main consequences in:

(a) the entropies H; (b) the transmissions T; (c) the interactions Q?

3.5. Functions with repeated arguments. Nominally distinct variables may be forced to become equal (as a copper wire between two points forces the potentials at these points to be equal). Simplify the following expressions:

(a) $H(A, A, B, \ldots)$;   (b) $H_A(A)$;   (c) $H_A(A, B, \ldots)$;   (d) $T(A : A)$;

(e) $T(A : A : B : \ldots)$;   (f) $T(A, B : A, C)$;   (g) $T_A(A : B)$;

(h) $T_A(A : B : C : \ldots)$;   (i) $Q(A : A : B)$;   (j) $Q_A(A : B : C : \ldots)$.

3.6. Adjudicate between the two arguments, (a) and (b), given below. They refer to a simple desk calculator that will take two numbers of 8 digits (denary) each and put their product into the main register of 16 digits.

(a) There *is* a loss of information on going from the two factors to the product, for (e.g.) 12 is given by 3×4, 4×3, 6×2, and others.

(b) There is *no* loss, for each wheel in one factor, with 10 positions, can convey $\log_2 10$ bits, that is, 3.32 bits. All 8 convey 8 times this, that is, 26.58 bits, and the two input registers make the total 53.15 bits. But the output register has $16 \times \log_2 10$, which is the same quantity.

3.7. Draw a 27-celled frequency table to represent three variables, A, B, C, each of which can take three values. Insert numbers (frequencies) to demonstrate that it is possible for the case to arise in which $T(A : B)$, $T(B : C)$, and $T(A : C)$ are all zero, but $Q(A : B : C)$ is not zero.

3.8. What single transmission term T^* will make true the identity

$$H(X) + H(Z) = T(X : Y) + T(Y : Z) + H_Y(X, Z) + T^* \, ?$$

3.9. What single term Y will make true the identity

$$T(A : B : C : D : E : F) = T(A : B, C : D, E, F) + T(B : C)$$
$$+ T(D : E) + T(D : F) + T(E : F) + Y \, ?$$

3.10. What single term Y will make true the identity (in $2n$ variables)

$$T(X_1 : \ldots : X_n : X_1' : \ldots : X_n') = T(X_1 : \ldots : X_n)$$
$$+ \sum_i T(X_1, \ldots, X_n : X_i') + Y \, ?$$

3.11. A ballet of 18 persons consists of a principal trio, who dance their steps, and five other trios who simply copy the movements of the principal trio. Considering the 18 variables of position, $X_1, \ldots, X_{18}$, what can be said of the highest-order interaction, $Q(X_1 : \ldots : X_{18})$?

3.12. "Broadcasting" may be regarded as an operation in which one variable, A say, forces others $(B, C, \ldots, N)$ always to take A's value. When this is so, find, in terms of $H(A)$, the values of:

(a) $H(B)$;   (b) $H(A, B, \ldots, N)$;   (c) $T(A : B : \ldots : N)$;   (d) $Q(A : B : C)$;

(e) $Q(A : B : \ldots : N)$.

3.13. Prove the identity

$$T(A : B : \ldots : G) = T(A : B) + T(A, B : C) + T(A, B, C : D)$$
$$+ \cdots + T(A, \ldots, F : G).$$

3.14. With the set $\{A, B, \ldots, G\} = \mathscr{A}$, prove the identity

$$T(A, \ldots, G : Z) = \sum_{I \in \mathscr{A}} T(I : Z) + \sum_{IJ \in \mathscr{A}} Q(I : J : Z)$$
$$+ \sum_{IJK \in \mathscr{A}} Q(I : J : K : Z) + \cdots + Q(A : \ldots : G : Z).$$

3.15. Prove the identity

$$T(A_1, A_2 : B_1, B_2) = T(A_1 : B_1) + T_{A_1}(A_2 : B_1)$$
$$+ T_{B_1}(A_1 : B_2) + T_{A_1 B_1}(A_2 : B_2).$$

3.16. Prove the identity

$$Q(A : B : C : D) = T(A, B : C, D) - T_A(B : D) - T_B(A : C)$$
$$- T_C(A : D) - T_D(B : C).$$

3.17. Prove the identity

$$Q(A, B : Y : Z) = Q(A : B : Y : Z) + Q(A : Y : Z) + Q(B : Y : Z).$$

3.18. Prove the identity

$$Q(A, B, \ldots, F, G : Y : Z) = Q(A : Y : Z) + Q_A(B : Y : Z)$$
$$+ Q_{AB}(C : Y : Z) + \cdots + Q_{A \ldots F}(G : Y : Z).$$

3.19. The error-controlled regulator shown in Figure 3.2, a machine with input, behaves in accordance with the following equations (showing the next states z' and r' in terms of the previous states z and r and the input or signal x):

$$z' = z + r - x, \qquad r' = r - 2z.$$
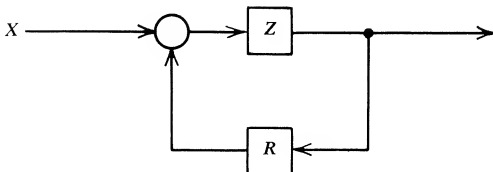
Figure 3.2. The error-controlled regulator (Problem 3.19).

The output (value) of z at successive times was:

Time:	0	1	2	3	4	5	6	7	8	9
z:		1	1	0	-1	2	0	-2	0	0	1

At $t = 0$, the value of R was 2. Deduce the sequence of values that x must have taken (ignoring those at the ends), and so demonstrate that the error-controlled feedback regulator is information conserving.

3.20. The effect, on T and Q, of failing to make distinctions. Variable X is able to take values $x_1, \ldots, x_k$; Y can take values $y_1, \ldots, y_m$. The $k \times m$ frequency table has frequencies n_{ij} for the compound event $\langle x_i, y_j \rangle$, totaling N, and showing transmission $T(X : Y)$. Now let another table be formed from it, by adding the frequencies for y_{m-1} and y_m, as would happen if the observer could not distinguish them. If the new table shows transmission $T(X : Y')$, if X and the set $\{y_{m-1}, y_m\}$ show transmission T^*, and if these last events have frequency N^*, show that

$$T(X : Y') = T(X : Y) - \frac{N^*}{N} T^* \qquad \text{(Conant [15])}$$

3.21. Similarly show that, if there are three variables with y_{m-1} and y_m again merged,

$$Q(W : X : Y') = Q(W : X : Y) - \frac{N^*}{N} Q^* \qquad \text{(Conant [15])}$$

3.22. It is well known that, if X and Y are statistically (probabilistically) independent, the transmission $T(X : Y)$ must be zero. Prove the converse.
(Conant [15])

3.23. If the set $\{A, B, \ldots, G\}$ of arguments of $Q(A : B : \ldots : G)$ can be partitioned into two sets such that all subsets of one are probabilistically independent of all subsets of the other, then Q must be zero. Prove this statement or provide a counterexample.

3.24. If a set $\mathscr{A}$ of variables is such that no subset of $\mathscr{A}$ can be increased in size beyond k variables without including at least one variable that is wholly independent of the others in the subset, then all interactions Q between $k + 1$ or more variables must be zero. Prove this statement or provide a counterexample.

3.25. Each of the four variables L_1, L_2, L_3, and L_4 may take any of the values -2, -1, 0, $+1$, and $+2$. What is the least safe capacity, as

$$T(L_1 : L_2 : L_3 : L_4),$$

if the values are always to be so coordinated as to maintain

$$L_1 + L_2 + L_3 + L_4 = 0 ?$$

3.26. A ganglion of 100 nerve cells provides the right-sized stimulus to a certain muscle when just 15 of the nerve cells fire. (*Which* 15 fire may be varied widely by other factors or conditions.) What is the least safe capacity within the ganglion that can ensure the restriction always to just some 15? (*Hint*: 42.2 bits per event is much too large.)

3.27. An aerial circus team of 12 aircraft goes through maneuvers in which the planes take all places in a $6 \times 6 \times 6$ "cube" (each "cell" of 1-plane size), except that (a) no 2 planes must try to occupy the same "cell," and (b) no 2 planes may be in the same line in the direction of flight (with 1 plane in the other's exhaust stream). What is the least safe capacity (in bits per maneuver) required by this coordination?

3.28. A certain centipede walks on rough ground: for safe ambulation its 100 legs must be coordinated. The rule is that the first 10 legs must move so that always some 5 of them are down and the other 5 up. (The same rule holds over the second 10, and so on.) The environment is likely to demand, at various times, all the permissible variations.

Each leg has only the two positions, up and down. Each "neural module" can transmit only 0.1 bit per stride. The nervous system has no other function and is minimal in size. How many "neural modules" will the histologist find in each ganglion (of the chain of 10)?

3.29. All 20,000 students on a campus want to go to the same vacation resort, some one to be chosen from eight possibilities. If communication between the students costs at least 0.02 cent per bit, what should be budgeted for the cost of the coordination?

3.30. Six supporters of the football club are proposing to signal secret coded instructions to the players on the field by each supporter having two cards, a black (B) and a white (W), and showing, at various times, one of five code signals:

BWB	WBB	WBW	BWW	WBW
WBW	WBB	BWB	WWB	WBW

What is the least safe capacity *between the supporters*?

3.31. A certain species can produce 24 phonemes, by means of a vocal apparatus of four muscles, each phoneme being produced by the successive contractions of some three of the four muscles, in a particular order. Thus, one phoneme is produced by the sequence

$$\begin{array}{c|ccc}
4 & - & + & - \\
\text{Muscle} \quad 3 & - & - & - \\
2 & + & - & - \\
1 & - & - & + \\
\hline
& 1 & 2 & 3
\end{array}$$

<div align="center">Time</div>

with muscle 2 contracting first, then 4, and then 1. Let the 12 variables be represented as $X_i{}^j$, which has the value 1 if muscle i contracted at time j $(i = 1, 2, 3, 4; j = 1, 2, 3)$.

(a) How much transmission $T(X_1{}^1 : \ldots : X_4{}^1 : X_1{}^2 : \ldots : X_4{}^3)$ is the least safe capacity for this coordination?

(b) Analyze this total, algebraically and numerically, into parts that distinguish, in some reasonable way, various types of "conduction" and "memory."

3.32. A system of three variables, A, B, C, has "memory" in the sense that its subsequent values, A', B', C', depend on the earlier ones, so that

$$T(A : B : C : A' : B' : C')$$

is not zero. To what physical effects might the following algebraic forms correspond:

(a) $T(A : A')$;   (b) $T(A, B, C : A')$;   (c) $T(B : A')$;   (d) $T(A' : B' : C')$;
(e) $T(A, B, C : A', B', C')$; (f) $T(A, A' : B, B' : C, C')$; (g) $T_{ABC}(A' : B' : C')$?

3.33. Each fish in a shoal tends to move variously as morsels of food are taken; hence a shoal tends to disperse, somewhat as molecules diffuse. Obtain (at least to an order of magnitude) the quantity of communication that must be used if the shoal is not to disperse.

REFERENCES

1. **Nyquist, H.,** "Regeneration Theory." *Bell System Technical Journal*, Vol. 11, p. 126, January 1932.

2. **Ashby, W. R.,** "Design for a Brain." *Electronic Engineering*, Vol. 20, pp. 379–383, December 1948.

3. **Ashby, W. R.,** *Design for a Brain.* London: Chapman and Hall, 1952; 2nd ed., 1960.

4. **Bourbaki, N.,** *Théorie des Ensembles.* Paris: Hermann & Cie, 1951.

5. **Ashby, W. R.,** "The Set Theory of Mechanism and Homeostasis." In *Automaton Theory and Learning Systems*, D. J. Stewart (Editor), pp. 23–51. London: Academic Press, 1967.

6. **Shannon, C. E.,** "A Mathematical Theory of Communication." *Bell System Technical Journal*, Vol. 27, pp. 379–423, July—October 1948.

7. McGill, W. J., "Multivariate Information Transmission." *Psychometrika*, Vol. 19, pp. 97–116, 1954.

8. Garner, W. R., and W. J. McGill, "The Relation between Information and Variance Analyses." *Psychometrika*, Vol. 21, pp. 219–228, 1956.

9. Garner, W. R., *Uncertainty and Structure as Psychological Concepts.* New York: Wiley, 1962.

10. Ashby, W. R., "Measuring the Internal Informational Exchange in a System." *Cybernetica*, Vol. 8, pp. 5–22, 1965.

11. Ashby, W. R., "Measuring Memory" (in Russian). In *Systems Organisation of Physiological Functions*, V. V. Parin (Editor), pp. 239–243. Moscow, 1969.

12. Ashby, W. R., "Two Tables of Identities Governing Information Flows within Large Systems." *Communications of the American Society of Cybernetics*, Vol. 1, pp. 3–8, July 1969.

13. Bremermann, H. J., "Quantal Noise and Information." In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, pp. 15–20, 1967.

14. Ashby, W. R., "Information Flows within Coordinated Systems." *In Progress of Cybernetics*, J. Rose (Editor), pp. 57–64. New York: Gordon and Breach, 1970.

15. Conant, R. C., "Information Transfer in Complex Systems with Applications to Regulation." *IEEE Transactions on Systems Science and Cybernetics*, Vol. SSC-5, No. 4, pp. 334–338, October 1969.

4. A Computer Approach to General Systems Theory

GERALD M. WEINBERG

School of Advanced Technology, State University of New York at Binghamton

EDITOR'S COMMENTS

Although no preliminaries are essentially required for Chapter 4, readers with a basic knowledge of computers will obtain far more benefit. For this basic knowledge, the following books are recommended:

Leeds, H. D., and **G. M. Weinberg,** *Computer Programming: Based on the IBM System/360.* New York: McGraw-Hill, 1970.
Hellerman, H., *Digital Computer System Principles.* New York: McGraw-Hill, 1967.

References [2] and [4] in this chapter, as well as the book

Boulding, K. E., *The Image,* Ann Arbor: University of Michigan Press, 1968.

are recommended as supplementary readings to Chapter 4.

4.1. THE COMPLEXITY OF THE WORLD

It isn't what we don't know that gives us trouble, it's what we know that ain't so. Will Rogers

Writing about general systems theory is like writing about unicorns. Even after you eliminate all the people who have never heard of them and all the

people who have heard of them but are not interested in the subject, you have to contend with the people who do not believe that they exist.

One reason for believing in general systems theory is that it has a history. There is an organization called the Society for General Systems Research, which was founded in 1956 by some otherwise very respectable men of science. Since that time it has acquired many members, most of them also respectable and most of them probably agreeing that general systems theory exists. About any more than that, however, there is an astounding lack of agreement.

Disagreement among the members of the Society for General Systems Research probably begins—but certainly does not end—over the interpretation of the words "general systems theory." There are two major schools, according to which noun, "systems" or "theory," the adjective "general" is taken to modify. With this chapter, a new school comes to the surface, one in which the adjective applies as well to the audience for whom the theory is presented.

General systems theory is an attempt to aid the human mind in dealing with a world which is too complex for that mind. It is true, as Robert Louis Stevenson once said, that

> The world is so full of a number of things,
> I'm sure we should all be as happy as kings.

On the other hand, James Thurber, who—in spite of his writing about unicorns—was much more of a realist than Robert Louis Stevenson, brought these immortal lines out of the world of children by adding, "And we all know how happy kings are." How much we appreciate Thurber's sentiment when we are faced with the complexities of living our lives!

Yet in spite of this complexity, we manage to go on living, perhaps with a bit more happiness than the average king. Moreover, we have managed to find some spare time along the twisting road of history to build an intellectual edifice which sparkles as a beacon of simple beauty amidst the fog of everyday affairs. That edifice is science. General systems theory is seen by some as a threat to that edifice, a regression to dark ages of mysticism and vitalism. Before we can dispel these fears, we shall have to consider how the edifice of science has been constructed, and why it has been so successful in bringing order to our chaotic world.

4.2. MECHANICS, MECHANISMS, AND COMPUTING MACHINES

In order to understand the successes of science, we can do no better than to examine physics—and particularly mechanics—for these sciences are often taken to be ideal models. The beauty of the mechanical model of the world was well expressed by Deutsch [1], who said that mechanism

. . . implied the notion of a whole which was completely equal to the sum of its parts; which could be run in reverse; and which would behave in exactly identical fashion no matter how often these parts were disassembled and put together again, and irrespective of the sequence in which the disassembling or reassembling would take place. It implied consequently that the parts were never significantly modified by each other, nor by their own past, and that each part once placed in its appropriate position with its appropriate momentum, would stay exactly there and continue to fulfill its completely and uniquely determined function.

Yet the beauty of this system is a bit dulled when we realize that mechanical systems ordinarily have only a handful of identifiable parts—most often two, but sometimes ten or perhaps even thirty or forty if they are highly constrained, as are the parts of a bridge. For if there are too many parts, we may write down equations which relate the behaviors of the different parts of the system, but we cannot solve the equations, even by approximate methods. High-speed computers have extended the range of mechanical systems whose equations can be solved (approximately), but only by a relatively small amount.

If the formal methods of mechanics are so limited, why is mechanics considered to be a model for the sciences? We must—if we are to have the answer—consider not the formal methods but the informal ones by which complex mechanical systems are reduced to simpler ones, for only then can they be subjected to the working of the formal methods. Consider, for example, Newton's achievement in explaining the motions of the bodies in the solar system. Rapoport [2], in speaking about this problem, pointed out, "Fortunately for the success of the mechanistic method, the solar system . . . constituted a special tractable case of several bodies in motion."

Although Rapoport's analysis is correct and pertinent as far as it goes, it does not penetrate deeply enough into the heart of Newton's success. The solar system, in the first place, does not consist of "several bodies in motion." We now know that there are thousands upon thousands of celestial bodies in our solar system plus other matter not in "bodies." Any analysis of planetary motions, however, begins by ignoring most of these bodies because they are "too small" to have a significant effect on the calculations. Although this seems a natural step—so natural that textbooks on mechanics do not ordinarily mention it—it is a step which happens to work only in very special circumstances. Yet any other circumstances are not considered proper systems for mechanists to think about.

Consider, for instance, the pineal body, a tiny piece of tissue in the brain. Can physiologists ignore this in their attempts to understand the behavior of the human body? Perhaps they can—the question is quite alive—and perhaps they cannot; but in any case no physiologist would think of arguing that, because the mass of the pineal body is small with respect to the mass of

the brain, it can be ignored on that account. The DNA in a living cell is a miniscule amount of the cell material, if measured according to mass; but understanding of cellular biology would be impossible without considering its role. The queen bee in a hive is only one of thousands of bees, and constitutes only a small fraction of the total mass of the hive, but no ethologist dare ignore her role in hive behavior.

Mechanics, then, is the study of those systems for which the approximations of mechanics work successfully. It is strictly a matter of empirical evidence, not of theory, that the human body cannot be understood by considering only the gravitational attractions between its parts.

But how important is the step of ignoring "small" bodies—the asteroids, comets, satellites, and other pieces of space flotsam—to the calculation of planetary orbits? Consider the equations describing a system with only two objects. We must describe how each object behaves by itself, which gives us two equations, or relationships, one for each object. We must consider how the behavior of each body is related to that of the other, which gives us another relationship. Finally, we must consider how things will behave if neither of the bodies is present (the "field" equation), making four relationships in all. In general, if we have n bodies, the number of relationships is 2^n. For 10 bodies, this means about 1000 relationships ($2^{10} = 1024$), and for 100,000 bodies, about $10^{30,000}$. Thus, by "ignoring small masses," we reduce the number of relationships from perhaps $10^{30,000}$ to approximately 1000, for 10 bodies. One thousand equations we might conceivably write down, even if we could not solve them.

Now, we have just promised to speak to general audiences, and here we find ourselves talking about solving a system of 1000 equations. But our readers do not have to know how to solve such a system of equations; they must understand only how much effort is involved.

Why are we interested in the amount of effort required to solve equations? In Newton's day, the impact of mechanics on thought was strong. Many philosophers believed, with Laplace, that, given precise observations on the position and velocity of every particle in the universe, one could calculate the entire future of the universe. They realized, of course, that they would need a large computing machine. But they lacked even the simplest computers. How could they possibly put a measure on the required computation?

Only in our lifetime have real computers come into existence, and with them philosophical thought has undergone a revolution. Anyone who doubts that there has been a revolution should read the debate on teleology between Rosenbluth, Wiener, and Bigelow, on the one hand, and Richard Taylor, on the other. (The four articles of this debate may be found in Part V.A. of Buckley [3].) Although this debate took place after World War II, Taylor, the philosopher, was still able to advance arguments which today could be

refuted by any undergraduate student in computer science. One aspect of this revolution, of course, was the concern with quantifying computational complexity and power.

Even now the question of quantifying complexity is not a closed one, but for our arguments we do not need exact measures. Instead, we merely want to estimate how the amount of computation increases as the size of the problem increases. Experience has shown that, unless some simplifications can be made, the amount of computation involved increases at least as fast as the square of the number of equations. Thus, if we double the number of equations, we will have to find a computer four times as powerful to solve them in the same amount of time. Naturally, the time often goes up faster than this, particularly if some technical difficulty arises, such as a decrease in the precision of results. For our present arguments, however, we may conservatively use the square law of computation to estimate how much more computing is required for one set of equations than for another.

In practice, then, there is an upper limit to the size of the system of equations which can be solved. Clearly, $10^{30,000}$ equations are far beyond that limit. And in Newton's day, without computers at all, the practical limit of computations was well below 1000 second-order differential equations, especially since Newton had just invented differential equations. Newton needed all the simplifying assumptions—explicit or implicit—he could get away with, just as do physiologists and psychologists today. We may note, in this regard, that old-time physicists now say that the "youngsters" no longer do "real physics." because they use the computer to solve large sets of equations, rather than applying physical "intuition" to reduce the equations to a form that can be solved with a pencil on the back of the proverbial envelope.

4.3. THE SIMPLIFICATION OF SCIENCE AND THE SCIENCE OF SIMPLIFICATION

I do not know how it is with you, but for myself I generally give up at the outset. The simplest problems which come up from day to day seem to me quite unanswerable as soon as I try to get below the surface.   Justice Learned Hand

Thinking about the practical problem of computation, then, can give us a new point of view about what mechanics is, or what any science is. Practical computation demands that implicit assumptions be brought out into the open; hence it is no coincidence that computer programmers are attracted to a systems theory which devotes itself to studying how people make assumptions. An excellent example of the kind of experience that computing people have relates to another assumption already made in our reduction of the solar system problem to 1000 equations.

We have assumed, as one always assumes in mechanics, that only certain kinds of interactions are important. In this case, the only important interaction was gravitational, which meant that each relationship gave only one equation. How do we know that only gravitational attraction is important in this system? How do we know that we can ignore magnetic effects, electrostatic forces, light pressure, force of personality, and so forth? One answer is that this would not be a problem in mechanics if those other forces were important, but that is merely begging the question. How do we know that it is a problem in mechanics?

As before, we know that it is a problem in mechanics because when we try these approximations they give us satisfactory answers—that is, answers which match observational data. If we had a problem for which they did not work, it would never make its way into the mechanics textbooks. Our practical computing example of this quandary is the calculations that were made of the orbit of the Echo satellite, which was a large, inflated Mylar sphere. After a few months, it was found that the classical solution of the gravitational equations was not doing a satisfactory job of predicting Echo's orbit. After much perplexing labor, the programmers realized that Echo, because of its small density, was much larger than any "normal" solar body of the same mass. Consequently, the pressure of the sun's light radiating on its surface could not be implicitly ignored, as it is in all "ordinary" orbital calculations. No, mechanics does not tell us which systems are "mechanical."

And yet, even having reduced—by applying deeply buried assumptions—the number of equations to 1000, we still may not be able to say that we have solved a particular mechanical system. The equations may still prove intractable, even for a large computer. We need further simplifications. Newton supplied an important one in his law of universal gravitation, which stated that the force of attraction (F) between two (point) masses is given by the equation

$$F = \frac{Gm_1m_2}{r_{12}^2}$$

where m_1 is the mass of the first body, m_2 is the mass of the second, r_{12} is the distance between them, and G is a universal constant. From the viewpoint of simplification, this equation says more implicitly than explicitly, for it states that no other equation is needed. It says, for instance, that the force of attraction between two bodies is in no way dependent on the presence of a third body, so that only pairs of bodies need be considered in turn and then all of their effects may be added up.

A psychologist, for one, would be tickled pink if he could consider only summed pair interactions. This simplification would mean that, to understand the behavior of a family of three, he would study the behavior of the father and mother together, the father and son together, and the mother and son

together. When all three got together, their behavior could be predicted by summing their pairwise behaviors. Unfortunately, it is only in mechanics and a few other sciences that superposition of pairwise interactions can be successful.

In the case of the solar system, pairwise superposition reduces 1000 equations to about 45—that being the number of ways in which 10 things can be taken in pairs. From a computational point of view, we have reduced the size of our task by the square of 1000/45—or about 400 times, at least. We might be willing to stop at this point, although Newton (perhaps because he lacked the computers that we have) went still further.

As it happens, the solar system has one body (the sun) whose mass is much larger than any of the other masses—larger, in fact, than the mass of all of the other bodies together. Because of this dominant mass, the pair equations not involving the sun's mass yield forces small enough to be ignored, at least considering the accuracy of the data Newton was trying to explain. (Discrepancies in this assumption led to the discovery of at least one planet that Newton did not know.) This simplification, which is made possible by the solar system, rather than by mechanics, reduces the number of equations to about 10, instead of 45—giving an estimated 20 times reduction in computation.

But Newton went even further, for he observed that the dominant mass of the sun enabled him to consider each planet together with the sun as a separate system from each of the others. Such a separation of a system into noninteracting subsystems is an extremely important technique known to all developed sciences—and to systems theorists as well. To understand the power of such a separation, we need only recall the square law of computation. If solving a system of n equations takes n^2 units of computation, n separate single equations taken one at a time will require only n of the same units.

At this point, Newton stopped simplifying and solved the equations analytically. He had actually made quite a few other simplifications, such as his consideration of each of the solar bodies as a point mass. In each of these cases, he and his contemporaries were generally more aware of—and more concerned about—the simplifying assumptions than are many present-day physics professors who lecture about Newton's calculations. Consequently, students find it hard to understand why Newton's calculations of planetary orbits is ranked as one of the highest achievements of the human mind.

But the general systems theorists understands. He understands because it is his chosen task to understand the simplifying assumptions of a science—those assumptions which delimit its field of application and magnify its power of prediction. He wants to go right to the beginning of the process by which a scientist forms his models of the world, and to follow that process just as far as it will help him in suggesting useful models for other sciences.

And why is the general systems theorist interested in the simplifications of science—in the science of simplifications? For exactly the same reason as Newton. The systems theorist knows that the square law of computation puts a limit on the power of any computing device, and he believes that the human brain is in some sense a computing device. Thus he knows that, if we are to survive in this complex world, we shall need all the help we can get. Newton was a genius, but not because of the superior computational power of his brain. Newton's genius was, on the contrary, his ability to simplify, idealize, and streamline the world so that it became, in some measure, tractable to the brains of perfectly ordinary men. By studying the methods of simplification which have succeeded and failed in the past, the general systems theorist hopes to make the progress of human knowledge a little less dependent on genius.

4.4. THE COMPUTER AS INTELLIGENCE AMPLIFIER

This new invention of printing has produced various effects of which Your Holiness cannot be ignorant. If it has restored books and learning, it has also been the occasion of those sects and schisms which daily appear. Men begin to call in question the present faith and tenets of the Church; the laity read the Scriptures and pray in their vulgar tongue. Were this suffered the common people might come to believe that there was not so much use of the clergy. If men were persuaded that they could make their own way to God, and in their ordinary language as well as Latin, the authority of the Mass would fall, which would be very prejudicious to our ecclesiastical orders. The mysteries of religion must be kept in the hands of the priests.
Thomas Wolsey

Except for a very few scientists, the mysteries of mathematics are today kept in the hands of the mathematicians. Mathematical reasoning is especially often a roadblock to thinkers who wish to pursue what Anatol Rapoport is fond of calling the " hard sciences," as opposed to the "easy sciences" of physics and chemistry. What a shame that the sciences which are hardest tend to be peopled by workers whose mathematical talents are weakest! But while we wait for a reform in our educational system to remedy this discrepancy, cannot something be done by the general systems theorists to help out? One possibility is to teach mathematics to adults, but this task is known to be well-nigh impossible. Besides, the " soft scientists " who take time out to learn mathematics usually discover that even the deepest mathematics is inadequate to their problems—because their science is so hard.

Another approach is to use the digital computer as the " vulgar tongue " —as a form of what Ashby calls " intelligence amplifier." This is another approach of our school of general systems theory. Not only is the computer our model against which we can study the processes of science, but also it is for us a conceptual model in its own right which can serve to clarify concepts

that are used loosely in many different fields. For an extended example of this kind of approach, we may turn to a simple system which Ashby proposed as a pedagogic device and see what kinds of insights we can obtain when we carry the analysis well beyond Ashby's purposes. By using a very simple computer implementation of this very simple system, we have been able to throw light on such diverse and important systems theory concepts as state spaces, closed versus open systems, stability, regulation, adaption, and particularly the role of the observer in all of these matters. These examples, then, should not be taken as the final or rigorous words on these subjects, but only as instances of how explicit algorithmic models can provide the basis for a kind of thinking of which even social scientists are capable.

State spaces. Imagine a machine which has the ability to store 100 digits, with each digit ranging from 0 through 9. The state of this system can obviously be represented by writing down a particular 100-digit decimal number; thus there are as many possible states as there are 100 digit numbers, or 10^{100} states. One of the nice features of this machine as a conceptual device is the ease with which we have generated a vast number of states.

For identification purposes, we may give each of the digits an index number, from 1 to 100. Thus the first digit is (symbolically) d_1 ; the second, d_2 ; and so forth. The behavior of the machine is determined by choosing two of its digits—say, d_i and d_j—multiplying them together, taking the last digit of the product, and using it to replace d_j. Thus, if the two digits chosen are numbers 28 and 35, as shown in Figure 4.1, $d_{28} = 3$, $d_{35} = 7$, 3 times 7 equals 21, and the low-order digit is 1; hence d_{35} takes on the value 1. The new state then differs from the previous one in just this single position, d_{35}. As soon as a new pair of digits is chosen, however, the state is changed again. (In Figure 4.1 the chosen digits are 38 and 30, yielding the new state as shown.) Thus, given enough time, the state of the system may differ from the starting state in any number of positions.

The pair of index numbers which select the two digits to be multiplied may be thought of as the inputs to the system. Given any initial state and any sequence of input pairs, we can determine precisely what the final state will be. Thus, although the system itself is not a state-determined system—since it is not closed—its behavior is perfectly determined once its initial state and input sequence are known. Because of this property, we can easily simulate this system on a digital computer, that is, we can make the computer behave as if it were this system. The computer stores 100 digits which record the current state of the system, and the computer program causes successive pairs of input numbers to be brought into the machine and used as indices to update the state of the system.
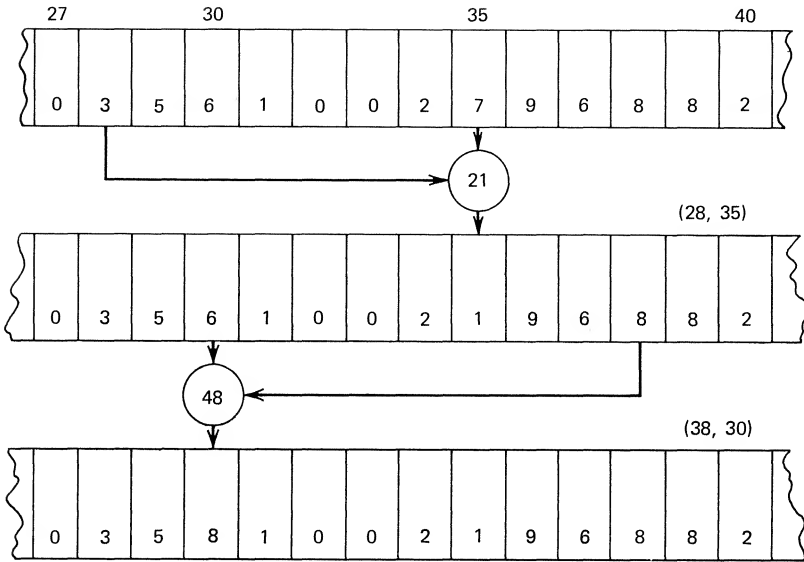
Figure 4.1. The structure of a simple machine.

Figure 4.2 shows a detailed output from a computer program which simulates this system for teaching purposes. Notice that the number of digits has in this case been limited to 40, using a parameter in the program which permits examination of the effect of state-space size on behavior. Using this detailed mode, the student can study the behavior of the system with as much care as he likes, until he fully understands its workings. Eventually, however, he comes to realize that this much detail is too much detail, and he begins to explore various points of view—various representations—which simplify the state space of the system. He has learned about kinematic graphs as representations of small discrete systems, but for 10^{100} states it is impossible for him to use this approach. He is thus prepared for the following types of arguments.

Consider first the state description, the set of 100 digits. If we had a much simpler system with only 2 digits, we could represent each state by a pair of numbers. On a plane, or two-dimensional surface, every point can be specified uniquely by a pair of numbers—in fact, that is what we mean by a two-dimensional surface. There is, of course, an infinity of ways of numbering the points in a plane. We can specify the location of a house in a city by giving

```
 I   J   |<---STATE----------------------------->|
         5955365408195661806513766175920539182950
15  39                  ∧                       ↓
         5955365408195661806513766175920539182900
15  20                  ∧   ↓
         5955365408195661806060137661759205391 82900
 6  11           ∧   ↓
         5955365408695661806013766175920539182900
21  24                       ∧  ↓
         5955365408695661806013766175920539182900
35  13                  ↓                    ∧
         5955365408695661806013766175920539182900
39   1   ↓                                   ∧
         0955365408695661806013766175920539182900
35   9                  ↓                    ∧
         0955365408695661806013766175920539182900
37  11                ↓                       ∧
         0955365408295661806013766175920539182900
39  10                ↓                       ∧
         0955365400295661806013766175920539182900
37   5                ↓                       ∧
         0955665400295661806013766175920539182900
15  21                  ∧       ↓
         0955665400295661806063766175920539182900
 1  17   ∧                   ↓
         0955665400295661006063766175920539182900
14  12           ↓  ∧
         0955665400245661006063766175920539182900
24  34                         ∧        ↓
         0955665400245661006063766175920534182900
38  38                                        ↓
         0955665400245661006063766175920534182100
35   5             ↓                        ∧
         0955665400245661006063766175920534182100
37  21                        ↓             ∧
         0955665400245661006023766175920534182100
35   6             ↓                        ∧
         0955665400245661006023766175920534182100
 6  26           ∧                       ↓
         0955665400245661006023766675920534182100
 6  29           ∧                          ↓
         0955665400245661006023766675420534182100
```

Figure 4.2. Sample output from computer simulation showing inputs and states for 20 cycles.

the house number and the street or by giving the two cross streets. Either system will do, but we can choose a standard one by common agreement. In a similar way, we could set up a correspondence between the points on a plane and the states of a two-variable system. If there are very many sub-states for each dimension, some orderly arrangement of the kinematic graph may appear to yield a continuous line for the behavior of the system. We must remember, however, that, since the assignment of states to points is quite arbitrary, the appearance of continuity may be entirely fortuitous. There is no a priori reason why the behavior of a system should be represented by a continuous line in this manner.

In the case where the substates do indeed range continuously through their possible values, the line representing the behavior of the system—the passage from state to state—will be a continuous one. This representation may be imagined to be a continuous form of the kinematic graph, with the arrows from state to state being infinitesimally small. In the case of our simple system, however, the substates would merely be 10 discrete points, or 100 points in all for a two-digit version. The lines of behavior would not be continuous at all, but would jump back and forth to represent the discrete changes in digit values. Thus we see that the representation of a system's behavior by a continuous line in this manner is strictly a special case, albeit an extremely important one for mathematicians.

Just as we can represent a two-variable system's behavior by a line in two-dimensional space, the behavior of a three-variable system can be represented by a line in three dimensions. Beyond three dimensions we cannot go with actual lines; but since we use this method of representation to help us think about systems, we can imagine such diagrams drawn in more than three dimensions, even if we cannot visualize them. An n-dimensional space, in which each "point" represents a state of a system, is called a *state space* of the system; and any smaller space which represents some subset of the variables is termed a *subspace* or *projection*.

Thus, if some of the variables do not change, or are not important, at a particular time, the signficant part of the behavior of the system can be represented in a subspace of the total state space of the system. For instance, we may represent the behavior of an airplane by a projection of its three-dimensional flight path onto a two-dimensional map of the surface of the earth. The route we describe in this way can be thought of as the path of the shadow of the airplane with the sun directly overhead, hence the use of the term "projection" for such a description. Naturally, many different flight paths will have the same shadow, or projection, so that some information about the flight—namely, the altitude—is lost in this representation. But projection is a useful representation in cases where the altitude is of no interest.

The state-space concept has the further advantage that we can expand a state space if we wish to expand our viewpoint of the system to include new variables. We merely add one new dimension for each new variable. Now the original state space becomes a projection of the expanded one, so that all our previous observations retain a meaningful interpretation.

One reason why we might wish to expand our view of a system is that the line of behavior of a continuous "closed" system crosses itself in our state space. Since we would like to have a state-determined system, as nearly as as we can, we would not like to have such crossings, for a crossing represents two different paths emanating from the same point, or two different successors to the same state. A continuous system whose line of behavior intersects itself cannot, therefore, be state determined. In a projection, on the other hand, crossings do not imply a lack of state determinedness. An airplane spiraling down for a landing may never return to the same spot in the air, but its shadow may cross and recross itself. Thus, the line of behavior in one representation may suggest to us that there are additional important dimensions to the state description.

To represent our 100-digit system directly would require a 100-dimensional state space. We may reduce the complexity of this representation in a number of ways, including a projection onto some subspace, which in this case would mean singling out a few of the digit positions for special attention. (We did this implicitly in Figure 4.1.) Although such a projection reduces the amount of information that we have to consider, it also tends to make the behavior quite mysterious, whenever digits outside the projection are involved. Many other transformations of our point of view suggest themselves. For instance, we might divide up the digits into batches and characterize each batch by the sum of the digits in it. To get a two-dimensional system, we might divide the system into two parts of 50 digits each. Figure 4.3 shows a typical trajectory of the system through this state space.

Chronological graphs. Although the state space is a useful concept for talking about systems, it does not meet our needs for representing the behavior of systems whose descriptions have more than a tiny number of variables. The use of projections can help us in handling these larger systems, but the state-space representation has a further disadvantage which is independent of the number of variables. Although time is often of central importance, the line of behavior in state space does not show us how fast the system is moving from state to state. To extend a previous example, the map of air routes does not distinguish between jets and propeller planes which travel the same routes.

One way of representing rates of behavior is to add another dimension to any state space—the *dimension of time*. Because time—for all systems—moves
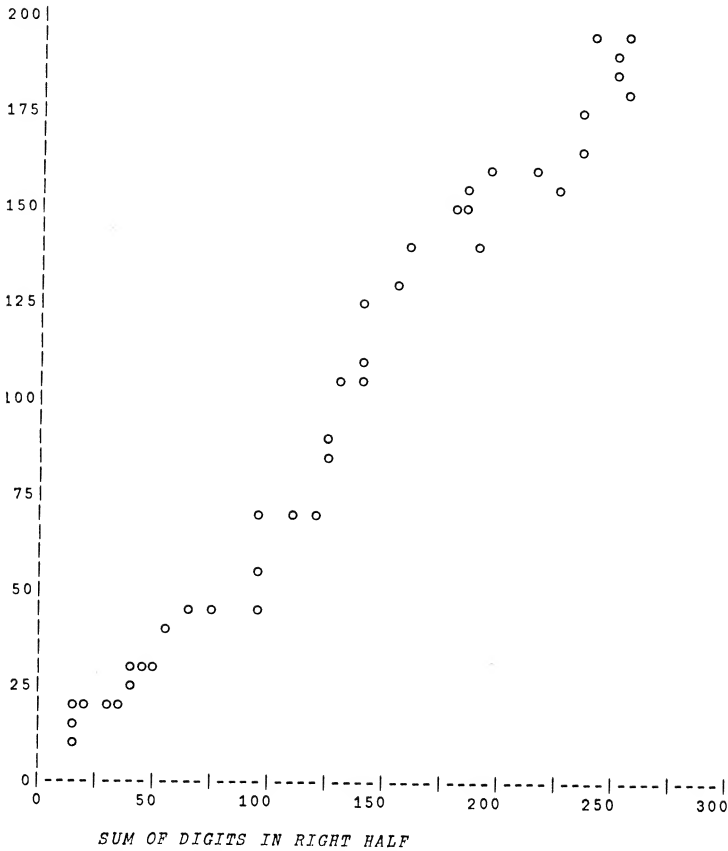
Figure 4.3. Trajectory through left-right state space.

on in one direction at a rate of 1 second per second, it has a way of unraveling twisted lines of behavior. More precisely, if time is one of the state variables, the system can never be in the same state twice, since time moves in one direction only. Thus lines of behavior in the time-expanded state space can never cross. Cyclic behaviors, for instance, are no longer represented by the same states being traversed repeatedly, but are represented by similar states being traversed at different times.

Adding the time dimension not only solves the problem of representing rates of behavior, but also enables us to use a simple technique for representing mutivariable systems. If we take a projection of the state space into two

dimensions, one of which is time, we get what is called a *chronological graph* of the other dimension. This chronological graph is the trace of one of the system's variables as it changes through time. Consequently, by using one graph for each variable, we can represent the total behavior of the system from a particular point of view. Figure 4.4 shows the two chronological graphs necessary to represent the behavior shown in Figure 4.3.

Chronological graphs are familiar to us in many forms, for they are the most effective practical way of reducing complex behavior to tractable representation. The electroencephalograph (EEG) is a chronological graph of the
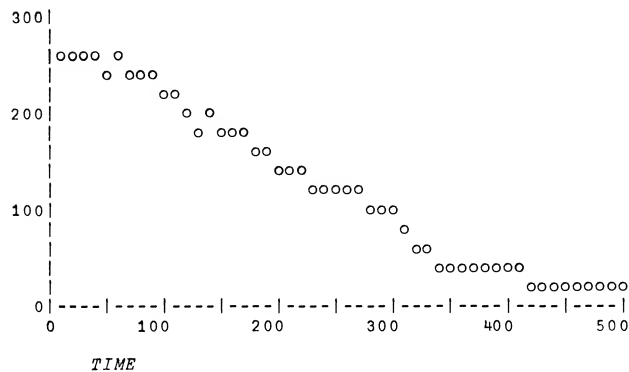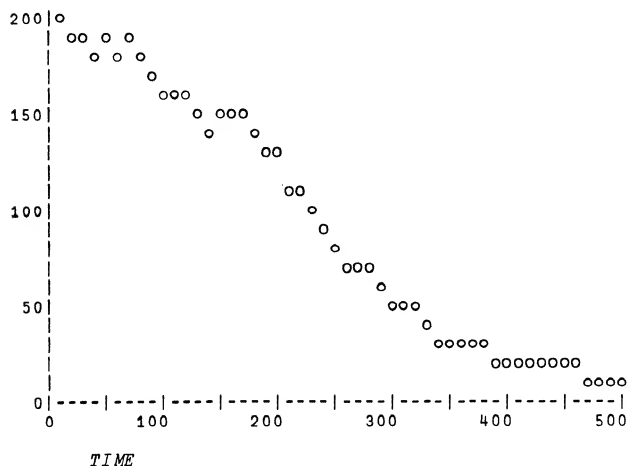
Figure 4.4. Chronological graphs in left-right state space.

electrical potential at a particular point in the brain of a living animal. A series of EEGs taken at different points in the brain gives us a more complete representation of this complex system than any single one provides, but we may still examine the individual EEGs with profit. Business indicators are often presented in terms of chronological graphs—stock market index, levels of retail inventories, wholesale price index, gross national product, and so forth—which managements use to obtain a more comprehensive picture of the economic system which is the environment of their enterprises. In turn, each business itself may be pictured by its management largely through chronological graphs of sales, inventories, production, and costs. To understand the complex system which generates our weather, we use chronological graphs of wind velocity, temperature, barometric pressure, and tides. The list could be expanded endlessly.

Like any other representation, chronological graphs embody certain distortions. For instance, we may represent different systems variables on different time scales—because they have widely varying rates of change—and then overlook the relationship among some of the variables. Even if the time scales are the same, however, the chronological graph forces us to separate the representations of the different variables. In doing so, it may mislead us into unjustified feelings about the independence of these variables.

Points of view. A set of variables is said to be *independent* if the behavior of any one cannot be predicted in any manner from the behavior of any other or the behavior of all of the others. To be independent, then, a variable would have to represent a closed subsystem, so that truly independent variables do not exist. Nevertheless, systems theorists always strive for the ideal of independence in choosing their representations for systems, for independent variables may be studied singly without loss of precision. Where there are dependencies, on the other hand, it may not be necessary to study all variables in order to reach a certain degree of precision in prediction. From a system such as the animal brain, thousands of different EEGs may be taken. In striving for a workable viewpoint of the brain, we should like to eliminate all but a few of these, from which the others, if necessary, could be inferred.

On the other hand, in reducing the number of variables we may be throwing away information which we cannot reconstruct from the variables we retain. The choice of system variables must always be a compromise between the convenience of independence and the necessity for completeness. Take our simple system as a sample. Our first view is certainly complete—we defined it that way—but it is difficult to use for extracting any patterns of behavior. Our division of the system into two halves and summing the digits of the halves reduced the complexity of the system and permitted us to see a certain trend in its behavior. On the other hand, we may have thrown away too much

in this transformation. Whether we did or not depends, ultimately, on what we want to know about the system.

Many other views might have been chosen. As a final illustration, suppose that we had decided on a set of ten variables, each variable being the number of a particular digit found in the state of the system. For instance, in the initial state of Figure 4.2 there are four zeros, five 1's, two 2's, three 3's, one 4, nine 5's, six 6's, two 7's, three 8's, and five 9's. This state would be represented as

$$(4, 5, 2, 3, 1, 9, 6, 2, 3, 5)$$

With this viewpoint, a rather different picture of system behavior emerges. Although we cannot represent the ten-dimensional state space, we can represent the entire view as a series of ten chronological graphs, as shown in Figure 4.5. Actually, these ten variables do not form an independent set, for we observe that once nine of them have been determined the tenth is determined uniquely.

If we had not synthesized this viewpoint, but rather had derived it from some higher-level observations of the systems, we would have discovered the nonindependence of the variables and made it into an inductive law of behavior for this system. We might have noticed that whatever else they tended to do, as one variable went up, the others went down. Eventually, we might have given the law a precise form, the form of a conservation law:

The sum of the variables of this system is a constant (100).

If we are inclined to put names to things, as some scientists are wont to do, we might call our variables a measure of "endigitry." Then the law could be stated in more elegant form:

Endigitry can neither be created nor destroyed.

Note well, however, that this law was discovered by choosing a particular way of looking at the system. The law did not even exist in our other representations; and though it might have been derived from the detailed view of the system shown in Figure 4.2, the "experimenter" would not need to know that much detail to discover it.

Other representations might yield other "laws." For instance, if the system had been reduced to two variables, by counting the even digits and the odd digits the experimenter could discover the law:

Evenness can never decrease.

This law is also available in our ten-variable representation, but it is not as easy to discover, for the experimenter must learn which state variables represent "evenness." When he finally discovers that variables 1, 3, 5, 7, and 9

should be added together, he will probably publish a paper and establish a world-wide reputation.

In effect, much of the progress of science comes in just this way—by learning which ways of looking at things yield invariant laws. Thus the laws of science can be thought of as a short description of the world, or else as prescription for how to look at the world. We really have no way of knowing for sure whether the law is about the world or about our way of looking at it.

4.5. STABILITY AND INVARIANCE

Nothing is permanent except change.   Heraclitus

The chronological graphs of Figure 4.5 show how the system behaves from a particular point of view, with a particular time scale and a particular sequence of inputs, and starting from a particular state. If the system were closed and state determined, the only way in which we could see different behaviors would be by coming upon the system when it was in a different state from one we had previously observed. Since a behavior of a closed, state-determined system is a single line in its state space, the only way to see different behaviors for such a system is by looking at different segments of that line.

But our simple system is an *open system*, not a closed one. Instead of its state at any instant being governed by the relationship

$$S_{t+1} = F(S_t),$$

it is governed by the more complex relationship

$$S_{t+1} = (S_t, I_t).$$

In words, the state at time $t + 1$ is a function of the state at time t and the input (I) at time t. Therefore what we see will depend on the input sequence which happens to be going on during our period of observation. We might, then, expect to see many different behaviors from the same system.

Behavior. A confusion about the term " behavior " exists in common speech and is carried over into some writing about systems theory. The confusion arises because " behavior " is used in two senses—for a particular action and for a set of actions which in some way characterizes the system under discussion.

For example, when Father interrupts Johnny's activites of finger painting on the living room wall and says, " I don't like your behavior, young man," he is using the word in its first—more specific—meaning. When Johnny's teacher gives him an F in Behavior, she is using the second point of view.

```
15|
  |
  |
  |
10|99 9999999988
  |8 9          9
  | 888      88 989
  |      888      8999888
  |   88        888   8888
 5|                    8888
  |
  |          99999988888    8
  |              9  88888 8888
  |              999999      888888888888
 0|----|----|----|----|----|----|----|---99999999999999999998888
  0        100      200      300      400      500      600
```

(a)

```
15|
  |
  |
  |
  |    7
10|88677
  |    6 6
  |  6 87       66666666
  |     6666  6        6
  |     7   66         6
 5|       77
  |         777777    6666666666
  |           77777       6666
  |             7            6
  |          7777 7777777777778777777777777777
 0|----|----|----|----|----|----|----|----|66666666666666666666668
  0        100      200      300      400      500      600
```

(b)

```
15|
  |
  |
  |   2
10|32   22 22
  |2  2  2  222
  |          22 2
  | 33     3332 2
  |  33    3  33 22
 5|    33333    333333332 22222
  |             333   2
  |             33   2222
  |              33333333333
  |               22222333
 0|----|----|----|----|----|----|----|----|33333333333333333
  0        100      200      300      400      500      600
```

(c)

```
15|5   4
  |  5    4
  |     4    55
  |  4 5 554
  |4    5 4  45
10|          4 4 4  4  44
  |            5  4   4
  |          5     4   4444
  |                44      4
  |          555         44
 5|            5           44
  |          55555555    44
  |               555555 55
  |                448885585
  |                  44  4888 4444444
 0|----|----|----|----|----|----|----|----|----|-555 5588 88888 8
  0        100       200       300       400       500       600
```

(d)

```
100|                                        00000000
   |                                   0000000000
   |                                 0000
   |                               0000
   |                             00
 75|                           000
   |                         000
   |                       0 0
   |                     0
   |                   0000
 50|                 00
   |               000
   |             0
   |           00
   |         00
 25|       00
   |     00
   |  ∪00
   |00 11
   |  1  11111111
  0|----|----|---11111111111111111111111111111111111111111111111
   0        100       200       300       400       500       600
```

(e)

Figure 4.5. Chronological graphs of counts of digits.

Nevertheless, the method whereby she represents a collection of individual behaviors by a single evaluation of Behavior is worth considering. Suppose that Johnny is an ideal pupil all semester (meaning that he does not cause any trouble for the teachers), but on the last day of school he sets fire to the library. He will get an F in Behavior.

In American society, we often characterize Behavior according to a single instance of behavior. In common speech, for example, we apply the name "liar" to anyone who tells a single lie; but we do not even have a word for

someone who invariably tells the truth. A multitude of other words—all carrying the implication of badness—are also applied for only a single act of the Behavior denoted: murderer, embezzler, loser, cheater, adulterer, and so on. In a way, the existence and use of such words indicate a peculiar value-orientation of our society, one which is often cruel and inappropriate, but one which, in systems design, has a special meaning. What would we think of an engineer who had just designed a bridge which was "safe" because "it won't fall down more than once in five years"?

In other words, whatever else may interest us in the Behavior of a system, we usually want to know about the chances of its ever displaying a previously unobserved and sufficiently disastrous behavior. No doubt we could build bridges for a fraction of the current cost if we were willing to let them fall down every few years, but usually a major share of the cost of a system is devoted to making the chances of such catastrophe sufficiently low.

Because of our fear of the unexpected, we usually observe a system for a period of time before we are willing to make statements about its complete repertoire of behaviors. The time we take to make these observations depends on a number of things, but particularly on our expectations, which are based on experience with similar systems. If we were to receive a box in the mail which emitted a loud, ticking sound, we might suspect, no matter how long we observed the ticking, that some unique form of behavior would be displayed at any moment. If the box contained a bomb, no amount of observation of its behavior would suffice to tell us when and whether it would explode, for that is precisely the idea behind a time bomb—namely, that its recipient have no idea of its capability for discontinuous behavior.

We are not, in life, constrained to passive observation of the behavior of systems. Although we cannot tell whether a ticking package contains a time bomb, we can behave in such a way that the odious consequences of it being a bomb are removed or diminished. We could, to name one possibility, immerse the package in a pail of water. This action works because a bomb is to some extent an open system; it can exhibit one of several behaviors, depending on its environment.

It is paradoxical that the openess of systems both prevents us from predicting their behavior with certainty and permits us to impose a degree of certainty on their behavior. We can better understand this paradox when we realize that there are two possible reasons for an open system displaying different behaviors—its initial state and its input sequence. Let us look once again to our simple system to illustrate the difference.

Effect of the initial state. Consider first the effect of initial state. Given a random input and almost any starting state, the behavior of the system will be much like that shown in Figure 4.5. Figure 4.6 indicates how the system

```
100|                                                    0000000000000
   |                                        0000000000000000
   |                                    000
   |                                 0000
   |                               00
 75|                            000
   |                          00
   |                         0
   |                      000
   |                     0
 50|                    0
   |                  00
   |                 0
   |               00
   |             00
 25|           0·
   |         00
   |00
   |
   |111111111
  0|----|----111111111111111111111111111111111111111111111111111111
   0        100      200      300      400      500      600
```
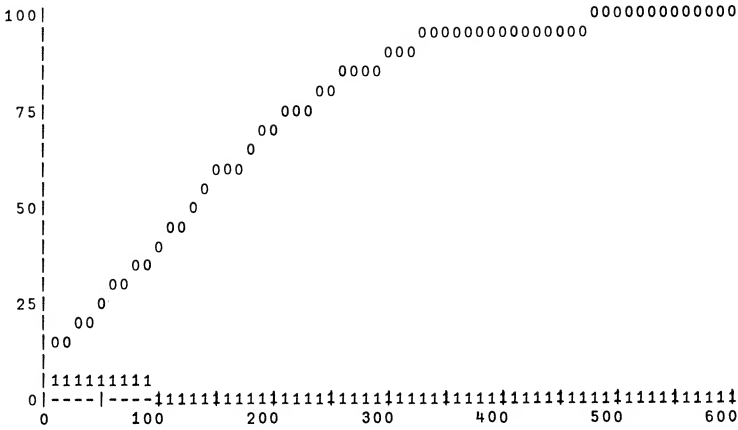
Figure 4.6. Chronological graph of counts of digits.

behaves with an entirely different starting state and input sequence, and is hardly distinguishable from Figure 4.5. In the end, when the system reaches the state (100, 0, 0, 0, 0, 0, 0, 0, 0, 0), it would no lo nger be possible to say in which of the two states the system has started.

This property of a system, which makes it likely to reach the same final state almost regardless of the initial state and input sequence, is called *equifinality* and is exhibited by many systems. Notice, though, that we said "almost regardless," for we can certainly prevent the system from reaching the equifinal state by starting it, say, in the state (0, 50, 0, 0, 0, 0, 50, 0, 0, 0). In that case, it will be driven by random inputs to the state (0, 0, 0, 0, 0, 0, 100, 0, 0, 0), as shown in Figure 4.7, not to (100, 0, 0, 0, 0, 0, 0, 0, 0, 0). For another case, consider what happens if we start the system in such a state as (0, 20, 0, 20, 0, 20, 0, 20, 0, 20). This behavior, leading with equifinality to the state (0, 0, 0, 0, 0, 100, 0 ,0, 0, 0), is shown in the chronological graphs of Figure 4.8.

If we study our system for a while, we can ascertain that there is one other equifinal state, (0, 100, 0, 0, 0, 0, 0, 0, 0, 0), which can be reached from the state (0, 99, 0, 0, 0, 0, 0, 0, 0, 1) by just the right pair of inputs (both specifying the single 9). This state will be reached eventually with random input to any starting state which contains no even numbers or 5's. In summary, then, we have four different equifinal states, which we can present in the following table:

Symbol	Equifinal State	Class of Starting States
S0	(100, 0, 0, 0, 0, 0, 0, 0, 0, 0)	Any state with one or more zeros Any state with one or more 5's and one or more even numbers
S1	(0, 100, 0, 0, 0, 0, 0, 0, 0, 0)	Any state with no even number or 5's
S5	(0, 0, 0, 0, 0, 100, 0, 0, 0, 0)	Any state with no even number and at least one 5
S6	(0, 0, 0, 0, 0, 0, 100, 0, 0, 0)	Any state with no zeros, no 5's, and and at least one even number

The mathematical trained reader will notice that, by defining the relation "leads to the same equifinal state as," we can partition all of the states of the simple system into four equivalence classes, which we can name after the equifinal state in each: S0, S1, S5, and S6. Thus, the set of all behaviors of the system breaks down into four different classes of behavior; if we wish, we can imagine the state space of the system as being partitioned into four regions, none of which can be reached from any of the others.

What is the chance of the system initially being in each of these four regions? That we cannot say, for we have not specified how the initial state is chosen.
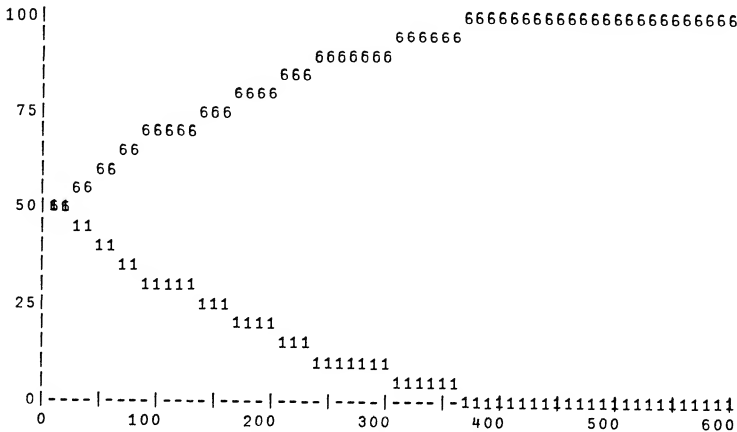
Figure 4.7. Chronological graph of count of digits.

```
100|                                                       555555
   |                                      555555555555555
   |                                 55555
   |                              55555
   |                          55555
   |                       555
 75|                     55
   |                     55
   |
   |               55
   |            555
 50|          555
   |         555
   |        5
   |      55
   |   5555
 25|555
   | 1111
   |1      11
   |       1111111111111
   |                  11111
  0|----|----|----|----|----|11111;11111;11111;11111;11111;11111;11111;
   0        100      200      300     400      500      600
```
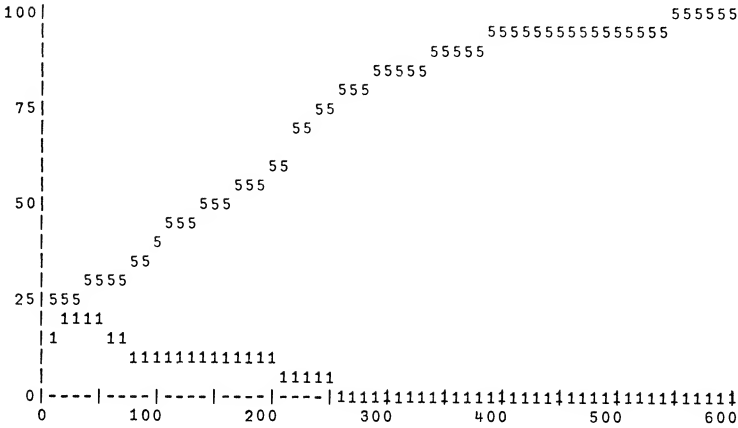
Figure 4.8. Chronological graph of counts of digits.

If it were chosen at random, then the probability would be about 0.9999999999 that it would be in region S0, about 10^{-10} of its being in region S6, about 10^{-30} in S5, and about 10^{-40} in S0. Thus, if the initial states were chosen at random, we should be unlikely ever to see any state outside of S0, and we should be content to say that S0 was the state space of the system. If, however, the starting state is not chosen at random, we might see any of the regions, and our picture of the system would be quite different. Such is the possible influence of the initial state of the system.

Effect of the input sequence. To study the influence of the input sequence on the behavior displayed by the system, let us for the moment restrict our attention to the region S0, where random input will eventually drive each of the digits to zero. Without asking the embarrassing question of exactly what we mean by "random input," let us consider what sorts of behavior might be exhibited if the input were not random in some obvious ways. For the first case, suppose that the input is constrained so that the second number of the pair is always the same number, while the first remains random. In this case, only one of the digits can ever change; hence the various states through which the system passes can never be very far from the initial state—far, that is, in the sense that none of the ten variables can change by more than plus or minus 1 from its initial value. Starting from the state of Figure 4.5, the system under this input would never reach S0, even though it is in that region of the state space.

Other nonrandom inputs, though not as drastic as our first case, would also prevent the system from reaching S0. For instance, in one of the computer simulations, the mechanism used for generating a sequence of "random" inputs happened to exclude certain numbers from ever appearing as j. The digits in those cells were never changed, so that S0 could be approached, but not quite attained. Indeed, it was the failure of the system to reach S0 that revealed the nonrandomness of the mechanism.

Much more subtle nonrandomness can have a similar effect, by prohibiting either certain points in the input state space or certain sequences. Another experience with an attempted computer simulation of the environment yielded a sequence which, though it permitted all 100 cell numbers in either position, did not yield all possible pairs of input numbers. As it happened in this case, the numbers 12, 13, 37, 82, and 94 never appeared as the second number of a pair unless one of the others was the first number. In one random starting state, cell 13 had a 5 and one of the others had an even number, so all five eventually went to zero; and similar behaviors were exhibited for several other starting states. Eventually, however, one starting state was generated in which none of the five cells contained either a zero or a 5. After a very long simulation, of course, they remained nonzero, much to the surprise of the experimenters—who then discovered the nonrandomness of the input mechanism.

The law of indeterminability. An observer viewing this case without complete knowledge of the system could arrive at one of two equally reasonable conclusions. Either the input to the system was nonrandom, or there was a subsystem consisting of five cells which somehow "resisted" the action of the input, or at least behaved differently from the rest of the system. Indeed, a simulation which was constructed along similar lines but which arbitrarily prevented cells 12, 13, 37, 82, and 94 from becoming zero (by overriding the basic rule in those cases) would behave in essentially the same manner as the unconstrained system under this constrained (nonrandom) input. We have, then, an important principle which is based on only the most general axioms of systems theory:

WE CANNOT WITH CERTAINTY ATTRIBUTE OBSERVED CONSTRAINT TO EITHER SYSTEM OR ENVIRONMENT.

This principle, which we call the *law of indeterminability*, does not, of course, prevent the observer from doing worse in specific cases, for he might be contributing to constraint himself, where none actually exists. The classic example of this type was proposed by Eddington, who imagined an oceanographic vessel whose crew, by classifying the specimens it caught in its nets, concluded that there were no creatures in the ocean less than 3 inches long.

The uncertainty as to source of constraint is of quite general importance, but it is especially important for a proper understanding of stability, one of the central concepts of systems theory. We must pay special attention to the concept of stability, because of the many confusions which always arise when a word is taken over from common usage.

In everyday speech, the word "stable" often connotes something which does not move, but we must not confuse "stable" with "immobile." Almost everyone would agree that the Empire State Building is stable, yet on a windy day the top sways appreciably from side to side. By "stable," then, we mean not necessarily complete lack of change, but rather change within certain limits. If the building sways a little, it is stable; if it falls down, it is, without doubt, unstable.

When there is no wind, most buildings do not sway at all. Conversely, if the wind gets strong enough, any building can be blown down. Must we conclude, then, that no building is stable? Not at all. Just as "stability" implies limits to the changes in the sytem, it also implies limits to the disturbances which the system is supposed to withstand. Thus, when we speak of stability, we are speaking of two things at once: a set of acceptable behaviors of the system and a set of expected behaviors of the environment. We are, to put it another way, defining a certain region of the state space of the environment and a corresponding region of the state space of the system. For example, we might define the stability of a tall building as "swaying of not more than 10 feet from the perpendicular in any direction in winds of up to 90 miles per hour."

At this point, we must call special attention to a certain lack of emphasis in our presentation of the concept of stability. We have scrupulously avoided any statement to the effect that the behavior of the environment was the "cause" of the behavior of the system. If we had done so, we would have had to assume some special powers of observation which would enable us to attribute the observed constraint to either system or environment. Since we cannot, in general, make this discrimination, we do not want our concept of stability to rest upon it. Instead, we merely require that certain behaviors be observed within certain limits—then the system can be presumed stable, within the limits of our inductive evidence.

By definition, the system is that part of the world that is of immediate interest to us. Consequently, we often elevate the system to undeserved special status in the system-environment relationship required for stability. If we are designing a system, we often overlook quite simple ways for producing stability by limiting the environment, such as building in a sheltered area or erecting some special baffle to divert the strongest winds, rather than trying to provide stability "in" the building itself.

When we say that a system is stable, we are talking about a relationship

between system and environment. More than that, whatever behavior of the system we are talking about—no matter what simplifying words we use—we are talking about a relationship. If we could only keep this equivalence between system and environment in mind, how many fruitless controversies would be eliminated!

Stability and goodness. Another conceptual difficulty we wish to avoid is the idea that "stability" is somehow equivalent to "goodness." In Ohio, there is a coal mine that caught fire 40 years ago and has been burning ever since —displaying high stability, but not much goodness. Moreover, since stability and goodness are both defined relative to a particular point of view, it is easy to see how the same situation can at once be "stable and good," "stable and bad," "unstable and good," and "unstable and bad." The form of a government may remain stable, even though various officeholders shuffle in and out. From the conservative point of view, the government is stable and good, whereas the radicals consider it stable and bad. The deposed officeholders think that it is unstable and bad, while the poor, abused citizens deem it unstable and thereby good—since the rascals get thrown out eventually.

Still, we do tend to feel that stability and goodness are related, and they are—in our minds. How does this association come about, and why is it so universal? When we seek explanations for such admittedly universal impressions, we should look for some equally universal experiences from which the impressions arise. In this case, our feelings about the goodness of stability probably arise from universal change as a personal experience. We are more likely to notice changes than lack of change. Furthermore, among the things we notice, the ones which cause us pain or discomfort tend to stand out as individual impressions. Thus, when a change takes place, the ways in which we feel we are worse off generally make a stronger impression, so we begin to equate change with badness and, by implication, stability with goodness.

Perhaps each of us has in mind an ideal world in which only the "bad" things change, while all the "good" things stay the same. But our world is not built that way, if only because we change our definition of what is good as time goes by. In the same way, we can change our definition of what is stable, and a system that was once "stable" can become "unstable" simply by our changing ideas as to what should be the range of its behaviors or what should be the range of environments in which it is stable. The change may come about gradually, as in the case of a parent who comes to accept new behaviors of his children; or it may be rapid, probably in response to an actual event which taxes the previous definition. Suppose, for example, that a building were to fall down on a day when the wind velocity reached 110 miles per hour. The owners of the building and the people injured in the event would accuse the architect of having designed an "unstable" building. At such a time, it

would not do much good to remind the owners that they approved plans in which "stability" was defined in terms of maximum winds of 90 miles per hour. The occurrence of a 110-mile-per-hour wind has changed their conception of stability, even though before the fact they might never have approved of the extra expense to build for a wind velocity not previously known.

Stability and observation. There are three main parts of the definition of stability, and all depend on the observer's point of view: the system behavior and the environment behavior depend on his selection of the system, and the critical limits assigned depend on his goals. Nevertheless, isn't there something special about the concept of stability, something that makes it one of the central systems concepts rather than simply a derivative and dependent one? The answer, of course, is yes, and the reason for this answer was supplied in a definition of "system" given by Parsons and Shils [5, p. 107] when they said, ". . . if a system is to be permanent enough to be worthy of study, there must be a tendency to maintenance of order except under unusual circumstances." In other words, a selection of variables out of the world does not have to exhibit stability to be a system, but the less stable it is, the less chance it has to be "worthy of study."

Implicit in our definitions of stability for a system is our inability to think of "change" as an abstract entity, our necessity to identify "change" with two "things," the old thing and the new—different—thing. In order to have a "thing," we must feel that there is a certain persistence of the characteristics by which we identify the "thing," and it is to get this persistence that we require stability. Perhaps someday we will be able to think in different ways about change, but for the present we require stability in our systems if we are to think about them at all.

In science, we often go to extreme lengths to find the needed stability. To a physicist, an isotope which exists for a millionth of a second is a system worthy of study, whereas "common sense" tells us that such a system really does not exist at all. In engineering, the study of linear systems has been elevated to high importance, largely because these systems have the property that, if they are stable for any input, they are stable enough if the input is doubled—or multiplied by any factor. Of course, the accepted range of system behavior must also be multiplied by the factor; but the important point to the engineer is that the system under study does not exhibit different kinds of behaviors, such as runaways or breakdowns, just because the magnitude of the input is increased. In common experience, however, we rarely come across a system which even barely approximates a linear one, for all systems break down when they are "overloaded."

Yet, everything considered, scientists have been very successful at abstracting

points of view which elevate in importance the more constant parts of the system—so successfully, in fact, that they forget that this constancy is largely a matter of choice, not chance. Darwin [6, p. 46] recognized this situation over a century ago when he observed:

Authors sometimes argue in a circle when they state that important organs never vary; for these same authors practically rank that character as important (as some few naturalists have honestly confessed) which does not vary; and, under this point of view, no instance of an important part varying will ever be found; but under any other point of view many instances assuredly can be given.

Darwin was, of course, thinking of anatomists and naturalists. It is interesting, however, to observe the same kind of circular argument being expressed more than 100 years later in the field of anthropology by such a noted theorist as Julian Steward [7, p. 184] (author's italics):

The present statement of scientific purpose and methodology rests on a conception of culture that needs clarification. *If the more important institutions of culture can be isolated from their unique setting so as to be typed, classified, and related to recurring antecedents or functional correlates, it follows that it is possible to consider the institutions in question as the basic or constant ones, whereas the features that lend uniqueness are the secondary or variable ones.*

In the end, nothing is constant. Everything changes, and only change is unchanging. For a person to make any sense out of such a world, he must introduce mental simplifications which ignore certain changes—or, rather, which classify certain changes as being irrelevant to the continued existence of the mental construction, or system. Even the word "person" denotes this kind of simplification, for who would deny that he is different in some ways from moment to moment? We cannot even express the thought of change without recourse to some implicitly unchanging person about whom we speak. Thus, though our expressed purpose is to study change and though we know that everything changes, we must study things which seem to remain the same, and we must do it in terms of things which we imagine are unchanging.

At bottom, then, this persistent search for the invariant in system behavior is a consequence of the inductive nature of our learning. If a system or sub-system is to be around long enough for us to learn about it, it must have certain invariant properties. If it does not, it becomes "something else" before we have had time to recognize it as a system. Consequently, a question of fundamental interest to general systems theory is, Why do some systems survive whereas others do not?

4.6. REGULATION, ADAPTATION, AND LOSS OF IDENTITY

If you look at automata which have been built by men or which exist in nature you will very frequently notice that their structure is controlled to a much larger extent by the manner in which they might fail and by the (more or less effective) precautionary measures which have been taken against their failure. And to say that they are precautions against failure is to over-state the case, to use an optimistic terminology which is completely alien to the subject. Rather than precautions against failure, they are arrangements by which it is attempted to achieve a state where at least a majority of all failures will not be lethal. There can be no question of eliminating failures or of completely paralyzing the effects of failures. All we can try to do is to arrange an automaton so that in the vast majority of failures it can continue to operate. These arrangements give palliatives of failures, not cures. Most of the arrangements of artificial and natural automata and the principles involved therein are of this sort. John von Neumann [8]

Why does a system survive? From the long-range viewpoint, a system survives because systems which do not survive are not around to be thought about. The systems we are accustomed to seeing are those which have been selected from all systems of the past; they are the best "survivors."

Lest we imagine that this is some sort of trivial quibbling, we should observe that surviving is a truly remarkable thing for a system to do. Our view is biased because we most often see systems which are good survivors, but the vast majority of systems do not survive for long—over any time span that we wish to choose. At the level of biological individuals, we have reason to believe that nobody lives forever: the most ancient living things we know are no more than 4000 years old. If we choose populations as our systems, so that the systems survive even when individual members die, the situation is not much better. Since life began on this planet, over 90% of all species that ever lived are now extinct—and there are but few species like the cock-roach, which has been around about 300,000,000 years. Human organiza-tions are even more puny. Most new businesses fail, and it is hard even to think of a business which has been in operation for more than a few hundred years. Organizations like the Roman Catholic Church, which is not quite 2000 years old, are of the greatest rarity.

Survival, then, is far from being a trivial property of system behavior. It is a property which every system must have for us to study it, and a property which not every arbitrary collection is likely to have. Consequently, it is important that we have a pellucid understanding of what we mean by "survival." Since survival is the continued existence of a system, we must, if we are to be precise about this property, examine the meanings of *continued* and *existence* in this context.

Survival time. "Continued" refers to the length of time that a system has to exist in order to be worthy of study. How long this must be is a question of relative time scale between system and observer—and thus, is related, at least indirectly, to the typical length of time the observer survives. In the case of man as an observer, the effect of time scale is not difficult to ascertain. We do not, for instance, ordinarily think of plants as moving about under their own power; however, if we watch a plant through the quickened time scale of a time-lapse motion picture, we see it as writhing about in a most active manner. Through the device of slow-motion photography, we can begin to empathize with worlds such as the microbiological one, where things otherwise are born and die before we can apprehend them.

Our simple system, which has served us so well in illustrating other points, can be used also as an excellent illustration of time scale. Recall the four regions in the state space, S0, S1, S5, and S6, so named after the equifinal states to which they led under random input. While we pictured the approach to states S0, S5, and S6, we only talked about the inevitability of reaching S1. The reason we did not show the approach was time scale, for if we show the behavior starting from such a state as (0, 25, 0, 25, 0, 0, 0, 25, 0, 25), the result is something like Figure 4.9, which does not at all seem to be approaching S1.

The reason the behavior does not seem to be approaching S1 is evident in Figure 4.10, where we have started the system in a state (0, 99, 0, 0, 0, 0, 0, 0,
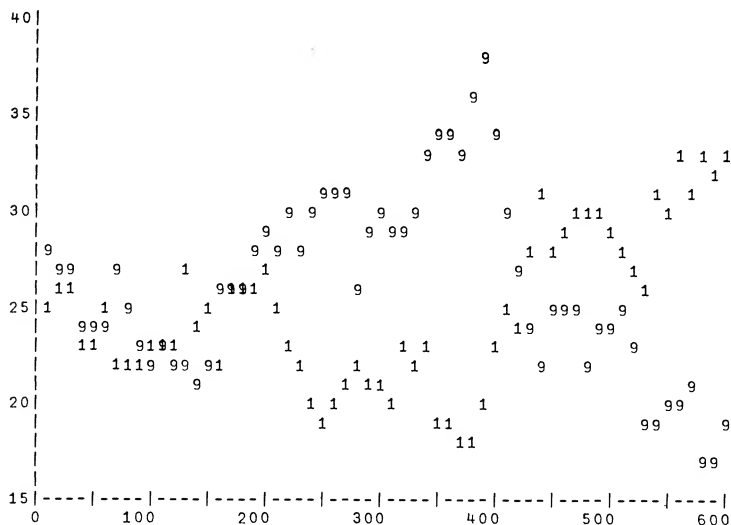
Figure 4.9. Chronological graph of counts of digits.

```
100|1111111
   |        111111
   |              11111
   |                   11111
   |                        1111
 75|                           11111
   |                               11
   |                                111
   |                                  11      1
   |                                  11111 1111111 1111111
 50|                                                     ∮
   |                                         99999 9999999 9999999
   |                                       99       9
   |                                     999
   |                                   99
 25|                                99999
   |                             9999
   |                          99999
   |                       99999
   |              999999
  0|9999φ99--|----|----|----|----|----|----|----|----|----|
   0        100       200       300       400       500       600
```
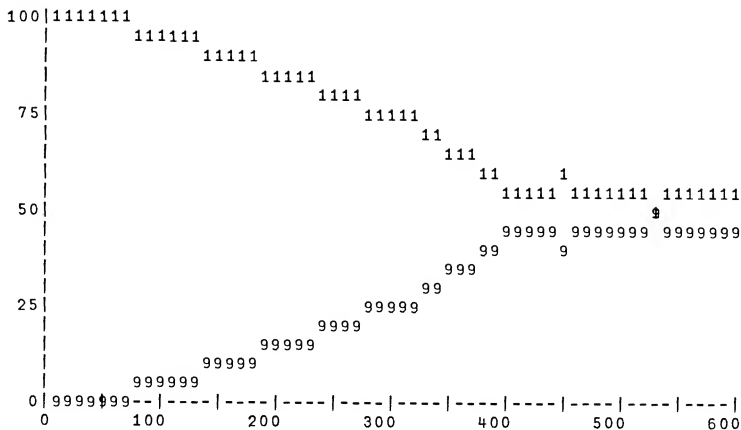
Figure 4.10. Chronological graph of counts of digits.

0, 1), the state which must precede S1. An input such as the fifteenth line of Figure 4.2 would be required to move the system to S1; however, that did not occur just at this time, and the system moved farther and farther away from S1, though remaining in a subspace of all 1's and 9's. We might have expected this to happen, for the probability of just the right input occurring before (0, 99, 0, 0, 0, 0, 0, 0, 0, 1) was lost is 1 in 100. Thus the system might reach the state (0, 99, 0, 0, 0, 0, 0, 0, 0, 1) many times before passing into S1, where it would "stick."

How long, then, would it take to reach (0, 99, 0, 0, 0, 0, 0, 0, 0, 1)? Although this state might be reached from (0, 99, 0, 0, 0, 0, 0, 1, 0, 0), the best chance of getting there would be from (0, 98, 0, 0, 0, 0, 0, 0, 0, 2), because the latter is a lumping of more states. Nevertheless, the probability of the required penultimate transition is only 0.04, so the probability of reaching S1 from (0, 98, 0, 0, 0, 0, 0, 0, 0, 2) is $0.04 \times 0.01 = 0.0004$.

We could continue the analysis, but we have made our point; starting from an arbitrary state in S1's region, it will probably be a very, very long time before we actually get to S1—though once we do, we shall surely stay there. How long is "very, very long"? If the computer we are using can make 1000 state transitions a second, we should probably have to wait much longer than the age of the universe to see S1. So, though we know that the system must, by force of logical necessity, reach S1, we would never characterize it in that way if we knew it only through observation of its behavior. In that case, Figure 4.9 would be a "typical" picture of the behavior of the

system, and we would say that that behavior has a tendency to survive—even though it is doomed to eventual extinction. And if we did ever happen to observe one of these systems reaching S1, after having known it for a long time, we would probably say that the system we knew no longer existed. Thus we come to the second part of our ideas regarding survival, the idea of existence.

Existence. To exist is to have an identity. Identity is indeed synonymous with viability, for nothing remains to be identified which is not viable, and a thing which changes its identity passes out of existence. But to have an identity is to have an identifier, and hence comes the difficulty of saying when a system exists or ceases to exist. Did Rome fall in 476? Some historians say so, but several million Romans live in the city today. Did Dr. Jekyll cease to exist when Mr. Hyde was roaming the streets? It all boils down to a question of how we establish the identity of a system.

Looking out my kitchen window on an early spring day, I saw a blue jay in the cherry tree. I recognized it as a blue jay because it was blue. I am not a professional (or even an amateur) bird watcher; if someone showed me a red blue jay, I would call it a cardinal. Furthermore, I feel safe in saying that most people would make the same identification. But an ornithologist sees the matter differently, for to him color is not the major identifying variable of blue jays. No doubt he would be surprised and delighted to find a red blue jay, but he would never mistake it for a cardinal; to him it would be a sport, a mutation, but still a blue jay. Now, who is right, the ornithologist or I? We are both right, of course. We can argue about whose set of identifying variables is superior; but even there, we are likely to come away from the argument holding our original views, for my criterion is perfectly adequate for my poetic purposes and completely inadequate for the ornithologist's prosaic ones.

In any case, suppose that the blue jay which inhabits my cherry tree turned red overnight. To me, a blue jay has disappeared and a cardinal has come into my world; to the ornithologist, the jay has changed its color. Until we can agree on just what states of what variables identify a blue jay and a cardinal, we will be wasting our time in an argument about what happened to the blue jay.

Perhaps this example seems ludicrous. It was chosen for clarity—not for subtlety—so as to avoid prejudicing the argument, as might have been done with numerous classical examples. When did the first man appear on earth? What happened to the Standard Oil Company? Who is a Negro? Does a revolutionary government have to honor agreements made by its predecessor? Is space empty or filled with an ether? In all these cases, the argument has meaning only because there is no agreement about the identifying variables

of the system in question. Still, we all are confident that we know a man when we see one, are aware what constitutes a company, and can recognize Negroes, governments, and empty space. Only when our fuzzily defined concepts are challenged by actual cases do we see how fuzzy they usually are; in everyday life we have no need for more careful delineation.

To discuss the problem of identity, we do not need to go to the trouble of making an explicit list of the identifying variables of any system; rather, we can separate the problem into two parts. First, we can assume that all observers will see when certain changes are made in the system. Second, we can explore the questions of how observers will disagree when they have different lists of identifying variables for the same system. We mean to imply not that people always carry around perfectly defined lists in their heads (or even that they ever do), but only that they do have some idea of what they mean when they talk about the continued existence of a system or its failure to continue existing.

But surely are there some systems on which everyone agrees? Probably not, if we ever really got down to cases; but to the general systems theorist it is simpler to assume that there may always be differences and to see where that assumption leads. If there is general agreement in a particular scientific discipline or culture or family or even the whole human race, well and good; but we do not need such agreement to talk about systems. We only need to agree that such agreement is theoretically possible so that we can talk at all.

Suppose, then, that we have agreed on how to recognize a particular system. We have then agreed on what variables are relevant in the recognition and on what values those variables must have (such as the redness of the cardinal). If our identification is based only on a selection of variables—identifying variables—and their values, we can decide on the question of identity by making a single observation. In that case, the system is recognized if all the identifying variables lie within the proper ranges, that is, if the system occupies a particular region of the state space at the time of our observation.

In other cases, a single observation may not suffice, for we often identify systems by the behavior they exhibit, not merely by their instantaneous states. If, of course, a system were closed and state determined, the observation of one state would always suffice, for it would be equivalent to observing all future behavior. Because of this behavior of closed, state-determined systems, we tend to think of identification as a process involving a single set of observations which could, in principle, be made simultaneously.

The problem of survival. In cases where identification proceeds in this manner, the problem of survival can be represented schematically as shown in Figure 4.11, a *diagram of immediate effects*, following Ashby. In this figure, the environment is represented as affecting the identifying variables through
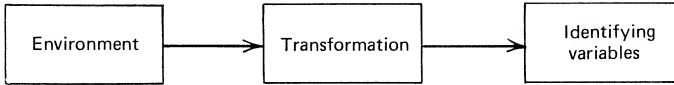
Figure 4.11. The general problem of maintaining identity.

the medium of a transformation. The transformation is the set of rules which determine how particular events in the environment will affect the entire state of the system and, in particular, the state of its identifying variables. In our formula

$$S_{t+1} = F(S_t, I_t),$$

the F represents what we have represented in this diagram as the transformation. As the formula shows, F determines the new state of the system, once the previous state (the system's contribution to the change) and the input (the environment's contribution) have been given.

The transformation rules stem partly from the "laws of nature" and partly from the way in which the particular system is put together. Thus, for example, as the outside temperature rises from 10 to 20°, the internal temperature of one animal may rise while that of another may fall. Similarly, a decline in the stock market may be profitable for some investors and disastrous for others. The transformation, in other words, is one of the factors which prevents us from speaking of environments as "good" or "bad." As Veblen observed, there is no change, no matter how generally beneficial, which does not harm someone, and no change, no matter how generally odious, which does not act to someone's advantage.

Now, in terms of the diagram of Figure 4.11, a system survives if its transformation continues to convert variations in the environment into values of its identifying variables which lie within the indentification region of its state space. Alternatively, a system survives in an environment if its identifying variables are stable in that environment.

Seen in this way, the problem of survival is a problem of having the right transformation at the right time. It would seem, however, that as time went by there would be fewer and fewer systems as the unsuccessful transformations fell by the wayside, one by one. Formally, this argument may be true, but on any reasonable time scale it does not have to be. The catch lies in the question of how a system is identified, and what the source of its transformation is.

In our simple system, as implemented on a computer to produce our descriptions of its behavior, the state of the system is represented by a set of

digits stored in the computer's memory. The environment is simulated by another set of digits, which is generated in another part of the same memory. In a way, then, the computer's memory represents a larger state space out of which our system and its environment were projected. The advantage of this outlook lies in the homogeneity of the computer memory as an embedding space, for the state of any system embedded in it is represented in the same way—as a set of digits.

If two of the boxes in Figure 4.11 can be thought of as digits in the computer's memory, what about the third, the transformation? In the computer simulation, the transformation is simulated by a program. In a computer, the program is the set of instructions obeyed by the computer which determine how it will operate on the numbers in its memory to produce new numbers in its memory. Since the numbers in the memory constitute (or simulate) a state space, the program plays the role of the transformation, that is, the set of rules determines how particular events in the environment will affect the state of the system. But the program is stored in the computer as a set of numbers, numbers which represent the instructions in coded form. Thus the transformation, too, is represented in this state space as a set of numbers!

Recognizing the homogeneity of representation in the computer's memory, we can begin to appreciate the power and suggestiveness of the digital computer as a tool in the study of systems. Examination of our simulation has led us quite naturally to recognize a fact which would be rather difficult to appreciate fully in a more natural (and thus less homogeneous) system. The new fact is simply stated: the transformation is part of the system. Symbolically, the equation describing the behavior of the system can be rewritten as

$$S_{t+1} = F(S_t, I_t) = n((V_t, T_t), I_t),$$

in which the state variables of the system have been partitioned into T, the variables affecting the transformation, and V, the other variables, including the identifying ones. The small letter n represents a transformation, too, but now stands for the transformation brought about by the laws of nature, that is, the universal laws which do not depend on the particular structure of the system at hand. What the formula says, then, is that the state of the system at any given time is obtained by applying the laws of nature to the state of a larger system, $((V_t, T_t), I_t)$, which includes the environment. If we wish, this formula may be taken as a definition of what we mean by "laws of nature."

In the case of the computer simulation, the universe of choice is the computer's memory, as shown in Figure 4.12. Consequently, the laws of nature are the programming rules which are built into the circuitry of the computer and which determine how different sets of digits are to be interpreted as instructions to change other digits in the memory. The computer itself is, of
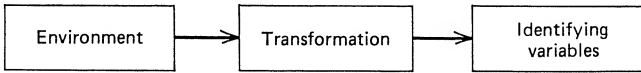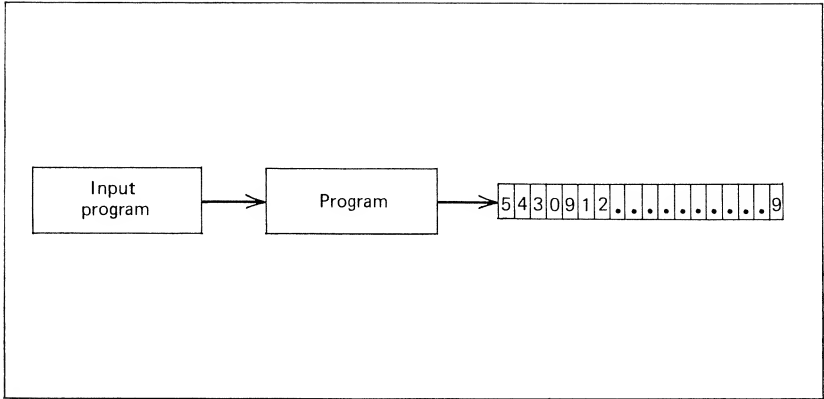
Figure 4.12. Embedding a model of a system in a computer memory.

course, embedded in an even larger universe of choice, a universe whose laws may be taken to be the ordinary laws of science, or some local application of those laws. The existence of this larger universe need not concern us in studying our simple system, because we have restricted our attention to an approximately closed system embedded in that universe. In our artificial universe, the laws of nature need not bear any resemblance to the laws outside.

Regulation and adaptation. Having expanded our awareness of the structure of our system, we can return to the question of survival and the number of possible transformations. Previously we implicitly assumed that the transformation was fixed, but our new point of view shows us that the transformation can be separated into a fixed part (the *laws of nature*) and a variable part (the *program* of the system). Since there is a variable part to the transformation, or program, one way in which a system might preserve its identity is by changing the transformation itself, rather than by using a fixed transformation to convert the environmental variations into acceptable values of the identifying variables.

When the system survives by making a change in its transformation, we say that the system is *adapting*. When it survives while retaining a fixed transformation, we say that the system is *regulating*. Regulation and adaptation are two central concepts of systems theory, for every system either regulates or adapts; otherwise it loses its identity.

By recognizing the possibility of changing the transformation, we solve the problem of using up all possible transformations. In our system, for instance, we previously recognized 10^{100} states, but we were using only 100 digits of the computer's memory to achieve this vast number. If the computer has a memory of 1,000,000 digits—a typical number—there are $10^{1000000}$ different programs (transformations) which can reside in that memory. Each of these programs will have some characteristic effect on the 100 digits we have singled out for attention, but many of them will have the additional property of changing some of their own digits, as indicated in Figure 4.13. Thus each transformation we observe is one of a set of transformations, each of which produces the same behavior over the period of our observation.
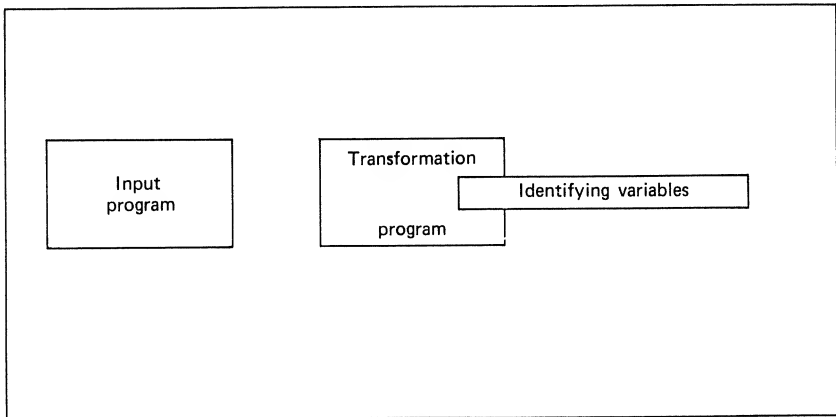
Figure 4.13. Overlapping of transformation and identifying variables.

As an illustration of the richness of transformation, let us return to the observed behavior of our simple system when in S0. After a reasonable amount of time, the system settles down into S0, never to show variation again. Never? Well, hardly ever. Figure 4.14 shows the behavior of a system which is very similar to our original one, but which has a special feature to its transformation. Whenever S0 is reached, the system waits until two equal input numbers are received and then suddenly restores a random state to the

```
100|          0000              00000000          00000              000
   |        00                00                00                000
   |       0                 0                 0                 0
   |      0                 0                0 0                0
   |     0                 0
 75|                                        0                 0
   |     0                 0
   |                                                           0
   |                              0
   |   0                                                     
 50|                    0                                    0
   |                   0                    0
   |   0                              0
   |                                  0
   |             0
 25|0                                                     0
   |                                  0
   |11                  0            1                1
   |  1                 1            1                1
  0|---1l11111l111-l11111l11111l11111-1l11111l11111l1--1l11111l11111l
   0        500       1000      1500      2000      2500      3000
```
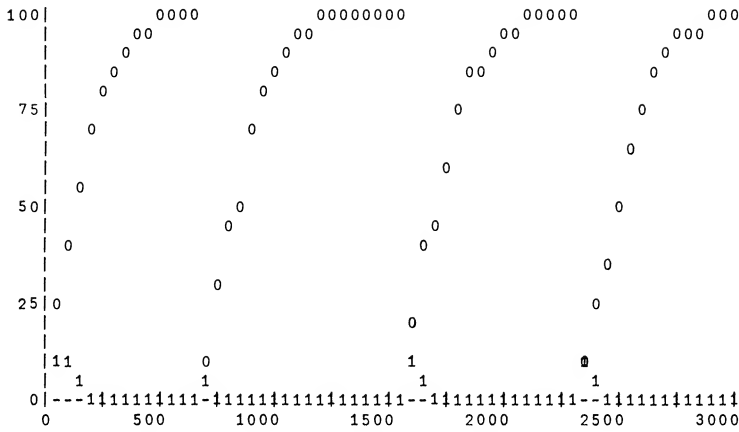
Figure 4.14. Chronological graph of digits.

100 digits. The behavior we observed in Figure 4.5 would be precisely the behavior of this system in an environment which produced random inputs with one constraint—that no equal pairs were produced as input.

We should not be surprised by this result, because of the law of indeterminability. We are surprised, however, because we tend to think of the transformation as a separate, or separable, part of the system. There are important reasons why we have this impression, but for now we need only observe that an inexhaustible supply of new transformations will always exist—"always" in the same sense that our simple system will "never" reach S1, even from the region of S1.

Identification by transformation. All of these conclusions were reached on the basis of a simplifying assumption—namely, that the system was identified by the states of certain variables, and not by the display of certain behaviors. We now see, however, that once the transformation is represented as part of the state, the identification by behavior—in a standard environment—is reduced to the identification by region of the state space. The only difference is that some of the transformation variables, T, are included in the list of identifying variables. Thus, the division into transformation variables, identifying variables, and other variables is not a true partition, for certain variables may be in two classes at once.

The assumption that the variables of a system can be partitioned cleanly lies very deep under many scientific arguments. From biology, we might say that this view represents the "anatomists." The anatomical view tries to

understand change through statics, which would be simpler, could it be done. But even species discriminations cannot be done on a strictly anatomical basis, and confusion reigned on the species question until a behavioral definition of "species" came into common use. Although the behavioral view cleared up certain questions in classification, it ruined the simplifying assumption of a clean partition between transformation and identifying variables. In effect, the concepts of adaptation, regulation, and failure to survive became no longer separable as they were under a purely anatomical view.

Once we have so identified the source of trouble, many of the major problems which have plagued philosophers and scientists simply disappear. In biology, does the species of fish which starts to be able to breathe air become a new species or is it simply adapting? In anthropology, does the group of people which adopts the language and some of the other ways of another group become a new culture, an adapted culture, or part of the culture from which the adoptions are taken? In organizational theory, does the acquisition of new tasks produce a new organization or is the result an adaptation of the old organization?

Since the method of breathing is part of the transformation as well as part of the identity of a "fish," since a language is part of the transformation and part of the identity of a "culture," and since the tasks performed are part of the transformation and part of the identity of an "organization," none of these questions can be solved—because the definitions of "adaptation" and "preservation of identity" rest on an assumption which they do not satisfy. If we tried to put these problems to a computer, as we should with our simple system, we would have to be explicit in our definition of the system, and we would find that we were unable to make the required partition. Of course, if we arbitrarily assigned the variables in a certain way in order to obtain a true partition, we would get a point of view under which adaptation had taken place; but if we threw the partition in the other direction, we would then have loss of identity.

Confusions of differences among the different viewpoints of a system may also lead to confusion between adaptation and regulation, or even between regulation and preservation of identity. These confusions are compounded because of the complementary relationship which exists between regulation and adaptation:

1. A system which is doing a good job of regulating need not adapt.
2. A system adapts in order to simplify its job of regulating.

For our final example, we choose the realm of psychology, where such paradoxical material is almost the trademark of the field. The sense of self, or ego, is often closely related to particular behavior patterns—or, in our terms, transformation variables. When an individual makes such an identification,

he may continue to exhibit the behavior in question even when signals from the environment seem (to us) to make such behavior inappropriate. Thus a man may continue to behave in a socially unacceptable way, even though he is in danger of losing his job, his friends, or his family because of it. To him, the particular behavior pattern represents his identity more than do the things whose loss is threatened. He thus screens out or avoids the warning signals from his environment; hence, from our point of view, he is regulating to preserve his identity, to survive. The more effective is this regulatory system, the less likely he is to alter the offending behavior; and the only hope for change is to change his method of identification.

4.7. PROSPECTS FOR THE FUTURE

I have but one lamp by which my feet are guided, and that is the lamp of experience. I know of no way of judging of the future but by the past . . . .   Patrick Henry

We have tried to show how the computer as a model can have a continuing influence on systems thought. This chapter, of course, represents only part a of the approach, and the contribution of only one person or of a small group of people. We have argued elsewhere [9] that the growing use of digital computers will create increasing numbers of workers who will be attracted to general systems theory. We have also written about how interaction with computer models can be explicitly educational [10, 11]. A third type of interaction between computers and systems theory is obtained by reversing the direction of interaction and applying systems theory insights to certain difficult computer problems [12, 13].

The achievements, then, in terms of published output, are modest, to say the least. Judging the future by the past, we would have to say that this is not a promising approach from the point of view of research output. Some potentially strong results can be obtained, such as the clear idea that regulation and adaptation can never be distinguished once and for all, but must remain as useful, but potentially conflicting concepts. But perhaps the nature of general systems theory is such that we cannot expect much in the way of "results" from any approach, for when approaches become "results" they cease to be general systems theory.

General systems theory, according to our view, is no more a collection of results than a science is. *A set of ways of looking at the world*—that would be a better description of general systems theory. Thus we feel that the major job of general systems theorists is changing the thought patterns of people, not publishing "results." Our "results" are the people whom we can teach to think a little more productively. If we can do that through published papers,

well and good. If we use interaction with computer models, well and good also. And what is the sense of arguing whether mathematics or computer languages constitute the best way to reach systems concepts? Let us hope that no one system of systems theory ever eliminates the rest—that no approach is elevated to the rank of dogma, and no group to the rank of high priests. Should we not "let a hundred flowers bloom . . ."?

PROBLEMS

4.1. In economics, one sometimes speaks of the "law of supply and demand," which says that, if demand is greater than supply, price rises until demand is diminished or an increase in supply is obtained, and that, conversely, if demand is less than supply, price falls. List and discuss several simplifying assumptions which might underlie such a law in economics, just as some of the assumptions which underlie a law of physics were discussed in this chapter.

4.2. Write a computer program for solving sets of linear algebraic equations, or find such a program in a book or in a program library. Make an empirical study of the behavior of the program (in particular, the execution time) for various-sized sets of equations and for various sets of coefficients within each size (e.g., with many of the coefficients equal to zero). How does the behavior of the program compare with the square law of computation? Does the program take advantage of any possible separation of the system into non-interacting subsystems? If it did, how much would the user gain?

4.3. In a business magazine and also in an engineering magazine, see how many chronological graphs you can find. Also see how many other forms of state-space representation you can find in each. In the systems described, identify the initial state and the input sequence, and discuss the effects which different ones might have on the behaviors shown. If possible, describe each behavior in terms of some other representation than the one given.

4.4. Make a list of some systems which you encounter in your daily life, such as a particular tree, animal, business, machine, building, or road. Try to establish a lifetime for each, and discuss the problems of attributing lifetimes to such systems.

4.5. Write a computer program to simulate our simple system. Have that program produce at least three new views of the system behavior in addition to the ones shown in the text.

4.6. Modify the computer program for our simple system so that the transformation follows this rule: Multiply the ith and jth digits, as in our simple system.

(a) If the result is a single digit, use it to replace the jth digit.

(b) If the result is more than a single digit, add those digits to obtain a new result. If this result is still more than one digit, add the digits again. Use the final result to replace the jth digit, as before.

Thus, for example, in our simple system, 7 times $4 = 28$, giving 8; but in the new system, 7 times $4 = 28$, giving $2 + 8 = 10$, giving $1 + 0 = 1$.

Study the behavior of this new system, and discuss its behavior in terms of the principles discussed in this chapter.

4.7. The most general transformation in a 100-digit system is simply a table of 100 digits, arranged in a 10×10 matrix. The result of digit a interacting with digit b is found by extracting the element in row a, column b, from the table.

Cast the program for Problems 4.5 and 4.6 in the form of such a transformation table, and create some different transformation tables which will display interesting behavior.

4.8. The transformation in Problems 4.5–4.7 is completely independent of the state of the system. However, since the state and the transformation each consist of 100 digits, interesting systems can be created in which the two sets of digits overlap, completely or partly. Write a computer program in which there is overlap between the state and the transformation, and study its behavior in terms of the principles of this chapter.

REFERENCES

1. **Deutsch, K.,** "Mechanism, Organism, and Society." *Philosophy of Science,* Vol. 18, pp. 230–252, 1951.

2. **Rapoport, A.,** "Mathematical Aspects of General Systems Analysis." *General Systems Yearbook,* Vol. XI, pp. 3–11, 1966.

3. **Buckley, W. F.,** *Modern Systems Research for the Behavioral Scientist.* Chicago: Aldine Publishing Co., 1967.

4. **Ashby, W. R.,** *An Introduction to Cybernetics.* New York: Wiley, 1961.

5. **Parsons, T.,** and **E. A. Shils,** *Toward a General Theory of Action.* Cambridge: Harvard University Press, 1951.

6. **Darwin, C.,** *On the Origin of Species* (Facsimile Edition). Cambridge: Harvard University Press, 1964.

7. **Steward, J.,** *Theory of Culture Change.* Urbana: University of Illinois Press, 1963.

8. **Von Neumann, J.,** *Theory of Self-Reproducing Automata.* Urbana: University of Illinois Press, 1966.

9. **Weinberg, G. M.,** "Systems Research Potentials Using Digital Computers." *General Systems Yearbook,* Vol. VIII, pp. 145–150, 1963.

10. **Weinberg, G. M.,** "Learning and Meta-Learning Using a Black Box." *Cybernetica,* Vol. XIV, No. 2, 1971.

11. **Weinberg, G. M.,** and **L. Hyvarinen,** "Computerised Hill Climbing Game for Teaching and Research." In *Optimization*, R. Fletcher (Editor), pp. 171–186. New York: Academic Press, 1969.

12. **Weinberg, G. M.,** "Natural Selection as Applied to Computers and Programs." *General Systems Yearbook*, Vol. XV, pp. 145–150, 1970.

13. **Gause, D.,** "New Considerations in Adaptive Programming Systems." In *Proceedings of the Sixth International Conference on Cybernetics.* Namur, Belgium, 1971.

Part II

Contemporary Systems Problems:
The Social Context

5. The Hierarchical Basis for General Living Systems

JOHN H. MILSUM

BioMedical Engineering Unit, McGill University, Montreal

5.1. INTRODUCTION

Until the last few decades a hierarchical structure was accepted almost universally as the necessary and sufficient basis for all the living systems encountered around us, including those of which we are part. Indeed, our everyday language and folklore are full of expressions implicitly involving the hierarchical concept, for example, "knowing one's place," "climbing the ladder," "the delicate balance of nature," and "the trophic food (web) pyramid." Even the long history of political struggle has largely been aimed at producing revolutions which only replace one hierarchy with another; indeed, the relatively recent democratic movements and revolutions aimed at establishing ideal egalitarian societies have, from the functional viewpoints of

how power flows and how control is exercised, resulted in only slightly modified hierarchies.

These last comments make no value judgment upon whether better or worse societies can result or have resulted from political revolutions, but rather state only that the underlying social structures seem to remain hierarchical. Consequently, it seems pertinent to ask whether a "biological imperative" applies in regard to the necessary structuring of human, and indeed of all living, societies, at least up to now, and given a certain level of complexity in the society.

At this point, a difficult but relevant question to resolve is whether there is an essential discontinuity in passing from the "involuntary" structures of man's internal ecology to the nominally "voluntary" structures of his external societies. We note first that human society is deeply involved in the natural ecologies comprising flora and fauna. Now, in at least those ecologies which we have not seriously perturbed, the imperative for a hierarchical structure, in particular for that of the food pyramid, seems generally accepted. Such a hierarchical structure can demonstrably favor the development of stability in the population structure of an ecology, especially when a large number of species are involved in the predator-prey reactions. Indeed, it seems tempting to infer that the operation of selective pressure tends toward population stability, possibly because this measure implies a minimization of the probability of population crashes, and therefore of species extinction; equivalently, it favors the probability of life continuing, in at least some forms.

A rigid hierarchical structure is also evident in many of the highly developed animal societies, especially those of insects. Later it will be argued that this structure may well be optimal for an "information-poor" society. On the other hand, in human societies the current information revolution based on electronic computers renders us, potentially, an information-rich society, so that at least the economic argument for a more or less rigid hierarchical structure is less urgently applicable. However, the release of this constraint does not necessarily guarantee that the ideal human society is nonhierarchical. Indeed, a gap seems necessary between the ideal form of society to which we nominally aspire and the practical form which we are realistically prepared to support.

The structure of our human societies has obviously been a matter of basic concern to us throughout our recorded history. Even earlier, mankind had myths about the origins of his society, and one is particularly pertinent to the present theme. Thus Fromm [1] writes:

... Bachofen [2] suggested that in the beginning of human history sexual relations were promiscuous; that therefore only the mother's parenthood was unquestionable, to her alone could consanguinity be traced, and she was the authority and lawgiver

—the ruler both in the family group and society. [Then] . . . in a long-drawn-out historical process men defeated women, subdued them and succeeded in making themselves the rulers in a social hierarchy.

These two forms of society are, respectively, the matriarchal and patriarchal orders.

In the matriarchal concept all men are equal, since they are all the children of mothers and each one a child of Mother Earth. The patriarchal system, on the other hand, considers obedience to authority to be the main virtue. Instead of the principle of equality, we find the concept of the favorite son and a hierarchical order in society.

There are still some human societies which are largely nonhierarchical, for example, the North American Indian and the Eskimo [3–5], but none is of the large, complexly interdependent, industrialized type. Thus, while the use of hierarchy cannot confidently be asserted as the necessary basis for industrialization, it seems impossible that any presently industrialized society, especially our own, would ever voluntarily and responsibly accept the risks implicit in a complete, or even large, removal of hierarchy. But since the revolution in information and control technology is now removing much of the physical rationale for the need of hierarchy, it becomes important to understand better how the underlying psychosociological factors affect the putative *biological imperative* favoring hierarchical societies. Indeed, since the technological revolution is catalyzing the other current revolutions which are attacking the discriminatory conditions of racial, generational, and sexual classes, this new understanding will be essential to our survival. Thus it becomes important to question whether complete elimination of all discrimination would not lead to an unacceptably dull, unmotivated, homogeneous society. Indeed, it is important to draw a very clear distinction between the idea of equality of opportunity for all, and the idea that all are equal; instead, in Ardrey's felicitous phrase, we need to recognize and accept the *innate random inequality of man* [6].

This chapter, then, will attempt to explore the need and limitations of a hierarchical structure in general living systems. The dual processes of cooperation and competition will be considered as essential, concomitant but complementary aspects of such hierarchical structures. The treatment will cut across the approaches from many different fields, for this idea of hierarchy has been recognized in a wide range of disciplines. This range might conveniently be categorized, for example, by noting the matriarchal-patriarchal concepts of Bachofen, already mentioned, at the human symbolic level, as compared with the formal mathematical treatment of hierarchical structures by Mesarovic et al. [7]. Indeed, we should note that the idea of hierarchy has not been limited to living systems; for example, one recent

book includes the observed clustering of the cosmic physical system into a hierarchical set [8]. However, this aspect is outside the scope of the present chapter, for here we are concerned to include that essential positive-feedback ability conferred by the evolutionary process whereby the complexity of living systems can apparently continue to grow. Specifically, we focus on the extent to which the use of hierarchical structure has been and may continue to be an optimal strategy in the growth of complexity and stability in our living systems.

5.2. DESCRIPTION OF GENERAL LIVING SYSTEMS

5.2.1. Natural Hierarchies

The word "hierarchy" comes from the Greek *hieros* (sacred) and *archos* (ruler).

(a) The human hierarchy. If the cell is considered to be the fundamental unit of viable life, a natural hierarchy follows, embracing all living systems. Of course, the cell itself constitutes a complex dynamic system of a hierarchical type in that there is a master controller, the DNA, and several levels of manufacturing responsibility. At present we shall concentrate upon modeling at the supracellular level, but later we will summarize the intracellular hierarchical system. In order, also, to avoid the complexity of differences between various organisms, we shall initially consider the natural hierarchy based on the human being (Figure 5.1).
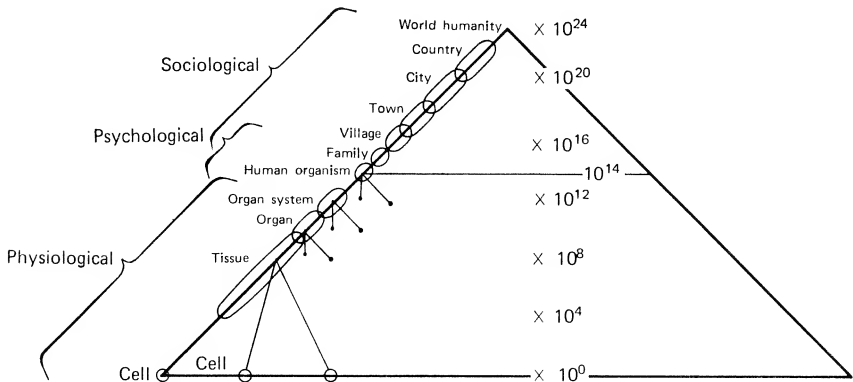
Figure 5.1. The human hierarchy. The multiple subsystems at any level are suggested by the several radiating lines. The variation in size at any level is suggested by the length of the link.

Cell. The human hierarchy starts with the human cell, which specializes during the organism's growth into one of some two dozen types, subserving many different functions. Notable examples are the *receptor*, transducing information about the organism's environment (and about its own *milieu interne*); the *neuron*, for information transmission and processing, both of incoming and of outgoing signals; and the *muscle* cell, for manipulating the organism or the environment. The number of cells within the human adult is not known with any certainty, but it is approximately of order 10^{14}. Three functionally useful hierarchical levels can be distinguished within this range, however.

Tissue, which is essentially a group of largely similar cells, specialized for some function. Thus an organ such as muscle consists of the actual contractile muscle tissues and of connective tissues, among others. Depending somewhat on its definition, a tissue may well have up to order 10^{10} cells.

Organ, which essentially comprises a set of different tissues, grouped together in a functionally recognizable way to perform one or more functions. Muscle provides an important example of an organ, comprising muscular, neural, connective, and epithelial tissues. Its primary evolved use is the generation of mechanical forces from chemical energy, but in addition it has evolved the shivering mechanism whereby all of the chemical energy converted is degenerated into heat for thermoregulatory purposes.

Organ system, which comprises a set of organs, usually operating as a recognizable negative-feedback control system. The neuromuscular system provides a pertinent example, comprising the organs of muscle, nerve, and central nervous system (CNS). There are of order 10^3 skeletal muscles in man, but many more if smooth muscles, such as those in the precapillary sphincters, are included.

There are perhaps of order 10^2 variants of organ systems, the main ones being, in addition to the neuromuscular systems, the cardiovascular, respiratory, thermoregulatory, receptive, central nervous, gastrointestinal, reproductive, and excretory systems.

Organism. The human being is the complete, independently functioning organism, the result of this hierarchical set of cells. It is relatively easy to visualize the evolutionary march upward through ever-increasing complexity of organization, because of the advantages gained by cellular specialization when this is harmoniously controlled in a hierarchical way for the "common good." However, when we ask for whom is the "common good," it is not easy to answer whether it is for the good of all the constituent parts, or of the whole organism only, or of both. In other words, is the system of these elementary parts greater than the simple sum of the parts? The deeply held

human idea of *soul* as distinct from the material *body* obviously suggests the answer yes, and so, in an apparently different context, does systems theory. In any case, when we start a further progression up through the hierarchical levels of human societies, the idea of *purpose*, which can be plausibly recognized as the basis for the organismic hierarchy, is no longer so defensibly pertinent.

Family. The family, of order 10^1 human beings, constitutes the purposeful unit by which the growth stage of the new generation, from conception to adulthood, has been ensured. Historically, the family has been of the hierarchical type, typically with three hierarchical levels of husband, wife (wives), and children. As such, it has been a unit in the patriarchal rather than matriarchal society. Of course, the value and the continued existence of this structure are now being re-examined very critically, at least in Western societies.

Village. This unit of order 10^2–10^3 human beings, is still of conceptual and practical importance in society. It may be considered to circumscribe the mutually supportive activities of a group of persons who interact *personally* with each other.

Town, City. These units are of increasing population size. In the city, the population is of order 10^6 and is too large, spatially and intellectually, for a single grouping of the various main facilities and activities. Such a grouping is still possible, however, in a town.

Country, World. Although a *country's* size and boundaries have been determined in part by ethnic homogeneity, a major factor historically has also been the extent of the spatial domain which could be effectively controlled, given the particular state of technology required for the movement of information, materials, and people. *Empires* have represented further groupings of countries, as regards hierarchies of control, but this particular level may now largely be extinct. The human *world* at present merely represents the total human biomass, rather than any coherently functioning total human system. However, this does not deny that some control processes are continuously at work through politicoeconomic channels; it denies only that the same effective political hierarchy of control exists as in any viable country.

(b) The intra-cellular hierarchy. The scheme for intracellular hierarchical control, as related through the hormonal systems to the supracellular systems, has been summarized by Weber et al. [9]; see also the discussion by Waterman [10]. As shown in Figure 5.2, the information flow is implemented in a multihierarchical system. However, the hierarchical flow of control information is reversed to some extent from that of Figure 5.1, depending on the
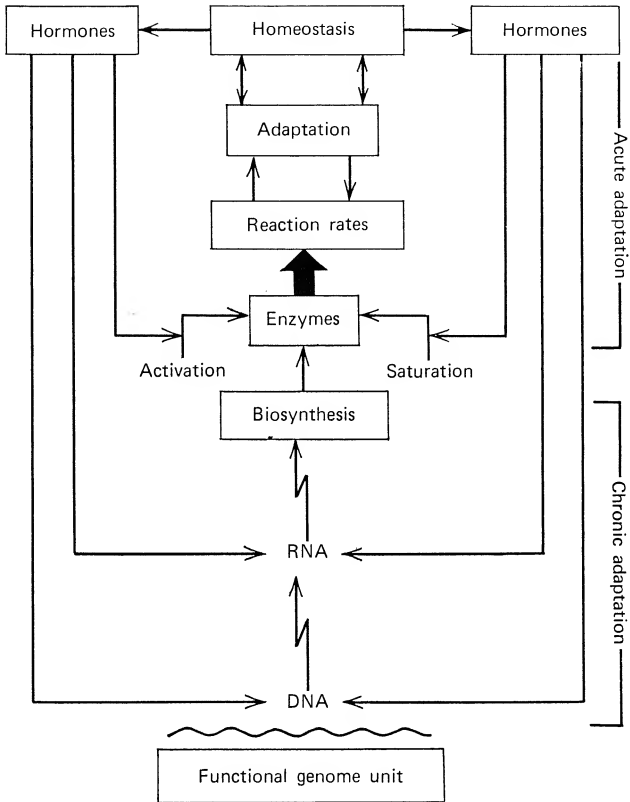
Figure 5.2. Hierarchical organization of genes, enzymes, and hormones in the overall homeostatic system. Antagonistic action of hormones control enzyme activity in short-term adaptation, and genetic mechanisms are responsible for their long-term adaptation. From Weber et al. [9].

observer's viewpoint. Thus, on the one hand, the hormones released by the negative-feedback controllers of homeostasis produce conventional "downward" constraints in the hierarchy onto the enzymes, so as to produce corrective homeostatic action. On the other hand, the flow intracellularly is that the DNA controls RNA synthesis, which in turn controls the enzymes. Finally, the DNA-RNA stages are controlled by long-term genetic adaptation, the *chronic* adaptation of Figure 5.2, rather than the *acute* or fast adaptation in the enzyme path. In the human hierarchy, the corresponding slow path would be the "social adaptation" by which different types of individuals or

communities become adapted or selected. The fast hormonal control of enzymes would correspond to the fast-acting social control of law and management.

   (c) The management hierarchy. In a slightly different human grouping, that of persons working in an enterprise, a hierarchical system normally symbolizes the form of management. This comprises conventionally the successive descending levels from the single office of "president" to much more numerous "workers" (Figure 5.3). This span typically covers a numerical range of about 10^4 in some four hierarchical jumps. (We note that there are usually "boards of directors" to guide the executive regarding policy,
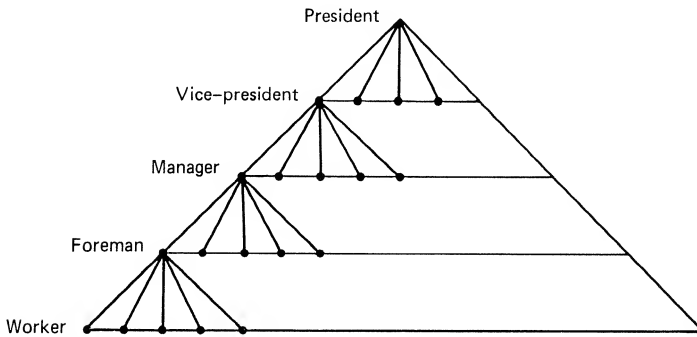
Figure 5.3. A typical administrative hierarchy.

but they are not essential in the present argument regarding executive control.) The essential structure has tended to remain invariant, independently of whether the enterprise is industrial, government civil service, military, or other types. Even in the university, where for good reasons the management function has been least clearly defined and centralized, there has been a generally similar administrative structure. This pyramidal form has, of course, the sociological rationale that at each level in the hierarchy the "manager" must operate through the group of his particular assistants—the prototype political "cabinet," in fact. A number of order 10^1 appears to be the largest that can be handled conveniently by any one manager, that is, a *convergence factor* around 10 appears to be sociologically ideal in moving up a hierarchy. As an interesting comment on this, it has been noted that a political cabinet has a strong tendency to grow over time, but that whenever its size exceeds about 21 it tends to generate another intermediate hierarchical level of "executive cabinet" [11].

(d) Mobility in the hierarchy. In this presentation of the ubiquity of social hierarchies, no comment has yet been made about the ability of individuals to move around, and especially upward, in these hierarchies. Indeed, the hierarchical nature of the structure itself should be clearly differentiated from the mobility of the individuals within the structure. One difference between a static society, such as a "feudal" one, and a modern industrial society concerns the ability of an individual, according to his "innate random inequality," to move upward in the hierarchy. We note further that there has been a tendency for an age-specific pattern to exist in which the movement is correlated to a considerable extent with increasing age. This is based on the belief that with increasing age comes increasing knowledge, wisdom, and usefulness, but in recent years this idea has been subject to considerable questioning. Specifically, the revolutionary nature of technological change has militated for the use of younger people in responsible positions. Indeed, it is sometimes questioned whether the accumulated knowledge and wisdom of the generation currently in managerial command has any social usefulness today.

(e) The experiential hierarchy. From the viewpoint of a given individual moving through the social hierarchy in time, it is interesting to consider also the situation of the individual moving through a hierarchy of ascending personal achievement levels. In general, the base levels are concerned mostly with fulfillment of material aspects (the *deficiency needs*, in Maslow's terms [12]), whereas the upper levels are concerned with the satisfaction of the integrated thinking and experiencing human being (the B-values, *the intrinsic values of being*, in Maslow's terms). In this view a person aims to ascend to the highest level of which he is capable, the reward being that experienced by him internally, for example, Maslow's *peak experiences* [12]. Graves has proposed a recognizable hierarchy of seven levels [13], while from a slightly different viewpoint Erikson presents a set of eight problem and achievement levels encountered during different phases of the individual's life (the *eight ages of man*), as summarized in Figure 5.4. Thus Erikson, and before him Freud, are considering the results of the individual's innate physiological-psychological growth processes, in which each individual moves through a hierarchy of personal achievement. Also, of course, this approach is compatible with that of many great religions, especially the Eastern ones.

It is tempting to speculate that, if the neurophysiological basis of these experiential hierarchies is ever discovered quantitatively, it will correspond to the growth of a hierarchical neuronal network. Certainly there is some suggestion for this in the way in which the CNS of animals has increased hierarchically through phylogenetic development. For example, man's neuromuscular system certainly has the basic spinal reflex arc, but on this
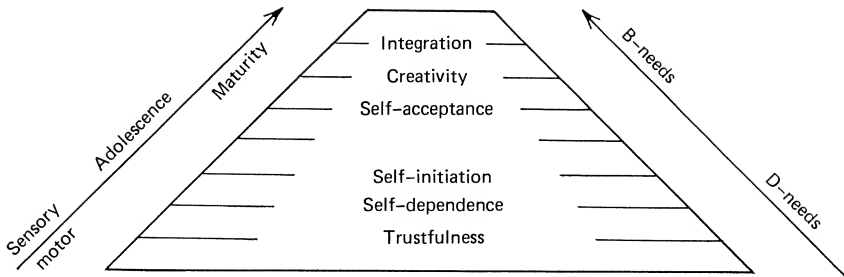
Figure 5.4. Stages of psychophysiological development in man—inferred from Erikson [14], Graves [13], and Maslow [12].

are superimposed successively higher "long-loop" pathways through the brain stem, cerebellum, and cerebral cortex [15]. Present indications are that the basic spinal reflex is retained and indeed is essential for fast protective reflexes, but that the hierarchical loops are used for successively more refined, skilled, and stable reflexes and voluntary actions. It seems probable, then, that evolution would also incorporate such a successful, economical strategy in other neural systems, for example, the CNS itself.

(f) The trophic hierarchy. Another important natural hierarchy is that of the *food web* or *pyramid* in an ecology. In general, the predator-prey interactive web is very complex, for many species may be involved, but the conventional pyramidal structuring is as shown in Figure 5.5. This pyramidal structure is based implicitly on the underlying energy flow, which is upward through the various *trophic* levels of the hierarchy, for this may be viewed as the essential "engine" powering an ecology. Since the "efficiency" of energy conversion between the different levels is much less than unity, only a very small, omnivorous biomass can be supported for a given biomass of photosynthesizers at the base level. The energy conversion efficiency may vary
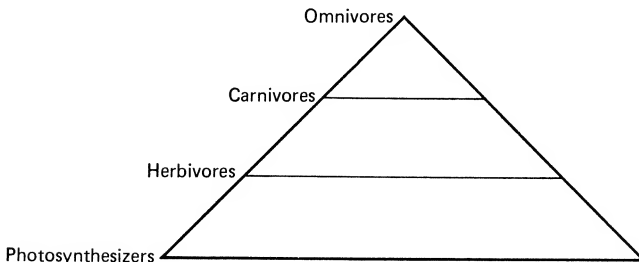
Figure 5.5. The food pyramid (trophic web).

somewhat among these trophic levels, and in any case is not known exactly. A figure of $\frac{1}{10}$ probably represents a reasonable approximation to an overall average, however [16], as a result of which we note that a convergence factor of about 10 again applies to each upward step in the hierarchy. In this regard it has been noted generally that the hierarchical structure provides a way of avoiding " direct confrontation between the large and the small in nature " [8].

(g) Interwoven hierarchies. Finally, in regard to natural hierarchies, we should comment that any individual entity in a system, the human being, for example, may well exist simultaneously in a multiple set of coexisting hierarchies which are different in either space, time, or essential variables. Thus the university professor has simultaneous places in the university hierarchy, the "invisible university" hierarchy of his codisciplinary workers around the world, the local sociopolitical hierarchy, the local hierarchies of age, sex, and race, and finally the food web of which he is a part.

5.2.2. Arguments for Hierarchies

Since the evolutionary process has amply validated the value of a hierarchical structure, let us at this point consider in more detail some of the arguments in favor of hierarchical systems.

(a) Elementary processes — few and simple. A few elementary processes or building blocks may still allow structures of great complexity to be built out of them. The major processes used in natural systems are, for example, *reproduction and growth* (DNA, RNA, enzymes, membranes, and cell, as in Figure 5.2); *energy conversion* (ATP and oxygen); and *organization* (cell specializations, homeostatic systems, neural networks). This paucity of basic building blocks requires less "capital investment" than other methods.

(b) Specialization. A particularly important economy and flexibility within hierarchies has been made possible by the ability of cells to specialize for different functions. Some important types other than those already mentioned (receptor, nerve, and muscle cells) are the following: epithelial, structural, glandular, blood, luminescent, and electrical capacitive. An especial advantage accrues when the specialization may be changed if necessary during the life of the organ.

(c) Economy of information flow. Information flow necessitates a finite cost, partly in capital investment and partly in an operating cost proportional to the rate of information transfer. When these costs are high relative to other important costs, the system may usefully be categorized as an *information-poor* or *information-scarce* economy. In an information-poor system, a hierarchical organization is particularly desirable, because the upward convergent

flows of information may be processed and the necessary control decisions made at a single hierarchical stage (or at a few stages) up from the initiating information. Only that processed information is then passed on to further hierarchies which is necessary for decision making at their successive levels. Such hierarchies are rather rigid in their behavior patterns, and although presumably very near optimal for insect societies they are not ideal for mankind. Nevertheless, the human cost of information processing is very significant, for example, in being physically present at committee meetings. Thus it may often be nearer optimal for individuals to delegate their authority in certain respects to an individual at one level up in the hierarchy; in this way they can be free to do their own preferred tasks. However, as we shall explore later, the electronic information-processing revolution opens up additional operational methods which were not previously available, in particular by the effective removal of many hierarchical decision makers, in that the previous constituency of each can now pool and process its information appropriately for decision making directly in a computer. Any such changes, however, will be the result of moving into an *information-rich* economy.

The preceding discussion has implicitly concerned the suprahuman hierarchies of Figure 5.3 or, in Figure 5.1, the levels above 10^{14}. Below this level in Figure 5.1, in the area of *physiological systems*, all the evidence seems to point toward a highly structured hierarchy. Thus we note the many local reflexes, and the comment that evolution has not yet entrusted the cortex with any biologically important functions (e.g., one cannot voluntarily hold one's breath until death occurs). Furthermore, we note the large data convergence in moving up the hierarchy, as exemplified by the eye having perhaps 10^8 visual receptors, but only about 10^6 neurons in the optic nerve carrying the visual information on to the cortex. On the other hand, the CNS does have the wide "alarm" powers of activating systems at any level when necessary through broadcast neural and/or hormonal commands, which is a capability implicit in the hierarchical structure. Flexibility of control here has clearly been important, and it is noteworthy that in the human being the brain, the "president" of the personal hierarchy, constitutes only about 2% of the body mass but receives about 16% of the oxygenated blood supply in resting conditions. In other words, selective pressure has favored a rich information-processing capability, even though its cost to the system is high.

In the ecological pyramid of Figure 5.5 the concept of information flow is less directly applicable, since as noted already the implicit assumption is that of energy flow. Indeed, the concept of an actively operating control downward through the hierarchy, based on information converging upward, is questionable. That is to say, the carnivores do not predate upon their prey, the herbivores, *in order to* control the prey population or to control some other variables. Rather, we are forced to infer that evolution has operated at the

ecological level, so that those ecologies have been selected which operate, through the genetic endowment of their partaking species, *as if* a control process were at work. In other words, the various species exhibit those characteristics of predation, procreation, and the like which are appropriate, and indeed which could well be those which would be programmed by an intelligent, active control structure. In this context, and over fairly short periods of time, these two types of system operate similarly, and it is not helpful to try to distinguish between them functionally. Thus we may say finally, in regard to economy of information flow, that ecologies behave *like* an information-poor system with a relatively strong hierarchical structure.

(d) Accumulation of information (knowledge, experience). The evolutionary process may be considered as a very complex, continuing *Monte Carlo* experiment. A crucial particular result of this process is that the ineffective solutions are rejected, and only the effective ones are retained. In consequence, a set of current *best estimates* to an idealized solution is made available. Upon these best estimates further complexity can be grafted, using the same continuing Mònte Carlo form of experiment. Although this method cannot in principle assure that a better solution could not be found to any current problem by starting over again, it would certainly seem to be the best strategy for an optimizing mechanism, working on the random search principle. Notice that a temporal hierarchy is being structured here, in which time is the ordinate of the hierarchy and the different biological variations of each species, ecology, and so on constitute the abscissa. Thus at any given time, corresponding to some level in this new temporal hierarchy, only small proportions of the variations for each species, ecology, etc., are effective and therefore are passed on upward in the temporal hierarchy.

This evolutionary process has interesting analogous features to the acquisition of human knowledge, especially through the method of science. In both cases transient successes of ineffective (wrong) solutions may occur, but the self-corrective nature of the process subsequently eliminates them. Indeed, it is even analogous in the sense that it proceeds by finite jump steps rather than by continual smooth accretion. In this regard Kuhn has pointed out that science actually advances fundamentally through breakthroughs or revolutions, usually with considerable opposition and turmoil [17]. In the biological analogy, breakthroughs have also provided the keys to the accumulation of new information and life forms. Thus the few basic processes noted in Section 5.2.2(*a*), such as DNA-controlled reproduction and ATP energy conversion, represent good examples of breakthroughs. An outstanding instance of such a breakthrough has been the biological "discovery" of the negative-feedback mechanism which is now so ubiquitous and essential a feature of all the homeostatic and neuromuscular control systems in animals.

This analogy should not be pushed too far, however. Thus in science it is hopefully unnecessary to repeat all past experimentation, and indeed the experimentation process itself is at best serendipitous rather than random. In nature, however, we assume that an unguided random experimentation never ceases. Of course, the individual in his own life cycle represents an intermediate case, for all practiced knowledge and skill must be gained anew by each generation, and the learning process seems inevitably to require the making of mistakes. In more general terms, no skill or knowledge can be known to be optimal or true unless the knower has sampled the conditions producing nonoptimal or untrue results.

It is also worth noting that essentially only since the invention of the printing press has knowledge begun the exponential growth [19] which, not accidentally, has been similar to human population growth, even though with a much shorter doubling time. This presently uncontrolled information explosion provides a stimulus to which the human individual and his society do not yet possess stable and satisfactory responses.

5.2.3. Competition and Cooperation

Competition and cooperation constitute key basic processes in living systems, and largely as a result of them living systems exhibit their many complex behavior patterns. Their influences are inextricable, and indeed often complementary, inasmuch as competition at one hierarchical level of organization may be recognizable as cooperation at the next higher level. The complex structures which living systems have evolved over such long periods imply that some optimizing principles of design have been effectively at work.

A restatement of a few basic ideas pertinent to the theme of competition/cooperation will be helpful at this point.

(a) Growth. All living organisms have high potential growth (birth) rates. In consequence, any early but transient stage of growth is characterized by *positive feedback* dynamics and is represented simply in differential equation form by

$$\frac{dx}{dt}(t) = kx(t). \tag{5.1}$$

However, unlimited *exponential* growth (Figure 5.6) is not possible; instead the growth rate must decrease as the population increases. Many theoretical formulations exist, but the *sigmoid* (or *logistic*) growth curve represents the simplest conceptual extension of Eq. 5.1 which is of some practical value:

$$\frac{dx}{dt}(t) = k\left(1 - \frac{x(t)}{x_0}\right)x(t), \tag{5.2}$$

where $x_0 = $ "saturation" population as $t \to \infty$.
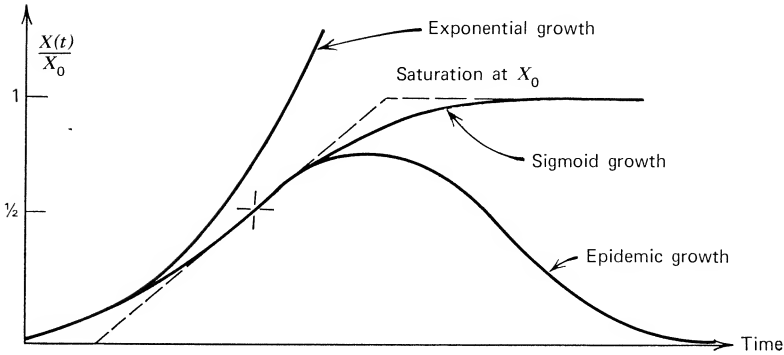
Figure 5.6. Three characteristic growth patterns.

A third basic growth pattern of interest is that of *epidemic* growth (Figure 5.6), in which, after an initial rise, the population peaks and then decays either to some small level or to zero. For present convenience this can usefully be represented by the Gaussian density function, although in detail we note that the latter is symmetrical and also decays to zero.

(b) Scarce resources. Given high potential birth rates, the inherent practical constraint in any physical system of having only finite resources (*scarce resources*) must eventually become effective. Thus the curve of population growth must be forced away from the unlimited exponential, as already exemplified in Figure 5.6.

(c) Competitive exclusion. When resources are scarce, the fittest of the systems competing for these resources survive preferentially. In consequence, and because of the inherent positive feedback dynamics of potential growth, this implies the phenomenon of competitive exclusion. A simple model for competitive exclusion, by which less fit systems are shut out [20], is sketched in Figure 5.7; the reader is referred to [21] for more detail. Note that competitive exclusion operates intraspecially so as to enhance the "breeding pure" tendency, and interspecially so as to exclude less fit species.

(d) Specialization and cooperation. Given this selective pressure, the process of specialization would have an advantage for survival in many conditions. Initially, this selection occurred around the cellular level but later operated also at successively higher hierarchical levels. At the species level specialization occurs in the context of filling particular ecological *niches* most effectively. The process of specialization amounts to selection for any complexity which provides better performance. As such, cooperation
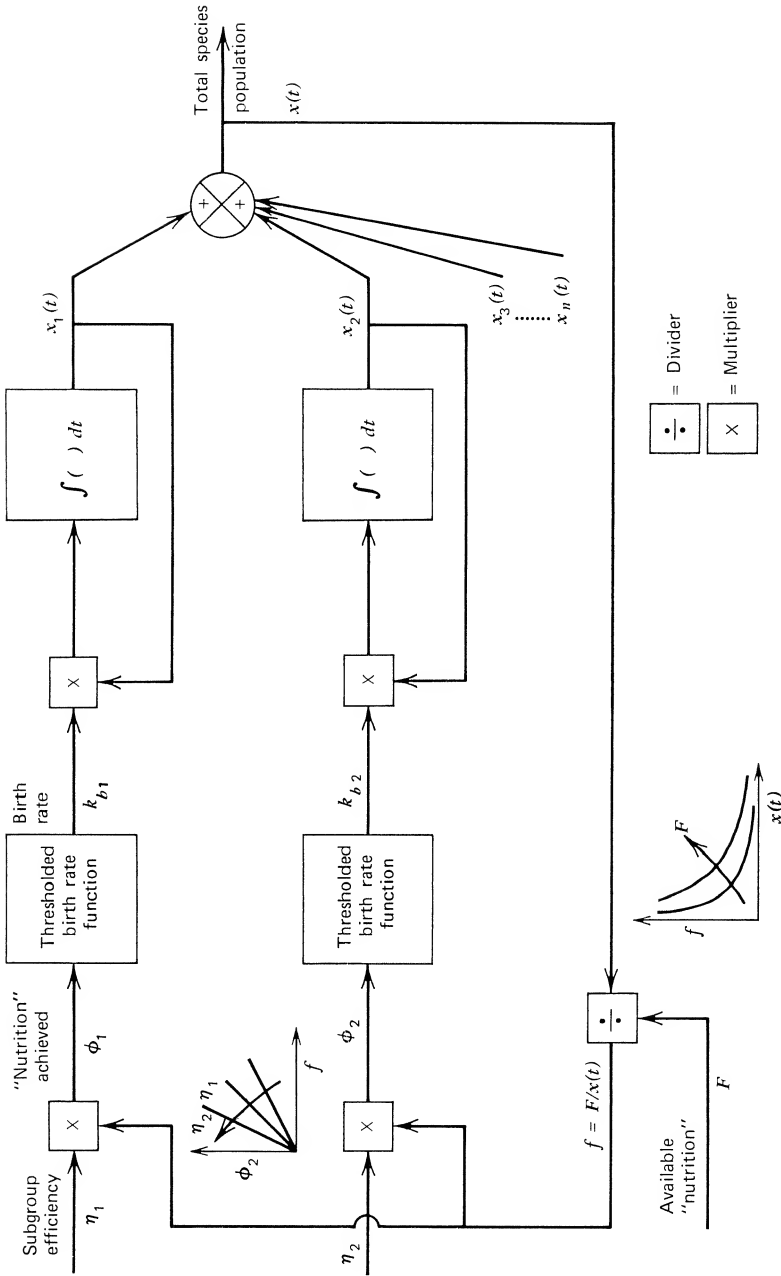
Figure 5.7. Simplified model of intraspecal competitive exclusion. From Milsum [20].

160

also includes the various *symbiotic* arrangements by which different organisms obtain mutual benefit.

It is interesting to note that this process of increasing specialization and complexity is paralleled in engineering. Thus the performance of some initial design is usually extended, or "stretched," as time passes, and this improvement is almost always obtained by successive refinements which, in fact, complicate the system. Thus the original turbojet engine for aircraft propulsion has been made more complicated by the addition of such refinements as automatic controls, afterburners, multiple and different-speed rotating parts, the "bypassing" of some gas flow, thrust reversal, and sound reduction equipment.

The viability of the more complex system depends typically on the reliability of the basic design having been made excellent before the additions are successively made, proof of system reliability being needed at each stage. Of course, the specific addition of safety devices can directly improve system reliability, but in both engineering and nature these tend to be used only after the basic reliability is good, and for particular purposes. The neuromuscular safety reflex attributed to the Golgi tendon organ provides an example of such a device, for it can intervene to protect the integrity of the ankle joint when unanticipated balance or footing problems suddenly occur. In the case of the jet engine, an automatic safety device can, for example, provide for extinguishing an unexpected fire.

When these considerations are extended to include the various interorganism and interspecial symbiotic arrangements that have evolved, a general impression emerges that evolution tends toward increasing complexity, and that this is made possible in considerable degree through specialization and cooperation. Some recent evidence concerning the evolution of brain size in mammals supports this belief, since brain size presumably correlates with system complexity. Evidently there has been a progressive increase in brain size during the period from 65 million years ago to the present. Here brain size is scaled nondimensionally as the size relative to that to be expected in a living mammal of a given size [22].

(e) Duality of competition and cooperation. As noted already, competition at one level may functionally be identical with cooperation as viewed from the next higher hierarchical level. For example, the predator-prey relation is certainly highly competitive at the personal encounter level, but its function in the integrated context of the food web is cooperative. Furthermore, in the civilized human activity of athletic and intellectual games, the competitive and cooperative aspects exist simultaneously at the same level; that is, the harder the competitors strive against each other, the more they feel the satisfaction of a cooperative interaction. The unfortunate fact that the animal

"will to win" may frequently in practice override the amateur ideal of "good sportsmanship" does not inherently upset the validity of this cooperative concept. In any case, even in mortal human combat the loser in a hard-fought battle can feel respect for his vanquisher. This feeling is, of course, reciprocated by the more fortunate victor in the conflict, and this "moment of truth" presumably contributes to the unshakable fascination that animal hunting, fishing, bullfighting, and similar sports and contests hold for many persons.

In another area, the "laissez-faire" theory of economics was based on the idea that unbridled competition would select the most efficient producers in society. Therefore it would provide the goods which society needed most economically, to the greatest overall benefit; that is, the *unseen hand* effect would be that of cooperation. Of course, the totally free enterprise system has, in the event, proved unsatisfactory in all societies, partly because it tends toward the result of monopoly, after which a change of mode occurs since efficiency is no longer a necessary criterion. Such a system proves socially unacceptable and is therefore unstable.

(f) Games and their rules. Rules have been evolved at many levels for the competitive/cooperative games. As examples, we note the phenomena of *territoriality*, to stabilize and maintain viable the breeding population; *symbolic competition and fighting*, to prevent fatal intraspecial fighting; *prudent predation*, to limit predatory take to that needed; and all the levels of human *games* (childish through to warfare). In general, as any system matures the rules become more formalized. In consequence, less effort and uncertainty is involved, with the whole system becoming more stable. Thus it is partly because of their relative newness that so many societal games are unstable, in the sense that the intellectually preferable condition cannot be reached by either party in a dispute. *Game theory* shows formally that this is often due to the fact that the ideal solution requires the invocation of mutual trust [23]. However, since neither disputant can independently afford to take the risk of trusting the other, each is forced to the nontrusting, and highly undesirable, solution. Arms races constitute one example of current and basic international concern. Clearly, an effective higher hierarchical level of control than the national level must be generated if the desirable stable solution is to be reached.

(g) Systems versus subsystems optimization. If particular subsystems of a larger system operate so as to optimize their own individual "good" (performance indices), the net result will almost never be overall system optimization. One particular problem is the "tragedy of the commons," to use Hardin's phrase [24]. In this situation, the commons represents a resource held in common by individuals. Each may draw from this resource, and no problem arises until it becomes "scarce." If each individual continues to act

as a unit, his best strategy is to seek a greater share of the resource. Unfortunately, however, this applies to all, and hence the amount or quality of the common resource is rapidly depleted to zero, to the mutual ruin of all concerned. Thus again, as in game theory, there must be constraints enforced on subsystems, in order that the common good may be realized.

(h) Evolution at ecological level. Evolution operates at the ecological level also, in the sense that ecologies which have evolved so as to optimize their overall performances, as measured by some pertinent indices, are better fitted than those which have not. This provides the basis for the apparent purposefulness which an observer is tempted to infer as existing in ecological systems.

(i) Performance indices. The indices by which ecological performances have been compared competitively in nature cannot be inferred by us with certainty; they seem, however, to be related more to *stability* of living systems than, for example, to maximization of biomass. It is still a moot point whether stability in this context means a decay to essentially steady conditions after being perturbed, or whether the normal condition may involve some steady-state cycling. Thus, on the one hand, it can be argued that any oscillations in such variables as population and biomass increase the probability of extinction occurring. On the other hand, any system which adaptively optimizes itself must generally continue to oscillate about the optimal point in order that any changes in this point due to disturbances entering the system may rapidly be *tracked.* In this respect, we note the pertinence of Christian's postulate [25] that species which exhibit large population fluctuations, such as lemmings and voles, would be able to exploit new environments rapidly and to evolve in them. For this dispersal process, social hierarchy would provide a major driving force, since the low-ranking members would be forced to emigrate. There have been important examples of this process in human society also, of course.

(j) Civilized society. Acceptance by individuals of rules has provided the basis for civilized society. The greater stability, safety, and productivity of civilization that resulted has been traded off against the loss of the individual's complete freedom. Among other disasters, civilized society has protected its individuals against the *tragedy of the commons* [see Section 5.2.3(g)]. Thus, on the average, society has both regulated the extent to which any individual could exploit a common resource, and increased the resource if necessary, for example, by providing distribution systems for utilities and transport. Of course, our civilizations are currently faced with a new set of common resource tragedies, most notably the loss of right to such unpolluted resources as land, air, water, sound, or space unpolluted with other people!

When a civilized society has been successful, it has also regulated its intra-societal games so that the desired outcomes of mutual benefit can be obtained, rather than the "all-lose" conditions of war and chaos. Now, on the one hand, the result of achieving this end normally has been a hierarchically ordered society, since the necessary controls have been complex, time-variable, and adaptive, requiring good management. On the other hand, this has required of the individual a significant loss of his complete personal freedom, and this loss may produce anxiety, frustration, or stress, to such an extent that even in the 1930's Freud was pessimistic about man's ability to live satisfactorily in these conditions [26].

It is pertinent that stress in the neuroendocrine sense of Selye [27] seems to be produced by overcrowding, among other causes. Thus, in increasingly crowded urban societies, the crowding stress suffered by an individual may correlate with the degree to which he loses personal control over his environmental interactions. As examples, increased density may mean less ability to predict the length of a commuting trip, successive interruptions by noises of the act of falling asleep, and in due course, perhaps, anticipation of these stresses even if they do not ultimately occur [28].

In summary, it is clear that the effects of transportation and communication technology, and of urban size and density of population especially, have combined to make current civilized society significantly worse in many respects, and possibly unacceptably so. The original advantages of specialization and cooperation remain, but increased size and complexity have produced multiple, interwoven, multilevel hierarchies which severely constrain and alienate the individual citizen.

5.2.4. Optimality and Hierarchical Control

Implicit in the hierarchical structure is the concept that a system at any given level controls the performance of its several subsystems at the level(s) indicated below. This control is based on optimizing a pertinent performance index of the combined system. Unfortunately, it is much harder to infer quantitatively the pertinent performance indices at the various levels through a hierarchical structure, like that shown in Figure 5.1 for example, than it is to accept the qualitative concept. Indeed, it seems that most systems must operate satisfactorily in several different modes at different times, so that a satisfactory performance index must surely be some weighted average, in which the weighting of each individual parameter is based on its probability of occurrence and/or its survival value. Fortunately, some evidence is available for systems at certain hierarchical levels and will be summarized below.

A distinction must be drawn between (a) the optimization of the *design* of a biological system through the long-drawn-out evolutionary process, as typified by the mammalian vascular tree, for example, and (b) the optimization of the *operation* of a biological system, which may change significantly and frequently during the life of the organism. In the above example, the adaptive operating optimization of the cardiorespiratory operational parameters is highly pertinent. Of course, the system design for this optimization has also occurred during the long-drawn-out evolutionary process, but its "fast" adaptive capacity within a minute fraction of the organism's lifetime makes it functionally different.

(a) " Static " optimization of design. In the mammalian vascular tree used as an example, there are clearly many geometric configurations by which the blood vessels could be branched successively from the single aorta into the capillary network (of approximate order 10^9 capillaries) necessary to provide for gas and material exchange to all the organism's cells. Provided a pertinent performance index can be postulated, it would seem likely that one (or some) of these many configurations should prove optimal. Furthermore, if this putative optimal configuration should match that observed naturally, this would encourage belief in the validity of the optimization index postulated. Unfortunately, there is a uniqueness problem here in that the agreement does not *necessarily* validate the postulated index as being the one which the natural system has, in fact, optimized.

In the vascular system a pertinent performance index seems to be the combined metabolic cost to the organism of producing both the cardiac pumping energy for circulating the blood through the vascular system, and the maintenance energy for sustaining the various cells in the blood and in the vascular walls (see [29] for a current review). As a result of trading off these two cost components for a combined minimum, various optimality conditions prevail, in particular, the radius of a vessel, and the radii and departing angles following branching.

This design optimization is called *static* here in that it presumably remains essentially constant throughout the mature life of the organism. However, through evolutionary changes it may certainly be slowly modified during the history of the given species. Of course, there is also a *dynamic* control problem in that the design must be created during the growth of the organism, but unfortunately this process is not adequately understood to attempt modeling here.

As shown in Figure 5.8, there are further hierarchies of system organization, notably up to the design of the combined cardiorespiratory system. The same arguments about *static* design optimization would continue to apply;
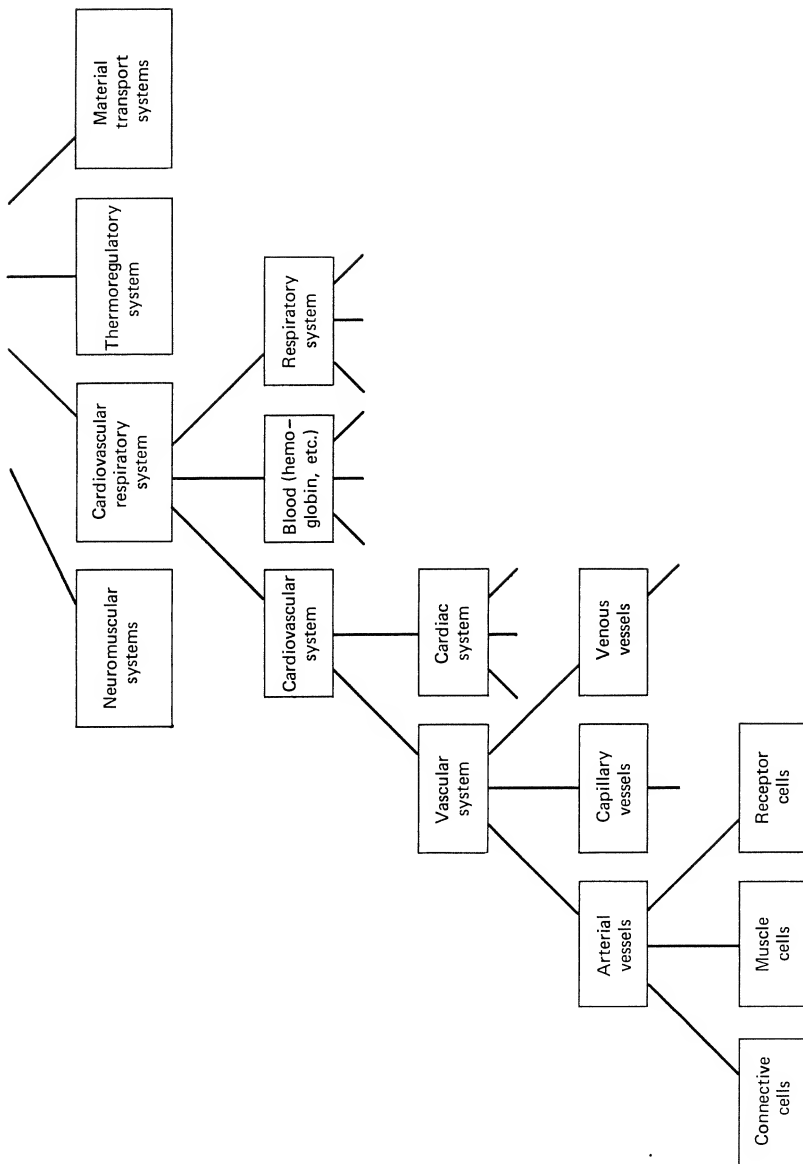
Figure 5.8. The cardiorespiratory system hierarchy.

166

as we shall show, however, *operating* optimization becomes an important function at the higher levels.

At supraorganism levels in animal societies, the same processes of evolutionary design optimizations would seem to have occurred, but they cannot be explored in this article. In any case, some aspects, such as territoriality and social hierarchy, have already been mentioned.

(b) Operating optimization through feedback control. Figure 5.9 illustrates in its upper feedback loop the minimal complexity necessary for homeostatic control. In the present example of the cardiovascular system (CVS) two
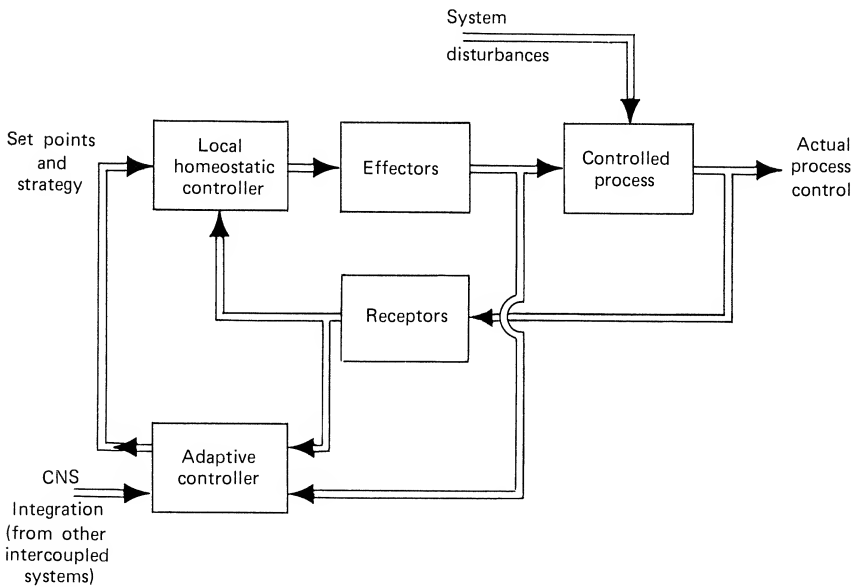
Figure 5.9. Negative-feedback homeostatic system, with adaptive control.

typical set points are systemic pressure and cardiac output. In general, there are multiple variables and components so that each information flow line is to be understood as a vector; for example, the heart and the vasomotor system comprise effectors of variable characteristics in the CVS. It is this multivariable nature of the control that gives rise to the possibility of some particular control strategy, from among the available set, being optimal.

For example, in the CVS the cardiac output (flow rate, Q) equals the product of the stroke volume (V) and the heart rate (f):

$$Q = fV, \tag{5.3}$$

while a similar relation holds for the respiratory system. There is evidence, reviewed in [29], that these systems can adaptively track any movement of the optimal strategy which may be required as a result of various system disturbances, in the sense that f and V can be varied reciprocally at constant Q until the condition involving minimal energy consumption is reached. The term *operating optimization* is used, since it can occur in the *real time* of an organism's living. As such, it seems to imply the necessity for an on-line adaptive feedback loop which is richer in information than that needed simply to ensure homeostasis (Figure 5.9) and is therefore, in the terms of this chapter, a hierarchical control.

Further hierarchical levels of this optimizing control can now be postulated; in particular, the combined operation of the cardiovascular and respiratory systems must be optimized with regard to their primary system task of achieving the necessary transport of oxygen to the organism's cells. Note that, as individual subsystems, their most economical operations would result from turning themselves down to zero output (Q). The key to an overall optimization lies, therefore, in the specification of such variables as the Q's. This involves a multiparameter optimization, including, for example, the important parameter of hematocrit [29].

The earlier comment about the multimodality of such systems can be amplified here. Thus the cardiovascular system has many other important transport functions, notably of nutrition (glucose, etc.), heat, hormones, and protective materials (white blood cells, etc.), while the respiratory system transports fluid and heat and in addition is utilized for other functions, such as the socially vital one of speech. Unfortunately, it is not clear at present how any optimality analysis could weight all these factors quantitatively.

We should also note that it may not be realistic to draw a sharp distinction between design and operating optimizations. Thus, for example, various organs such as heart and skeletal muscles increase or decrease in their capacity according to their state of use or disuse. This hypertrophy-atrophy aspect may perhaps be viewed as a continued use of those mechanisms basic to the original control of the organism's growth.

(c) Open-loop control or preprogramming. The basis for homeostatic and optimizing control shown in Figure 5.9 is provided by negative feedback. This makes possible the comparison between actual and desired conditions which is characteristic of conventional *closed-loop* control. However, we have noted previously that the evolutionary process is essentially *open loop* at one

level, in that particular systems either may or may not survive, according largely to their "genotypic" endowment. In other words, in its elementary form, the stage is set at the beginning, and the action plays out to success or failure, without control measures being available during the process. Of course, and in contrast, the larger but longer evolutionary process at the species and ecological level can certainly be considered to incorporate feedback action.

An important form of biological control can be operationally defined from these considerations, namely, an *open-loop* "*reflex*" *system*, together with a longer-term *closed-loop monitoring system*. The elementary escape reflex by which an undesirable contact produces a rapid (and possibly undirected) mechanical escape action provides a basic example, being used by many organisms from unicellular to the human. The reflex arc is characterized by a receptor to recognize the undesirable stimulus, followed by transmission to effector organs to power the escape movement. If satisfactory escape is not achieved by this "open-loop" movement, feedback is then effective, since the stimulus will still exist to trigger another reflex action. In a control theory context it can be considered as a form of *sampled-data* control (for more detail see [30]).

In many animals, including man, this basic reflex arc has been incorporated in an important variety of neuromuscular reflexes, sometimes in a more refined form involving some built-in feedback control, as, for example, in the neuromuscular stretch reflex. In any case, it has evidently provided a very effective and economic basic process in a hierarchical control context. The economy arises in that the necessary information processing is relatively small and discontinuous.

We may trace a whole hierarchy of ascending human functions based on the reflex arc. This hierarchy starts with the various levels of neuromuscular skills which the baby develops for feeding, moving its limbs and eyes, and locomotion. As exemplified by the fast eye movements (saccades), the control is characterized by a *preprogramming* in which the desired movement to the target is estimated and then the muscular action is released. If the open-loop action is insufficiently accurate to place the eye on the target, a second and correcting saccade is generated after an interval of about $\frac{1}{4}$ second [21]. It is important to emphasize that in a preprogrammed movement the control information is all generated before the movement is undertaken. In information-processing terms, it can then be released at a lower hierarchical level than the CNS's main processor, and therefore frees the latter for other and hopefully more important tasks.

This neuromuscular hierarchy continues to build up through very complex and refined neuromuscular skills, for example, the throwing, kicking, striking, and other highly skilled actions used in physical sports. However, most

notably perhaps, it builds up in the communication skills of human speech, writing, and behavior.

It is now possible to relate the preprogramming nature of developing neuromuscular reflexes and skilled actions to the developing of cognitive reflexes that may not result in immediate actions, but may rather be stored and processed only neurally in the CNS integrative centers. Thus cognition can be viewed as the result of preprogrammed neural reflexes and, furthermore, the intellectual processes of thinking, evaluating, and decision making. As one example, we note that racial discrimination results essentially from a preprogrammed reflex, which presumably had survival value at some point in man's history. This represents one example of a reflex which was appropriate in the levels of hierarchy then prevailing, but which now, at the higher levels of the present human social hierarchy, has become inappropriate. Perhaps the recognition that racial discrimination has a deep instinctual basis may allow us to undertake its exorcism more effectively.

Another aspect related to neuromuscular preprogramming is the *unfolding* characteristic of human *memory*, especially as related to tasks which involve significant time periods for their completion and also need to be repeated a number of times. We can note, as examples, the mental recovery of a route as we pass over it, in a just adequately anticipatory form to make turns, avoid hazards, and so on; and the continual mental unfolding of the score of a known symphony as time passes, in a similarly anticipatory manner. Note that by operating through a small but adequate *time window* of recovery the required information-processing capacity is reduced. The corresponding penalty that we cannot arbitrarily recall some particular segment of the whole memory trace at any given time has presumably not been of survival value to us, since usually we have only needed to recall in the unfolding sense of actually living the experience. Incidentally, we noted the hypertrophy-atrophy aspect of muscles resulting from their use-disuse, and an analogous facilitation-decay aspect exists for memory.

The mentally capable adult needs considerable flexibility in his use of memory, however, and, for example, becomes capable of "not losing his place" after an external interruption. It is interesting, therefore, to note that this facility is apparently gained only as a result of need and of considerable practice. Thus the young child must start his sentence or story all over again after being interrupted and of course finds this inefficiency very frustrating.

The same general form of preprogrammed reflex can be recognized at many human social levels in many different contexts, but we cannot pursue the matter here. As a closing comment, we should note that the claims just made about a developing hierarchy of reflexes, based successively on the lower ones already achieved, are entirely compatible with the somato-sensory-emotional hierarchies proposed by Erikson (Figure 5.4) and others.

5.3. CONSEQUENCES AND NEEDS IN OUR LIVING SYSTEMS

Up to now this chapter has emphasized the processes underlying the hierarchies of living systems. In this concluding section we shall focus on the consequences produced in our social systems and on the research aspects necessary if we are to work toward a viable world hierarchy, taking into account the different dynamic processes which seem *biologically imperative*.

5.3.1. Bandwagons, Initial Conditions, and Threshold

The inherent positive feedback implied in the saying "Nothing succeeds like success" characterizes the growth tendency in our living systems. However, the conditions under which growth first starts, and later under which it ceases, are of considerable interest to us.

First, the forces tending to initiate change are often effective in a *multiplicative* rather than an *additive* way, and thus it is often useless to seek the *primary cause* for a change. Unfortunately, in our sociopolitical systems we often try to do precisely this when the growth is in an undesired direction. This seeking for simple answers to complex questions can result, for example, in the finding of *scapegoats*. Second, once the growth has started, the self-sustaining nature of the process may render irrelevant even those initial forces which can be identified, that is to say, there is no point in trying to find and remove the match with which the fire was started. Third, a threshold is often present in either of the following senses.

1. The system has some stability which resists the initiation of change until some threshold of forcing stress is reached, after which growth seems to occur explosively. Certainly once the threshold has been passed and the response released, it cannot typically be annulled immediately. The initial subthreshold depolarization, followed by the spike action potential in the neuron, provides an interesting analogy.

2. Alternatively, the initial growth occurs smoothly but is not perceived by society until some threshold level of activity has been reached.

Once the *bandwagon* stage of established growth has started, increasingly large forces and/or changes in the system processes become necessary in order to reverse the growth. If any such control or restructuring activities are intended, it is obviously important that the planners have a good quantitative understanding of the dynamics of the process. This understanding should also include the end part of the growth process, that of stagnation or decay (Figure 5.6), since reasonably accurate prediction of it may sometimes permit

the conclusion that the process can be allowed to run its course harmlessly. In regard to this end stage, Boulding's pertinent remark is, "Nothing fails *in the end* like success."

These comments would seem to imply that society's general need is to retard exponential growth and to stabilize the system. Certainly the general "good" will tend to be perturbed continually by small "selfish" perturbations, but on many occasions change may be perceived as necessary by an increasing part of the society. In this case, the comments of this section can be taken the other way round to indicate effective strategies for producing change.

5.3.2. Growth, Decay and Cycling

The various large living systems generally appear to the superficial observer to be either statically maintained or recently growing more or less steadily. However, within any such system there is an underlying hierarchy of cyclic processes. For example, at the biochemical level there are the Krebs cycle and the hydrogen-ion transport system. At the level of the cell, the basic viable unit of life, an interestingly diverse situation holds. Thus in man some cells, notably epithelial and blood, are in a steady flux of dying and replacement, with relatively short life spans of the order of one month. On the other hand, the life-span spectrum includes some cells, notably those of the CNS, which are not replaced, even though they are subject to a small percentage mortality rate throughout the individual's life. Finally, at the level of our societies, the human being is the basic viable unit, so that an underlying renewal cycle must exist with a period of one generation, namely, some 25 years.

It is interesting, therefore, that the various hierarchies of man's societies seem to be characterized by further and longer growth-decay processes. Thus Forrester [31] finds that the modeling of a city involves a growth to maturity in some 200 years, typically with a tendency to decay thereafter, unless some new stimulus to growth exists. The growth of empires has followed a similar pattern but over a somewhat longer time [32]. Incidentally, Price [19] has shown that the growth curves of many social system aspects consist not of a single sigmoidal pattern, but rather of several such patterns superimposed as new stimuli to growth occur successively, typically because of technological breakthroughs.

Since a complete growth-decay process in any one system is usually coupled with similar processes in competing and/or cooperating systems, the overall system may appear to be cycling. This can be modeled by successive repetitions of the epidemic growth curve in Figure 5.6. Now cycling (or oscillations) appears to be a very general phenomenon in living systems, perhaps to a

much greater extent than is justified by the inherent growth dynamics just presented. Thus within the human organism there are cycles based on events at the following levels, among others: capillary, cardiac, pulmonary, circadian, water balance, menstrual, and annual. Many of these could be characterized as *relaxation oscillations* in control theory terminology [29], with the implication that such theory may be helpful in designing any attempts to harness or control such cycles.

It is certainly tremendously important for us to try to understand the growth and oscillation processes within both ourselves and our societies. Especially must this knowledge be applied at the political level. In this regard we note that the typical term of elected office, whether in city or in national government, is only a few years. Now, on the one hand, there are good reasons for this in regard to the psychological characteristics of men in power; on the other hand, however, the shortness of the period is not well matched to the underlying time constants required for the necessary long-term policies to be formulated, implemented, and accepted by society.

In particular, we must devise new ways to organize our various hierarchical societal levels so that physical growth is not a built-in criterion of success. Indeed, the greatest challenge may be to engineer *decreases* in some major societal variables, such as population and certainly pollution levels, without incurring social demoralization. This problem may not prove insuperable, however, if we can learn to understand and then harness the competitive and cooperative processes underlying our societal system. In any case, a very important aspect will be to re-emphasize personal attention to our own experiential hierarchy, so that as a society we become much more inner-directed.

5.3.3. Purpose, Striving, Unfolding, Reverence, and Leadership

(a) Purpose and striving in individuals. The comment has already been made that living systems at almost any level seem to behave *as if* they had a purpose, even though there may be no central intelligence in the system where this purpose is stored or any issuing of control commands which make the system work toward the purpose. Man thinks himself unique in that he tries to articulate his purpose, but unfortunately a purely intellectual evaluation often evokes the pessimistic judgment that no rational purpose exists. Fortunately, at least from the optimistic view, man's own biological hierarchy imperiously requires him to strive as if he had a purpose, and in the normal healthy human being this effectively produces purposeful living and striving.

This argument suggests a biological basis for our sense of purpose, or performance index. Now indeed the various recent discoveries concerning sensory deprivation and neural reward: punishment centers strongly suggest

that continued rich stimulation of neural centers is vital for good performance of the human system. In particular, the inputs from sensory organs and muscular activity and the self-inputs of cerebral activity are very important. Thus the concept of a top-hierarchical neural center suggests itself, where the various inputs are weighted in order to evaluate overall performance; this would then constitute the level of human "satisfaction." In this context no person is irrational; instead only the particular weightings adopted by the psychologically "not-well" individual are significantly different from those accepted by "normal" human beings. Again these comments are compatible with the body-mind hierarchy of performance implied in Figure 5.4. Thus it is important to have a well-functioning muscular system in order to provide the basis on which further hierarchies of experiential satisfaction can be constructed.

Maslow has pioneered an insightful theory concerning human motivation, purpose, and striving [12, 33], some aspects of which are implicit in the foregoing comments. Several of his basic points must therefore be stated [12] and amplified in the present context. The work of some other persons along similar lines has also been mentioned [13, 14].

1. The human being has intrinsic psychological needs as important as his physiological needs.

2. These needs are related hierarchically, in such a way that the fulfillment of one need tends to allow the next one at a successively less basic level to be tackled. Some important increasing-level needs are food, safety, love, and self-actualization. In particular, the first two are physiological (D-needs) and the last two psychological (B-needs).

3. There seems to be a single ultimate goal toward which human beings strive: *becoming* as fully creative, human, and self-fulfilled as each person *can* become.

4. To the aspiring human individual involved in his climb through this hierarchical set of goals, it usually seems that his ultimate goal is the one toward which he is currently striving. However, once he has struggled to this summit, a new and higher one begins to be visible up ahead through the mist. (This limited window of vision is reminiscent of the time window mentioned in regard to memory recall in Section 5.2.4(c).)

5. The struggle to achieve each successive hierarchical goal is sufficiently and increasingly self-rewarding, so that we do not need to be frustrated by the apparently unachievable height of an ultimate goal. *Becoming* provides *peak-experience* rewards at successive transient states of *being*; or, in Maslow's phrasing, "If they are not mountain-peak-experiences, at least they are foothill-experiences, little moments of absolute, self-validative delight, little moments of Being" [12, p. 154].

6. Our physiological and psychological capacities "clamor" to be exercised. Their well-use is satisfying and produces harmony and health; their disuse is irritating and frustrating, and may even produce a disease center. The term *well-use* could perhaps imply the tendency to operate optimally, as already explored at the neuromuscular level, for example.

7. Man's higher nature and aspirations are founded on his lower nature and would collapse without it. Therefore, to develop his higher nature, man must first fulfill his lower nature, that is, he must gratify rather than renounce his instinctual needs.

8. In the growth of the child the normal, desirable pattern is to feel initial delight with each newly acquired skill, but then to become bored and "eagerly" to strive toward new, more complex delights without feeling unsafe. Note that in terms of *skill* these successive growths correspond to successive superimposed sigmoid curves (Figure 5.6), while in terms of *delight* or *satisfaction* they correspond to successive "epidemic" growths (Figure 5.6). The latter is approximately the derivative of the former, and indeed this relation is expressive of the likelihood that the satisfaction characteristic at the neural integration level will be *adaptive*.

9. One sentence from Maslow [12, p. 55] summarizes well: "The single holistic principle that binds together the multiplicity of human motives is the tendency for a new and higher need to emerge as the lower need fulfills itself by being sufficiently gratified."

(b) Unfolding. The unfolding aspect of our individual lives is also deeply involved with our purpose and striving. Our life unfolds irreversibly as we move through the cycle of birth, maturity, decay, and death. Although we can to some small extent look forward, and can certainly learn from others already older, the aperture of our experiential time window effectively covers only the regime from the present back through an increasingly filtered past, a "rear view mirror," in McLuhan's term. In consequence, each individual necessarily exhibits an asymmetry of outlook, dependent on the stage of the life cycle he has presently reached. In particular, no individual can anticipate, perhaps fortunately, his future outlook as his experience irrevocably grows. Thus, like his neuromuscular skills, his philosophy grows according to the challenges he has accepted, and most inappropriate solutions cannot be recognized as such without having first been tried. One result is the inevitable tension and discord across generations. Unfortunately, these have become less containable in recent years because the explosive extent of technological change has resulted in members of the young generation growing up in such a different world that they feel justified in questioning and largely rejecting the conventional integrated wisdom of their elders. Although this is understandable, it would involve inherent danger if they were allowed total

responsibility for society, while the mature generation abdicated its role. The individuals in each new generation will inevitably mature as their lives unfold and will then tend to accept many previously unacceptable concepts. In the interim, the "torch" must be carried on as well as possible.

In order that youthful lives can unfold in a good biological way, experimentation must be facilitated; but to whatever extent it is irresponsible it should also be peripheral to the mainstream of society's management. Hence it is pertinent to note that our mature professionals learn many of their new skills through "irresponsible" simulators, or games; for example, the pilot first learns to fly a new airplane on a simulator. Here he may crash it repeatedly and harmlessly in order to learn the limits of his viable control ranges. In the same way, good winter drivers are not those who never skid, but rather those who do so purposefully when they can find "safe-play" conditions. By thoughtful development of new simulators appropriate especially for the adolescent, and by blending them into the real situation whenever practicable, much present intergeneration frustration could perhaps be removed, and better maturation ensured.

This unfolding of the lives of successive generations imparts an effective cycling to the larger social system. However, within the individual's unfolding life there are certain faster renewal cycles. One of them seems to be an *interest* cycle, lasting in the order of 4–6 years. Thus the challenge of a new situation is typically accepted, mastered, and used up in this period. The actual characteristic is something like the epidemic growth curve of Figure 5.6, perhaps because of the same underlying functional processes. Note also that this interest cycle undoubtedly relates to the social interest cycles in such areas as fashions.

Another important example is the annual climatic cycle, to which most of mankind has evolved. Although many would superficially welcome elimination of this cycle, the history of discovery and social advance has largely been shaped by the more competitive people from relatively rigorous cyclic climates. It is not necessary to make value judgments on this past in order to recognize that we should be very wary of casually developing the apparently ideal weather-controlled environments which are now technically feasible, for example, through domed cities.

(c) Purpose and striving in society. These various concepts about the individual human being have some immediate general applicability to the several hierarchical levels of human society, namely, society's psychosocial versus material needs; the tendency of a society to develop increasing levels of aspiration; the frustration of a chaotically disorganized, ineffective society versus the attractiveness of banding together for a common purpose or challenge (even the destructive one of warfare); and, in general, the relevance of such concepts as the national mood, the national objective.

One important current failing is that challenging but achievable social goals are not being set, and certainly not being set into a hierarchy of continuing social needs. Indeed, the short renewal cycle of our political electoral process contributes to the difficulty of doing so, as already noted. The difficulty is accentuated in the industrially developed Western nations, especially in the United States, by the fact that almost all of society's material (D, or deficiency) needs are perceived by the managers of society as having been fulfilled. Then, since no higher B-needs have been articulated, the nation has no worthwhile goal sequence to motivate it. Hence many subnational groups are forced to seek their own goal sequences, and these are often mutually incompatible. Again we note that the subsystems themselves cannot perform system optimization without first being provided with some minimal mutual goals and restraints. Developing countries which have recently undergone social revolutions are more likely to achieve an initial national consensus on goals, especially because these goals typically involve catching up to the known level of competitors. It is much more difficult for the runner who is out in front to pace himself.

Another problem of our society is the tendency to provide material affluence for individuals at increasingly young age, without requiring them to strive to achieve these goals in successive stages as their lives unfold. Thus the current tendency is to fulfill such objectives as new cars, furnished homes, and " exotic " vacations immediately upon entering working life, through advance credit financing. Since healthy growth seems to depend on the successive unfolding of growth needs, this process of "fulfillment" without effort could prove disastrous to the emotional growth and maturation of the individual recipients and their society.

On the other hand, the unremitting " rat race " to climb essentially arbitrary " success " ladders is a no less satisfactory aspect of our society. Clearly, we must give much thought to structuring our institutions so that a sequence of attractive but challenging goals can be chosen by each member of society, appropriate for his personality and abilities. In general, this sequence should probably be arranged to peak in late middle age, and such strategies as then developing " second careers " should be socially encouraged.

(d) Reverence. Historically, the development and codification of human purpose has been entrusted in considerable measure to religion. Religion has typically evolved " transcendentals," or revealed truths, about the nature, purpose, and right behavior of man. Among other aspects, these encourage an attitude of reverence, which seems very salutary for healthy psychic growth in man [34]. As Fromm [1] notes, these revealed truths are essentially myths which express the " experiences of the soul " in symbolic language. Since they are largely prescientific in origin, these myths have apparently been proved invalid by successive new scientific advances. The conflict engendered has

been sufficiently strong that even in recent times it has been believed that scientific advance may have to be suppressed in order that the transcendentals necessary for psychosocial health may continue to prevail (see, e.g., [35]) However, once it is realized that the intended truth of the myth lies in its *latent* rather than *manifest* content, this conflict becomes largely avoidable. With the resolution of such conflict, science can prove a valuable cooperator with religion. Thus, in particular regard to the aspect of reverence, it is important that the discoveries of science and the search for them all tend to encourage a feeling of reverence in the scientist for this marvelous, almost infinite system which is nature. No matter how many peaks of discovery are scaled, a new range of them always becomes visible up ahead, and one dare not even surmise where the peak of ultimate truth may be.

(e) Authority and leadership. Unfortunately, this potential cooperation from science requires from religion and philosophy an understanding of the concepts and methods of science which has not generally been forthcoming. Since science and technology are the agents most responsible for the current massive modifications in our society, undoubtedly the scientist and the technologist must make major efforts to "bridge the gap" separating them from their humanist colleagues.

With religion, and its revealed truths about the purpose and conduct of life, have been associated *authority* and *leadership*. As already noted, this authority has historically evolved into fairly rigid hierarchical structuring, however irrelevant this may have been to the founder's philosophy. Indeed, this point provides another example in which the inherent growth dynamics of a process renders the initial condition not very important to the outcome (Section 5.3.1).

With the development of human knowledge and technology, the natural base of social authority has moved overtly to the secular, political process, involving variable degrees of democratic participation. In consequence, the leaders must now be selected largely through their own characteristics, and this raises many questions for society to resolve in order to obtain leadership of the kind it needs, especially since the type of leadership required is changing.

In regard to the hierarchical theme of this chapter, we need only inquire here into the leadership effects of the random innate inequality of human beings. Certainly society should enable each person to find his own appropriate experiental hierarchy and to work in it. The spectrum of such hierarchies will undoubtedly require much continued organization of society and therefore the selection of leaders. Now an unresolved problem is that these leaders, by whatever process they are currently selected, are not usually motivated toward the self-actualizing society to which we would ideally aspire. Instead, they may rather be motivated by the wish to exert personal power, as an end

in itself. Indeed, there is the complementary tendency that the B-need oriented person, who can already fulfill himself, would rather not become involved, and possibly sacrificed, in the political process of managing society. The result has been that the societal framework is generally more conservative and less experimentally minded than the innately superior members of society would have organized had they exerted their natural leadership [36].

Indeed, this again raises the difficult problem, which is probably rationally unresolvable, of how the needs of the different components of society should be weighted. Thus, although there is a tendency in the long run for societies and natural leaders to select each other mutually, it has remained politically unfeasible to enforce such a process. In this regard, we should note Maslow's tentative suggestions [12] that healthy people probably choose on the whole what is "good for them" in both the physiological and the psychological senses, and that these choices would also be good for the others, were they better choosers. He then proposes "that we explore the consequences of observing whatever our best specimens choose, and then assuming that these are the highest values for mankind" [12, p. 169].

5.3.4. Optimization, Tradeoff, Cost : Benefit, and Discord

We can summarize the optimization process (Section 5.2.4) as follows:

1. Optimization involves searching the available parameter space for the extremum of a relevant *performance index* (e.g., maximizing the profits or minimizing the costs).

2. The performance index comprises a *weighted* set of *component indices*, and most of the difficulty in system optimization theory concerns the appropriate allocation of these weights.

3. Once an optimization has been achieved, there is still the *adaptive* problem of finding a control strategy by which to track any movements in the optimal condition due to system changes or disturbances.

4. In human systems especially, an important aspect is the extent to which degrees of freedom in optimization are available at the lower hierarchical levels, for this deeply affects the motivation level of the human being and hence his performance.

In practice, optimization involves a *tradeoff* process in which, as manipulatable parameters are changed, some of the component indices decrease while others increase. This implies that the optimal solution is not an absolute, unchangeable imperative, but rather a negotiated compromise. Clearly, the bargaining of trading nations, of management and labor, and of psychological transaction theory exemplifies this process. However, our societies are still

tempted to lay down principles which must be honored "at all cost." In fact, however, such honoring would clearly be absurd and is not practiced. As an example, the optimal value of a city's crime rate cannot be zero, even if some moral principles dictate that crime is evil and must not be tolerated. Instead, a tradeoff must be practiced between the increased social and economic cost when civil "protection activity" is increased, and the corresponding decreased cost of the crimes committed.

Actually, the existence of significant mismatch between the idealized and the real indices may indicate that the optimization is being performed on an invalid basis or on an incomplete system. In this case the whole spectrum of connected systems should be evaluated, from the possible uses of technology through the advancement of educational and cultural aspects.

Cost: benefit analysis provides an important technique for evaluating what the members of society consider the value to them of any given system compared to its cost. Many of the data are necessarily empirical, however, since no theoretical modeling yet seems possible. Thus again we note that an ideal such as perfect safety is not sought by society in practice. Specifically, Starr [37] has shown that the underlying statistical risk of death from disease apparently determines society's psychological yardstick for other risks. For example, individuals accept "voluntary" risks (as in sports) at a rate some 1000 times higher than this basic "involuntary" rate.

Another implication of the tradeoff process is that a nonzero amount of *dissatisfaction*, *discord*, or *frustration* must remain even at the optimal condition. Now there are many basic discords in any society, partly because of the lack of complete communication across groups, and partly because of their necessarily different objectives. Thus we note particularly the discords between the different classes of race, generations, sexes, urban-rural dwellers, economic conditions, and workers. On the one hand, at extreme levels of discord, it is undoubtedly true that an old social system is proving unsatisfactory and will undergo a "hierarchical restructuring," in Platt's term [18], and our current society may well be approaching this point [38]. On the other hand, moderate levels of dissatisfaction are not undesirable in that they can provide the graded negative feedback necessary for corrective action to be undertaken. Indeed, as such they represent important stimulating inputs for the system, whether for the single individual or his society, and the system must learn to use them creatively in order to be optimal. In this regard, note the earlier comment that the human being needs a rich neural stimulation to remain alert.

Moreover, if all known frustrations and discords could be removed from a society or an individual, it seems entirely probable that each would invent new ones in order to prevent such dull agreeableness. Perhaps this helps to explain why few people ever seem really motivated to build and sustain the putative utopias which have been proposed throughout our history!

5.3.5. Simulation, Gaming, Computer Utilities, and Reduction of Hierarchy

Society's problems are so immense and variegated that no overall utopian blueprint seems conceivable, even if such were desirable. However, scientific discovery and technological application make possible some new approaches, especially modeling techniques utilizing electronic computers. Furthermore, when the information-*processing* abilities of computers are teamed with (a) the information *transmission* abilities of telephone, microwave, and satellite, and (b) the information *display* and inputting techniques associated with television, the computer *utility* (social information utility) emerges as an important potential tool for implementing our new societies. In particular, it must be emphasized again that our transition to an *information-rich* society opens up for the first time tremendous possibilities for the members of this society to become self-fulfilled individuals enjoying previously impossible freedom of access to knowledge, instruction, entertainment, and other needed resources and stimulants. Of course, this potentiality poses tremendous new educational problems to be solved before the resources can be fruitfully exploited by society. Indeed, there is the real danger of these potentialities being subverted, and of society becoming regimented in a rigid hierarchy under total electronic control. However, such alternative outcomes of greater good or greater "bad" have existed throughout history in exploiting each new technological breakthrough.

One hopeful strategy for achieving the good result is to involve the individuals of society in its study and improvement. Specifically, the advent of simple programming techniques makes it possible to generate *simulation models* of large systems relatively easily. These can then be used for *gaming* studies by almost any interested group of individuals, as well as the professional planners and managers for whom gaming is becoming a primary tool. As a result of "hands-on" experience they can develop a feel for and involvement in their own society, at any desired hierarchical level. Indeed, the desirable next step is then to carry on in the classical scientific method:

1. Study the available imperfect model of the system, and game with it to discover various possibly desirable strategies, as well as improvements in the model.

2. Carry out the implied experiment in the group's community, with the full understanding of the community that this small-scale experiment may not prove successful. On the other hand, since an amateur competitive game is being engaged in, the overall outcome is that the results of a set of parallel "best-guess" experiments can become available to the larger community. One advantage of these small-scale pilot experiments is that no great capital investments are involved. But more importantly, the exciting sense of involvement and challenge should benefit the individuals of society tremendously.

3. Continue the iterative cycle of improved simulation and experimentation as long as is desirable. Clearly, the whole process is open ended, since worthwhile possible improvements are endless and continue to unfold as each new improvement is incorporated.

Another social problem of hierarchical systems can be tackled by these techniques, namely, the difficulty, on the one hand, that if subsystems optimize themselves without regard to the total system needs such nonoptimal conditions as " mutual ruin " may occur; and, on the other hand, the difficulty that subsystems comprising living persons cannot remain challenged and motivated unless they can retain considerable degrees of freedom for optimization purposes, as well as the constraints which must inevitably be sent down. Computer gaming by subgroups will encourage them to look at their unit interactions with the whole system, and therefore to begin to accept some responsibility for the fate of the whole system.

In this way the subgroup will embrace the vision and responsibility of several hierarchical levels at once. Perhaps this will provide a working approach to how our society's structure may be reduced in the extent and number of its effective hierarchal levels, and yet remain viable. Indeed, it is interesting to note that this reduction in hierarchy would make human society functionally more like ecologies, which, as we have already noted, behave *as if* they were hierarchically controlled, although in the daily routine sense they operate as individual units. The computer gaming experience would provide the equivalent of the evolutionary process by which the " right " behavior is evolved in ecological systems so that each subsystem operates harmoniously for the common good in the process of fulfilling itself.

The potential advantages of gaming can be summarized as follows:

1. The cultivation of an experimental outlook in our society can reduce alienation and increase the individual's involvement in his society.

2. Our material affluence can afford us the luxury of many experimental probes coming up with unsatisfactory solutions. In any case, no great capital is tied up in any single probe.

3. The conduction of multiple small experiments in parallel makes it likely that good solutions will be arrived at faster than if one supreme group were entrusted with finding the optimal solution and then imposing it on all. Furthermore, the obtaining and implementing of multiple good solutions will provide the diversity (*requisite variety*, in Ashby's [39] term), which improves the survival probability of the system when the environment changes. In other words, many options are kept open.

4. The competitive coexistence of these diverse good solutions can help prevent the bandwagon effect from taking over, and therefore can keep society from becoming too monolithic.

5. The fruitful use of the social information utilities can ensure that the relatively free subsystems operate harmoniously with, and in full knowledge of the needs of, the other competing/cooperating subsystems.

6. This whole process is suitably biological in that it constitutes a form of socially directed and accelerated evolution. The fact that we can use our intelligence for prediction and serendipity are only added advantages.

7. The harsh exclusion characteristic of biological evolution can be experimentally modified with some reasonable confidence when enough experimentation has been performed. In particular, we should be able to tackle the festering problems of increasing disparity, for example, as between developed and underdeveloped nations, and between races and classes within one country.

5.3.6. Conclusions

The concept of a hierarchical society is under vigorous attack, because of its perceived rigidity and consequent hampering of man's efforts to realize himself fully. Whatever may be the truth in the myth that a matriarchal society preceded the patriarchal one of our recorded history [1], there is now much pressure to discard the patriarchal principles and embrace the matriarchal ones. For example, Slater [38] writes, "The old culture, when forced to choose, tends to give preference to property rights over personal rights, technological requirements over human needs, competition over cooperation, violence over sexuality, concentration over distribution. . . . The new counter culture tends to reverse all of these priorities."

This chapter has advanced the thesis that competition and cooperation are vital complementary processes in a hierarchical structure which remains essential for any viable society that accepts technology. However, this acceptance does not deny that society must direct and control the technology so as to receive only that part which is judged to contribute the maximum benefit. In justification of this view, we have claimed that the basic hierarchical structure has been so efficient that evolution has incorporated it in the whole spectrum of living systems, up to and including human experiential growth. As such, we would now discard it completely only at our great peril. However, by encouraging a multiple experimental approach to our society, especially through intelligent use of our new computer tools, it seems possible that a satisfactory new structuring will be discovered. Although the pattern can at present be only dimly discerned, it would seem to consist of a functionally flexible hierarchy operating with such information-richness that the individual can enjoy not only responsibility for controlling the system, but also adequate freedom to work toward his personal fulfillment.

ACKNOWLEDGMENT

PROBLEMS

5.1. Consider a simple hierarchy in which each lower level has 10 times as many elements as the level above; furthermore, each element at any level has 10 elements converging on it from a lower level. Let the " 10-to-the-zeroth " level be the top, where total convergence has occurred to 1 element; the nth level below therefore has 10^n elements. We now deal with two informational structures:

Type (i): Each element processes the information converging on it from its 10 elements, and then passes information on up the hierarchy. For simplicity, assume that the upward information flow rate of every element is the same, b bits per second, independently of hierarchical level.

Type (ii): Only the elements on the lowest level (N) generate information (again b bits per second), and the elements of the higher levels operate only as "postboxes," that is, pass on but do not process the information.

The same conditions can apply for "downward"-flowing information, and the computations to be performed are not affected by including these flows also.

(a) For three hierarchical levels $(N = 3)$ compute the total information rates, B_1 and B_2, in the system for types (i) and (ii), respectively. Here B is to be computed as the bits per second transmitted across a single hierarchical stage.

(b) Find how the ratio (B_2/B_1) behaves as a function of N, and in particular show that it tends to $0.9N$ as N increases.

(c) Generalize your results for any arbitrary convergence factor, C, rather than for 10.

(d) Consider further aspects in the modeling, such as the costs of information processing in the two methods, in order to make the model more realistic.

5.2. Consider more fully the comment in [8] that the rationale for a hierarchical structure is to provide a way of avoiding "direct confrontation between the large and the small in nature." Is the tendency of social organization to generate a new "executive committee whenever the basic committee becomes large" (p. 152 and [11]) an illustration of this principle?

Establish for as many hierarchical examples as you can the approximate convergence factors which apply. Is the convergence factor typically constant within a hierarchy, or does it depend on the relative information rates to be processed?

5.3. The hierarchy of the food web is quoted as having an approximate energy convergence ratio of 10 [16]. Does energy flow indeed represent the appropriate information measure for making the food web hierarchy analogous to the other information hierarchies? Do such other measures as biomass and numbers of organisms have any greater pertinence?

5.4. Nominally an organization such as a factory follows a strict hierarchical structure. Make reasonable estimates for the convergence factors of the following levels: individual machinist, charge hand, foreman, manager. Furthermore, guess at the orders of magnitude of information rates up and down these hierarchies in normal operation. Assume initially that all information flow is "vertical"; then guess at the modified information flows when the realistic "horizontal" information flows are included.

5.5. The basic information flow for industrial production-inventory-sales could once be considered hierarchical with all or some of the following levels: head office, production plants, warehouses, wholesalers, retailers, customers. Unfortunately, the nominally efficient feedback of sales information upward through the successive hierarchical stages, in fact, provided at least one reason for a tendency to instability in the inventory-production control system. In detail the time lags of each stage can produce such instability in a closed-loop system. This problem has been solved in simple but effective ways. Discuss some of these ways, and their significance for the effectiveness of hierarchical structures. Could the dynamic instability exhibited in this example apply to any of the previous problems for which stable equilibria have been implicitly assumed?

5.6. Another aspect of hierarchical information flow is seen in our neuro-muscular organization especially. Thus preprograms seem to be developed at the lowest viable level for skilled but repetitive actions—for example, the muscle actions necessary for locomotion (whether ambulation, swimming, or flying), eye movements, and speech. Again, the retina does much preprocessing of visual images. Indeed, more generally the neural receptors often seem to have the characteristics which are already essentially those needed for the pertinent systems they will control (e.g., warm skin thermoreceptors vary their signal with temperature in almost the same way that the muscles subsequently respond in producing heat by shivering). In general, this amounts to following a principle of decentralization; in computer analogy we would say that nature uses small dedicated computers with relatively fixed programs at many decentralized sites. Thus the time-sharing information-processing

demands at the more central computers are minimized. Does it seem that this difference is due to the different "technologies," or is perhaps some basic organizational principle illustrated here? In your considerations note also that one of nature's important criteria seems to be that, in the event of breakdown higher up in the system, subsystems at each lower hierarchical level should remain viable.

REFERENCES

1. **Fromm, E.,** *The Forgotten Language.* New York: Grove Press, 1951.

2. **Bachofen, J. J.,** in *Der Mythus von Orient und Okzident,* M. Schroeder (Editor). Munich: Ch. Becksche Buchhandlung, 1926.

3. **Dechert, C. W.,** "Political and International Systems." In *Positive Feedback,* J. H. Milsum (Editor). Oxford: Pergamon Press, 1968.

4. **Maruyama, M.,** "A Democracy Model Based on a Non-Western Epistemology." *Cybernetica,* pp. 214–220, 1969.

5. **Mead, M.,** *Co-operation and Competition among Primitive Peoples.* Boston: Beacon, 1961.

6. **Ardrey, R.,** *The Social Compact.* New York: Atheneum, 1970.

7. **Mesarovic, M. D., D. Macko,** and **Y. Takahara,** *Theory of Hierarchical Multi-Level Systems.* New York: Academic Press, 1970.

8. **Whyte, L. L., D. Wilson,** and **D. G. Wilson,** *Hierarchical Structures.* New York: American Elsevier, 1969.

9. **Weber, G., R. L. Singhal, N. B. Stamm, M. A. Lea,** and **E. A. Fisher,** "Synchronous Behaviour Pattern of Key Glycolytic Enzymes: Glucokinase, Phosphofructokinase, and Pyruvate Kinase." In *Advances in Enzyme Regulation,* G. Weber (Editor), Vol. 4, pp. 59–81. Oxford: Pergamon Press, 1966.

10. **Waterman, T. H.,** "Systems Theory and Biology—View of a Biologist." In *Systems Theory and Biology,* M. Mesarovic (Editor), pp. 1–37, New York: Springer Verlag, 1968.

11. **Parkinson, C. N.,** *Parkinson's Law.* Boston: Houghton Mifflin, 1957.

12. **Maslow, A.,** *Towards a Psychology of Being.* Princeton, N.J.: D. Van Nostrand, 1962.

13. **Graves, C.,** "The Deterioration of Work Standards." *Harvard Business Review,* Vol. 44, September—October 1966.

14. **Erikson, E. H.,** *Childhood and Society.* New York: W. W. Norton, 1963.

15. **Eccles, J. C.,** "Long-Loop Reflexes from the Spinal Cord to the Brain Stem and Cerebellum." *Atti della Accademia Medica Lombarda,* Vol. 21, No. 3, 1966.

16. **Slobodkin, L. B.,** *Growth and Regulation of Animal Populations.* New York: Holt, Rinehart & Winston, 1966.

17. **Kuhn, T. S.,** *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1970.

18. **Platt, J.,** "Hierarchical Restructuring." *General Systems Yearbook,* Vol. XV, pp. 49–54, 1970.

19. **Price, D. J. de Solla,** *Science Since Babylon*. Princeton: Yale University Press, 1961.

20. **Hardin, G.,** "The Competitive Exclusion Principle." *Science*, Vol. 131, pp. 1292–1297, 1960.

21. **Milsum, J. H.,** *Biological Control Systems Analysis*. New York: McGraw-Hill, 1966.

22. **Jerison, H. J.,** "Brain Evolution: New Light on Old Principles." *Science*, Vol. 170, pp. 1224–1225, 1970.

23. **Rapoport, A.,** *The Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press, 1965.

24. **Hardin, G.,** "The Tragedy of the Commons." *Science*, Vol. 162, p. 1243, 1968.

25. **Christian, J. J.,** "Social Subordination, Population Density and Mammalian Evolution." *Science*, Vol. 168, pp. 84–90, 1970.

26. **Freud, S.,** *Civilization and Its Discontents*. New York: W. W. Norton, 1962.

27. **Selye, H.,** *The Stress of Life*. New York: McGraw-Hill, 1956.

28. **Konrad, K. W.,** *Crowding Stress*. Personal communication, 1970.

29. **Milsum, J. H.,** and **F. A. Roberge,** "Physiological Regulation and Control." In *Textbook of Mathematical Biology*, R. Rosen (Editor). New York: Academic Press (in press).

30. **Milsum, J. H.,** "The Role of Unidirectional Rate Sensitivity in Biological Control System Stabilization. In "Rein Control or Unidirectional Rate Sensitivity," M. Clynes (Editor). *Annals of the New York Academy of Sciences*, Vol. 156, Art. 2, 1969.

31. **Forrester, J. W.,** *Urban Dynamics*. Cambridge: M.I.T. Press, 1969.

32. **Taagepera, R.,** "Growth Curves of Empires." *General Systems Yearbook*, Vol. XIII, pp. 171–176, 1968.

33. **Maslow, A.,** *Motivation and Personality*. New York: Harper, 1954.

34. **Maslow, A.,** *Religions, Values and Peak Experiences*. Columbus, Ohio: Ohio State University Press, 1964.

35. **Weaver, W.,** *Ideas Have Consequences*. Chicago: University of Chicago Press, 1948.

36. **Comfort, A.,** *Sex in Society*. London: Penguin Books.

37. **Starr, C.,** "Social Benefit versus Technological Risk." *Science*, Vol. 165, pp. 1232–1238, 1969.

38. **Slater, P.,** *The Pursuit of Loneliness: American Culture at the Breaking Point*. Boston: Beacon Press, 1970.

39. **Ashby, W. R.,** *An Introduction to Cybernetics*. London: Chapman & Hall, 1956.

6. A Systems Approach to Epistemology

WALTER BUCKLEY

Department of Sociology, University of New Hampshire, Durham

6.1. INTRODUCTION

The focus of this chapter, by a sociologist, on some perennial problems traditionally dealt with by the philosopher reflects the current critical state of sociology and some other social sciences. Questions are being asked, and demands made on this area, that transcend its current level of traditional theory and methodology. The latter are thus under critical attack on many sides, and a basic change in the underlying theoretical paradigm of the field (to use Kuhn's term [1]) is blowing in the wind.

Although the ultimate goal for sociology proper is a more adequate conception of the structure and dynamic processes of complex sociocultural systems, it appears that such a conception requires a better understanding of the microprocesses underlying the macrolevel. Among many other areas, current concern is with such problems as the "social construction of reality," group interpretation or "definitions of the situation," and the nature and role of language as a mediator, if not creator, of such constructions and interpretations. On the assumption that the larger social structures and processes of society represent the net outcome at any one time of these group constructions and interpretations continually occurring in the underlayers of the give-and-take of daily social life, including the undertakings of the scientific endeavor itself, it becomes important to focus on the knowledge process in general, that is, the problem of epistemology.

This chapter will attempt only to make a start at mapping out a model of the epistemological process as seen from a systems theoretical view, the main emphasis being on the knowledge process as (not unreasonably) primarily an information-processing system. Many scholars already think in these terms and may find little that is novel here; others, however, use a wide variety of thought-schemes and terminologies, and for them it is hoped that a system model, because of its great generality, may provide a framework for more common and cumulative discussion across perspectives and disciplines.

One other disclaimer is in order. This chapter will touch on many philosophical bases, but the analysis can hardly be taken as adequate from that point of view: it attempts only to skim off some of the basic epistemological arguments, with the hope that the ignored refinements and subtleties do not seriously damage the position taken.

In essence, the sketch or model of the epistemological process presented here pictures a flow of information (in the modern broad sense) from the physical and social environment through the various transformations, codings, and processings of the human sensory, linguistic, and other mental or neurophysiological mechanisms, to the decision-making and consequent motor output apparatuses, and thus to actions or behaviors that constitute transactions back again on aspects of the physical and social environment—often changing the latter and hence the nature of its later inputs into the system (see Figure 6.1). The main points that I wish to argue below may be summarized as follows:

1. Adequate epistemological (as well as ontological) analysis must focus on the *total* system as a complex on-going whole, for the information selection, transformations, or codings that occur at any point or linkage in the system depend not only on prior events and processes in the system, but also on feedbacks from later points. A well-known example is the fact that the information attended to or selected for processing during activity is continually changing as a function of the on-going intentions, decisions, and actions of the individual. Perception, as well as conception, is at best a continual sampling out of the extensive potential informational cues available in the external (or internal) environment.

2. Thus the total system, when operating fully, is what I have referred to elsewhere as a transactional system, with morphogenic (structure-changing) as well as morphostatic (structure-preserving) capabilities [2]. What this means, among other things, is that knowledge is not passively and finally given merely through information input to the sensory apparatus, but rather is actively constructed and reconstructed through continual interchange between the individual and his physical and social environment. Cognitive, emotive, decision-making, and instrumental motor energy are also required
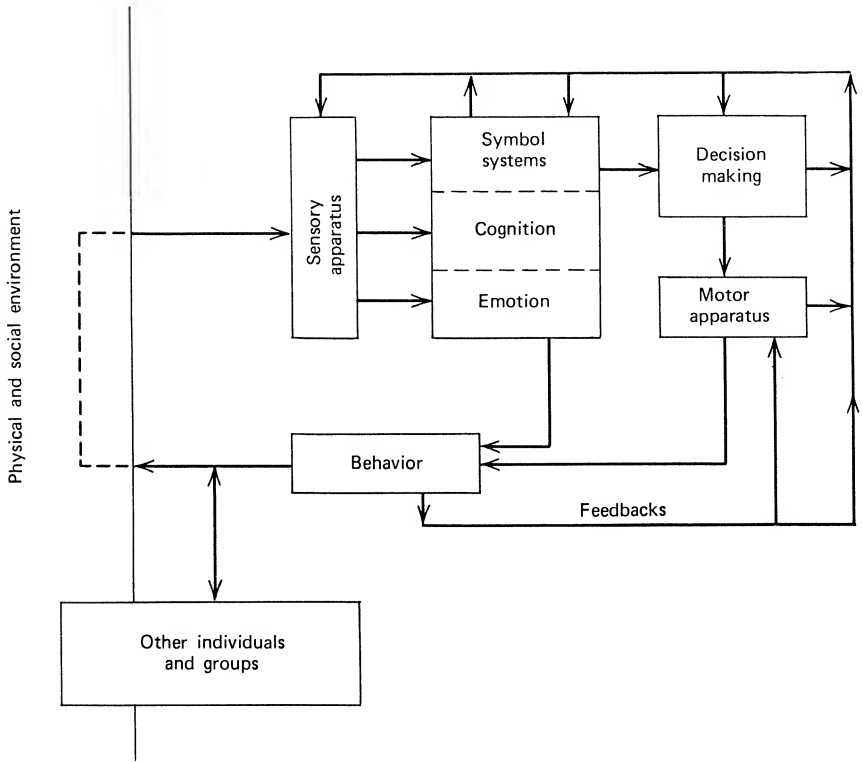
Figure 6.1. Schematic model of epistemological system.

to drive the system. Each of these subsystems contributes to the structuring and operation of the others.

3. Consequently, the classical philosophical approaches to epistemology and ontology, to which we still tend to adhere, are seriously incomplete and deficient, focusing as each does on only one or two links and transformations of the total system. I refer here to early empiricism and idealism, phenomenalism, phenomenology, logical atomism, and varieties of positivism (see the excellent review of these positions in [3]). They have moved along parts of the sequence from information inputs from an assumed (or studiously denied) external world to the sensory apparatus, to its transformation into sensations, sense data, or percepts, and thence to mental phenomena, concepts, ideas, images, or language and logical symbols. One or more of these transforms

is taken as the "ultimate" or "primitive" raw data of knowledge, which are then to be verified either through internal "intuitive" operations or by some reductive mapping against objects and relations of the external world.

6.2. A MODEL OF THE EPISTEMOLOGICAL PROCESS

In elaborating on these points, I will discuss (a) The aspects of information theory that seem especially pertinent to our model; (b) the transactional perspective; (c) the classical epistemological positions mentioned above; and (d) some implications for a modern theory of knowledge, including science.

(a) The modern theory of information and communication (c.f., among others, [4]–[6]) teaches us that information is inherently a relational concept, involving, let us say, a mapping of one subset of elements organized in a certain way (e.g., words of a vocabulary organized into written sentences) into some other subset of elements organized in a correlative way (e.g., mental concepts organized into complex ideas). If the words happen to be in a language foreign to the receiver, there is no mapping and hence no communication of information. Thus, when we speak of "information" from physical events in the external world being transformed by the sensory nerves, we are really using the term loosely, since such inputs are not information unless they map into—can be related systematically to—the internal, mental reference set; otherwise they constitute only "noise," without meaning. Also, if two or more receivers have significantly different internal reference sets, the same input signals may impart significantly different information, since the relational mappings will be different. Furthermore, since the organization of the internal reference set is a function, not only of prior external signal inputs but also of internal operations, such as cognitive and emotive operations, as well as of feedback signals deriving from active transactions of the individual with the external world, we must conclude that those philosophers are wrong who assumed that knowledge derives only from sense data from the external world.

Modern information theory also teaches us that the essence of information is the *pattern* or *organization* of the signal elements, regardless of the substantive nature of these signals—electromagnetic vibrations, air vibrations, spatial or temporal arrangements of objects or events, and the like. This means that signals may go through many transformations and appear in various substantive forms, but as long as the organization of the various elements remains invariant over the transformation, the potential information provided by the signals is preserved and requires only the final mapping into the receiving reference set to deliver its message. As a familiar example, we might take radio and television broadcasting and receiving, or the reproduction of events

via disc recordings or audio and video magnetic tape. Thus sound signals from air vibrations are transduced by the microphone, which transforms the pattern of air vibration into a nearly isomorphic pattern of mechanical vibration of a diaphragm; this in turn transforms the pattern into electrical signals which, after several more transformations, may preserve the pattern as wiggles in a record groove or magnetized sections of a tape. Eventually, if the fidelity is adequate, the pattern may be transduced by a loud speaker into air vibrations recognized as the vocal or musical message that originally entered the microphone.

Now, this illustrative model is not intended as a very close analog of the human knowledge process (which is much more of a transactional matter), but it is similar in that in both cases the input signals go through many transformations and the original pattern is preserved throughout with some degree of fidelity. Our main point here, however, is that the modern information theoretical framework provides a critical base for assessing the earlier philosophical analyses of the epistemological process. Thus of immediate import is the question of the significance, or lack of it, of attempting to localize the "really primitive or fundamental data of experience" as did the empiricists, idealists, phenomenalists, phenomenologists, and others. To localize such data in "immediate sense data," "immediate mental impressions" from sense data, "phenomenal mental configurations," language symbols, or other cognitive constructions is probably not by itself of any greater significance than to ask whether, in reproducing music via the phonograph, any one transformation of the signals—mechanical, magnetic—provides the more "fundamental or primitive" data (or information) about the original input signals. Theoretically, the question is not meaningful. If the signal pattern is well preserved in each case, then the potential information is as "fundamental" in any form. And just because the substantive form of the signals has changed, the potential information about the external world is no less direct or indirect, and is just as good for knowledge purposes regardless of the transformation, as long as the fidelity of pattern is maintained. Of course, the signal pattern is not "the same as" the external events that produced them, but for the knowledge game this is irrelevant, since the object of the game is to know of and about the external world, not to reproduce its "substance" in the mind.

In sum, since "information" (including "meaning") is inherently relational (a mapping between knower and known), it becomes meaningless to ask what the "real world" is like apart from a knower. And given the notion of information transmission as the preservation of pattern over transformations, there is no question, in principle, about how we can know the external world. That we "only" know it through its effects on our senses is no block at all to our knowing it, and in the fullest sense of that word. There is no

sense to the notion of knowing the external world "directly," just as there is now no sense in claiming that we know our selves or phenomenal experiences in some more "direct sense" than through the internal processing of "information" in the broad sense.

(b) The difficulty we actually face, then, is to be seen, not in the principles involved, but in the practical mechanisms of information transmission and transduction and their very complex organization in the higher organism. This leads us to a consideration of the fully transactional nature of the relationships between knower and external world.

We have already argued that the individual is not simply a passive receiver and recorder of incoming signals and sense data, but actively contributes additional information as well as helps to construct the particular framework or organization of the internal knowledge reference set that alone gives meaning to additional signals generated from without or from within (e.g., by thought or emotion). Additional information and knowledge structure are no doubt added, then, by the basic physical structure of the peripheral, central, and autonomic nervous systems; by on-going feedback from various phases of the total transaction of the organism as an open system adapting to or goal-seeking in its environment; and by the sociocultural processes, including language and other symboling, in which the individual and his information-processing activities are constantly embedded.

In the last few decades the work of neurophysiologists, psychologists, and the modern linguists have given new life to Kant [7] by suggesting the innate structuring of the nervous apparatus as a contributing factor in the structuring of perception and conceptual thought. The analysis of the nervous system as a complex of nerve nets processing data in the manner of logic circuits, with the higher centers acting to coordinate the various kinds of information from peripheral processes and from memory storage and to integrate it into plans, decisions, and actions, gives us a rather definite picture of a *construction*, rather than mere reproduction, process. We must await further work by the neurophysiologists and neuropsychologists to tell us the extent to which the peripheral, central, and autonomic nervous systems act as neutral transducers passing and transforming signals without adding to them significantly; the degree to which they add information that shows up in our picture of the world; or the extent to which they provide a framework (e.g., spatial and temporal), a kind of "coloration" which does not, however, alter the information pattern significantly. Studies strongly suggest that, if they do not actually add, subtract, or distort information patterns being processed, they most certainly contribute substantially to their structuring or format at the level of perception and conception.

It has been argued for some time, starting at least with transactionalists like Dewey [8] and Mead [9], that the world as we see and act on it is to a

great extent created by us, in the sense that we gradually build up a construction of it by interacting with it. A good deal of research has corroborated this view. Some studies, such as those on extreme sensory deprivation and those in which a subject wearing inverting prismatic goggles gradually comes to reconstruct an upright perceptual field, show that, not only does one have to learn or relearn a meaningful perceptual world, but also in order to construct such a world one must actively manipulate it—engage in transaction with it. Other studies, with a broader behavioral focus, have broken down the older stimulus-response model and have shown that attention, perception, conception, decision, and action all constitute a system of complexly interlinked components of the on-going act, or transaction, and none can be fully specified apart from the others. The experimental work of Piaget and his colleagues [10], cumulative over the last few decades, has done more than probably any other single group to support empirically and theoretically this transactional view. The development of perception, intelligence, and thought in the child is seen to be a matter of complex organization and construction involving the interaction of external data and internal operations, and these operations—which transform or modify the external data—are built up from the continual sensorimotor *actions* and *coordination of actions* normally performed on objects and their interrelationships.

As far as the peripheral sensory apparatus is concerned, the external world is a matter of sampling and the reception of intermittent, partial, and temporary cues, which must be constructed into larger patterns and wholes and related to other events in certain ways before they become meaningful or useful information to the organism. Thus Piaget and others have shown that even the percept or concept of a stable, constant object is not immediately and passively given but is constructed through interaction with bits and pieces of information organized from different perspectives. Even logic, a purely relational matter and one usually explained as a mere convention of language, is argued by Piaget to be, on the contrary, the basis of the structure of language. The roots of logic lie in the transactional experience of the child; logical concepts are built from operations abstracted, not from the perception of objects, but from the general coordination of actions performed on objects. This coordination of actions tends to be isomorphic with the *relations* among the objects dealt with and to build up, starting from the sensorimotor level, the internal schemas (nervous system routines) that develop further on the higher levels to structure thought and language. In coordinating its actions on the objects and events the child is engaging in acts of uniting, arranging in order, classifying, conserving some quantity or quality, including or excluding from classes, negating, and so forth—some of these depending on earlier experience with others. Thus these actions, which at first involve primarily the sensorimotor mechanisms and later gradually involve the

cognitive centers, presumably provide their share of "information" (perhaps including proprioceptive feedback signals) which, along with the data from the external world, lead to the build-up of internal representations and organizing subroutines (to use MacKay's terms [11]) that underlie the logical work characterizing intelligent thought. The fully developed "logico-mathematical" and linguistic operations of the adult are thus a mode of experience involving a large element of construction and having a transactional origin similar to, though distinguishable from, the experience of external physical objects.

Any attempts to go beyond the traditional theories of epistemology and ontology, or to speculate on the innate, conventional, or experiential bases of logic and language, must come to terms with the transactional theories and research studies mentioned above.

(c) To complete the discussion of the major sources of additional "information" and knowledge structuring besides those stemming from data of the external world, we must also mention the input from the sociocultural transactional process in which the individual, including the scientist, is intimately involved. Language, as a sociocultural product and process of the first order, has, of course, long been considered an important contributor to the structure, and perhaps the content, of knowledge and thought processes. However, too little research has been done since this theme was developed, especially by Whorf and Sapir [12, 13], to tell us much more than this.

The situation is not much better with the various versions of the sociology of knowledge. This field has provided strong arguments for the general view that social structures and processes significantly influence idea systems. Again, however, we do not know much about the extent or conditions under which this may, or must, occur.

It is at the microlevel of the basic social psychology of interpersonal transactions that social influences on knowledge have been most convincingly argued. Thus G. H. Mead has tried to show that the higher human mental processes depend on the ability to manipulate symbols and the ability to take one's self as an object—both of which develop together in the child through the give-and-take of social transactions. Through the responses of others to one's own actions, and using symbols as a vehicle for holding the self in the mind as an object, one's own actions and interactions with others become mentally manipulable. Thus both the social and symbolic processes make possible the higher knowledge processes involving the physical and social world and also make probable the introduction of additional structuring and content into these knowledge processes.

Piaget has argued that, whereas earlier intelligence in the infant is a sensorimotor intelligence based on the internalization of actions, higher intelligence requires the transformation of these internalized actions into operations, as

discussed earlier. This poses the problem of the mental representation of what has been absorbed on the level of action, which in turn involves a constructive process of "decentering" the child from his initial focus on his own body and actions to a state in which his body and actions become objects on a par with other objects and events in his external world. Once again, this implies the requirement of the development of a sense of self as an object, which in turn involves a long socialization process of interpersonal transactions.

It should be noted that, whereas the philosopher dealing with epistemology has focused almost exclusively on knowledge as knowledge of the external physical world, he must eventually face the equally important problem of knowledge of the sociocultural world. Here we have to deal with knowledge of a wide range of social and cultural settings, and of the expectations, intentions, wishes, and other internal states of persons with whom our actions must be coordinated. "Knowledge" thus becomes an extremely complex matter of the interpenetration of multiple perspectives and the continual interpersonal validation and verification of events in the day-to-day transactional process. Epistemology thus requires a group level reference, rather than a purely individual one.

Even if knowledge of simple physical objects did not contain a significant constructional component, the world of human experience involves much more that depends on the complex organization of separate and fuzzy components extending over time and space and intimately related to internal emotive, cognitive, and moral schemas. It is well recognized that, as data from the external world become more complex and ambiguous, their internal and verbal representational construction involves a greater and greater social and psychological component. It is no wonder, then, that knowledge, including scientific theories and interpretations of the world, tends to harden into "conventional wisdom" or "scientific paradigms" that require a great deal of social, as well as purely cognitive, energy and creativity to change. One result of the slow recognition of this fact has been the rejection of the traditional epistemologist's tendency to fall into a solipsist position and a turning toward the intersubjective verification view of knowledge. But this position has barely been developed. From the view of science as simply a recording and logical organizing of external data we are coming to recognize it as a fully morphogenic process, in which the sociocultural aspect may act, not only as a distorting process, but also as a purifying, noise-filtering device by gradually filtering out the idiosyncratic, the purely subjective, and the empirically false, and leaving a residue of the relatively verified and true.

Having discussed some modern viewpoints relevant to a system model of the epistemological process, we will look briefly at some traditional philosophical positions.

We have already said enough to act as a review of the fundamental inadequacies of the empiricist positions of Locke and Hume or the idealist view of Berkeley. A model of their image of the knowledge process would simply show input signals from " properties " of external physical objects being transformed in some unspecified way by the sensory apparatus to produce " sense data " or " impressions " whose relation to the external world remained mysterious, or at least a block to " direct " knowledge of the " real " world. The sense data in turn were impressed on the mind in some way as ideas, with little or no construction (for the empiricists) or with all being an internal construction (for the idealist). No consideration is given to the other links in the total transactional knowledge process, as suggested in our Figure 6.1.

The same restriction applies to the phenomenalist, for whom knowledge of the external world is reducible to sense content alone. All statements of the world are reducible to statements whose meanings can be given only in terms of sense percepts as the ultimate primitives. The model here similarly shows a flow from external objects to their transformation into sense data and an apparently immediate translation into symbolic thought. Thus, for Mach, at one stage in his thinking, a " thing " was a thought-symbol for a compound sensation [3]. Conceptual, or theoretical, construction not only did not, but also should not, play a significant role. The empiricism is rather raw in this view, with perception as well as conception quite immaculate. This position is a very seductive one for the empirically oriented scientist and makes sense if we cling to the view that knowledge is only a matter of perception and that we perceive the world only through sense data. But, as we have tried to show, neither of these assumptions is true since the full transactional process introduces much information and construction of a non-external nature into both the perceptual and the symbolic conceptual processes.

The phenomenologist simply moves the focus from sense data to mental configurational entities as unique, private, primitive givens of experience. He appears to recognize the constructive transformation of sense data into the mental realm, and appeals to private intuition as the arbiter of knowledge; however, he assumes that mental events are free from any transactional structuring of the several kinds discussed above. This assumption, of course, rests on very thin ground. There is little reason for believing that such phenomenal data are any " purer," more basic, or less abstractive than consciously thought-put theoretical constructions that are interpersonally validated. And, from our information-theoretical point of view, there are no grounds for claiming that private, intuitive mental configurations map the external world with greater fidelity. In the last analysis, its main contribution may lie in its agreement with the transactional viewpoint in warning against the atomistic reduction of complex experiences to any assumed elemental sense data.

This brings us to the logical atomism of Russell [14] and early Wittgenstein [15], and the logical positivism that was heavily influenced by it. In focusing on language and logic, and on the scientific role of these as providing a mapping of the external world, the logical atomism of Russell and Wittgenstein moved us an important step further into the transactional knowledge-system we are modeling.

Based on the view that language pictures the world, as developed by the early Wittgenstein of the *Tractatus*, and on Russell's purification of logic in his program, with Whitehead, in reducing all of mathematics, logical atomism argued that the world is made up of atomic facts and that the scientists' molecular propositions about the complexities of the world could always be reduced to and verified by way of atomic propositions that map these atomic facts in a one-to-one manner. Substantive words map to objects, and the appropriate logic constructs their interrelationships to any required degree of complexity.

Although modern information theory renders the mapping aspect understandable and plausible in principle, a number of problems arise. Russell and Wittgenstein believed that logic is conventional and quite independent of and not derivable from relations in the external world. Thus logic and fact, word and object, belong to two different universes, and consequently, whereas words picture external objects, logic is left with the mysterious task of relating these objects in empirically meaningful ways—of classifying, ordering, inferring, etc., as well as relating them into the complex wholes we seem to experience. Objects are given ontological status, but not relations, despite the fact that much of science is concerned with establishing such relations rather than the mere existence of the objects. In addition, the assumption that the world is reducible to a set of independent atomic facts can be questioned in terms of the transactional system model we have outlined.

Language, our model suggests, involves at least a three-way mapping: of individual mind to other individual minds and each to some relatively common world of experience. Since language systems develop only in a group-validating context (which is true also of the scientist's special symbol system), the resulting mapping is a complex one interrelating external objects and relations, individual cortical-symbol systems, and the common group symbol system and current understandings. Consequently, the picture theory, which is based on a simpler direct mapping between object and language, misses the full transactions in which not only do objects provide the referents for words, but also words help construct objects. Words do this because they develop in a social and manipulative transaction with the external world, involving also, of course, the constructive contributions of the sensory and higher mental apparatus. And if we accept the view that logic has an experiential base and maps the relations between objects in the external world, then

the logical atomist's (and later positivist's) program of purifying language and logic in order to map and structure the world more accurately must fail. Such purification must involve the sloughing off of needed potential information about objects and relations that show up in the open-ended richness of language. Purification is certainly necessary to clear away confusion and error, for the transactional system we picture surely contains noisy channels, but the main question is how to avoid clearing away useful information and meaning as well. To be adequate to the heavy demands put on it, language and logic must be rich enough to map the various levels of complex systems and dynamic relations of the world, and we cannot afford to beg the reductionist question.

It is such considerations as these that may underpin the modern rejection of extreme operationalism, the criticisms of a strict deductive-nomothetic model of explanation, and the insistence on maintaining open-ended concepts and theoretical constructs in the scientific vocabulary. Any positivistic appeal to "immediate data" of experience, or reliance on "ostensive" definition of atomic terms in scientific propositions, ignores the constructive process of perception and conception that underlies any experience or ostensive pointing. A similar argument can also be made against any rigidly held verification theory of meaning (cf., among others, [17] and [8]).

Logical atomism comes to grief especially over the question of what are to be taken as atomic facts—as basic ontological units which are to map into atomic propositions. Modern systems theory, with its emphasis on wholes and relatively autonomous levels, throws into serious doubt the atomist's denial of "molecular facts" (e.g., systems whose components derive at least some of their properties from being parts of the system). If we apply transactionalism to the interactions of objects and events of the external world (leaving aside the role of knower), we may conclude that the properties that define or make knowable such objects and events are always relational and are not totally inherent in the object itself. That is, a property or an attribute of an object refers to the resultant of the interaction of the object with something else, and the something else we select will help establish the property we tend to attach to the object alone. This may apply not only to so-called dispositional properties, but also to those assumed to be inherent. Thus it appears true not only of such properties as color (interaction of the structure of the object with light waves) but also, modern science suggests, of what used to be thought of as "primary qualities," for example, weight (mass), size, or shape. If this is the case, then the presumed object or event varies in its properties, depending on its relational or systemic context. Atomism would seem to be applicable only to a static, highly entropic world of no transactions and becomings. Wittgenstein's later rejection of this *Tractatus* with its atomism and picture theory of language, and his consequent development of the

"language game" notion and the public use-theory of meaning in his later *Philosophical Investigations*, can be interpreted as a shift to a transactional view of the world and the knowledge process. However, the rejection of the basic principle of a mapping process in favor of a vague use-theory of language seems an unwarranted overraction; our model argues for the compatibility and necessity of the two aspects.

6.3. CONCLUSION

By way of conclusion, it must be reiterated that this has been an exploratory sketch, which barely touches on the wide variety of difficult issues and problems to be raised in any systematic, interdisciplinary attack on a modern theory of epistemology. We began with the view that many problems in the theory and methodology of human behavior require such an overhaul of traditional epistemology. The main purpose of the chapter has been to suggest and argue for a fully systemic and transactional model of the knowledge process as a basis for the reconstitution of epistemology. This calls for the theoretical integration of many new developments of recent decades in the diverse areas of information and communication theory; neurophysiology; linguistics, psycholinguistics, and sociolinguistics; aspects of psychology, social psychology, and sociology; and the philosophy of science. A systems model seems clearly indicated here, since the genius of modern systems theory is its continually developing potential for cutting across and unifying disciplines that are battering more and more against the artificial barriers separating one from another. We have not attempted to review these newer developments here (even assuming our competence to do so), though an integrative review is definitely in order and would provide an invaluable service to the scientific community.

Our task has been the more modest one of suggesting a model that emphasizes that knowledge development is a complex morphogenic process involving interrelated, goal-seeking, and adaptive individuals in constructive interchange with each other and their more or less common environment. As a fully systemic and transactional model, it insists on taking into account (a) the total flow of "information" and its various filterings, codings, mappings, processing, and utilization—from environmental input back to environmental output and manipulation, in a more or less unified and on-going act; (b) the way in which each phase or link in the circuit provides feedback or feedforward to affect earlier or later phases in important ways; (c) the manner in which each transformation or processing stage introduces extra-environmental structuring, information, or noise: the sensory, cortical, linguistic,

motor, social group, and environmental manipulation stages all seem important aspects of the transactional construction of knowledge, apart from raw environmental inputs as such; (d) the extent to which some invariance of pattern is maintained through the various transformations, codings, or mappings, to account for the veridicality attaching to knowledge despite the transformations and structurings that occur (this is not to be identified with the older "psychoneural isomorphism" or copy theory). We have also introduced some rather radical suggestions which we feel require serious consideration: that logic has an empirical and experiential base, even if it also involves a conventional component; and, correlatively, that ontological status might be accorded relations in the external world as well as "objects" and events, especially considering that they all involve some degree of construction in any case.

It is our belief that the model roughly outlined is applicable not only to knowledge acquired in the everyday manner but to the procedures of formal science as well. Consequently, we offer it as a potential aid in the current recasting of the philosophy of science as well as in carrying forward the study of human behavior.

PROBLEMS

6.1. To say, as some philosophers have, that we can only know the external world "indirectly" implies that it could make sense to talk about knowing it "directly." Discuss what this might possibly mean.

6.2. Some have argued that ontology and epistemology are independent problem areas. Discuss the questions of how it would be possible to develop a theory of one without having a theory of the other.

6.3. Discuss the question of whether the "signals" from objects and events in the external world potentially available to a receiving individual can correctly be referred to as "information" in any strict sense of the term.

6.4. Discuss whether it may be merely an accident of human evolution (biological and/or cultural) that we tend to give full ontological status to objects but not to the relations that we believe we find empirically between objects.

6.5. Einstein and Infeld, in their study *The Evolution of Physics*, describe the decline of the mechanical view and the rise of the field view of the nature of reality. Discuss the possible bearing of this development on the previous question.

6.6. Discuss the implication for epistemology and the philosophy of science of according full ontological status to relations obtaining between objects as well as to the objects themselves.

6.7. If logic (the study of relations) is only a convention of language, discuss why it seems to work so well when applied to the external world (e.g., through applied mathematics or as the basis of the modern computer).

6.8. Discuss the implications of general systems theory for the philosophical arguments concerning emergence theory: the question of whether higher levels of organized phenomena can, in principle, be predicted or explained entirely in terms of the properties of lower-level phenomena.

REFERENCES

1. **Kuhn, T. S.,** *The Structure of Scientific Revolution,* 2nd ed. Chicago: University of Chicago Press, 1970.
2. **Buckley, W.,** *Sociology and Modern Systems Theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1968.
3. **Turner, M. B.,** *Philosophy and the Science of Behavior,* Part I. New York: Appleton-Century-Crofts, 1967.
4. **Buckley, W.** (Editor), *Modern Systems Research for the Behavioral Scientist,* Part IV. Chicago: Aldine Publishing Co., 1969.
5. **Cherry, C.,** *On Human Communication,* 2nd ed. Cambridge, Mass.: M.I.T. Press, 1969.
6. **Smith, A. G.,** *Communication and Culture.* New York: Holt, Rinehart and Winston, 1966.
7. **Kant, I.,** *Critique of Pure Reason.* New York: Macmillan, 1963.
8. **Dewey, J.,** *Human Nature and Conduct.* New York: Holt, 1922.
9. **Mead, G. H.,** *Mind, Self, and Society.* Chicago: University of Chicago Press, 1934.
10. **Piaget, J.,** and **B. Inhelder,** *The Psychology of the Child.* New York: Basic Books, 1969.
11. **MacKay, D. M.,** *Information, Mechanism and Meaning.* Cambridge, Mass.: M.I.T. Press, 1970.
12. **Carroll, J. B.** (Editor), *Language, Thought and Reality.* New York: Wiley, 1956.
13. **Sapir, E.,** *Language.* New York: Harcourt, Brace & World, 1921.
14. **Russell, B.,** *Logic and Knowledge.* London: G. Allen, 1956.
15. **Wittgenstein, L.,** *Tractatus Logico-Philosophicus.* London: Routledge, 1922.
16. **Pitcher, G.,** *The Philosophy of Wittgenstein.* Englewood Cliffs, N.J.: Prentice-Hall, 1964.
17. **Achinstein, P.,** and **S. F. Barker** (Editors), *The Legacy of Logical Positivism.* Baltimore: John Hopkins University Press, 1969.
18. **Colodny, R. G.,** *The Nature and Function of Scientific Theories.* Pittsburgh: University of Pittsburgh Press, 1970.

Part III

Formal Theories of General Systems: A Description of Some Inductive and Deductive Approaches

7. On an Approach to General Systems Theory*

ROBERT A. ORCHARD

Department of Mathematics and Computer Science, Fairleigh Dickinson University, Teaneck, New Jersey

EDITOR'S COMMENTS

Reference [1] contains all necessary preliminaries for Sections 7.1–7.8. The following references may help the reader to a deeper understanding of the material in Section 7.9:

Neumann, J. von, *Theory of Self-Reproducing Automata.* Urbana: University of Illinois Press, 1966.

Codd, E. F., *Cellular Automata.* New York: Academic Press, 1968.

Yamada, H., and **S. Amoroso,** "Tesselation Automata." *Information and Control*, Vol. 14, No. 3, pp. 299–317, March 1969.

* The approach taken in this chapter is derived essentially from: **Klir, G. J.,** *An Approach to General Systems Theory.* New York: Van Nostrand-Reinhold, 1969, and is more fully elaborated therein.

Yamada, H., and S. Amoroso, "A Completeness Problem for Pattern Generation in Tessela-
tion Automata." *Journal of Computer and System Science*, Vol. 4, No. 2, pp. 137–176,
April 1970.
Yamada, H., and S. Amoroso, "Structural and Behavioral Equivalences of Tesselation
Automata." *Information and Control*, Vol. 18, No. 1, pp. 1–31, February 1971.
Burks, A. W. (Editor), *Essays on Cellular Automata*, Urbana: University of Illinois Press,
1970.

7.1. INTRODUCTION

If we consider for a moment the diversity of disciplines which have been
attracted in varying degrees to at least a reading of the materials available
in general systems theory, an immediate question comes to mind. With so
many divergent interests to be addressed, what could possibly be presented
which would not be devoid of specific usable content? The majority of
readers appear to be concerned with specific and generally quite complex
problems, and their motivation is usually the gaining of insights or a particular
solution to a particular problem. *Content and broad applicability* should be
primary characteristics of a general systems theory if it is to be expected to
come into widespread use. The approach taken here will specifically address
itself to this probelm.

The notion of *system* has enjoyed widespread use in numerous fields, and we
will rely on an intuitive understanding of the term until it becomes necessary
to elaborate it further by precise definitions. A *system theory* can be given
(developed) for a given meaning of the word "system." This has often been
done, resulting in a spectrum of theories as indicated in Figure 7.1. A general
system is essentially an abstract model of an already existing (physically or
conceptually) system which reflects (to the degree we wish it to reflect) all the
basic or fundamental systemic traits of the original. Its precise meaning will
become clear once some general principles of systems have been explored.
It is, however, not unique and is directly related to the definition of the
system that it is to model.

In order to indicate more completely the meaning of the spectrum in
Figure 7.1, we characterize more fully the various types of system theories.

Type I. This category consists of *specific theories* of mechanical, chemical,
biological, social, and economic systems; of electrical, linguistic, mathe-
matical, and archeological systems; and so on. Type I theories deal with
particular traits of interest within the boundaries of the discipline involved
and in general ignore those traits of the system under study which are mani-
fested by virtue of its being a system. The content of these theories is usually
taken to be of interest only within the discipline involved. This information

Area of General Systems Theories

IV	III	IĪ	II	I
Highest level of generalization. Mathematical theories of abstract systems.	System theories which attempt to capture all basic or fundamental systemic traits.	System theories with some general system results.	System theories which generalize one or more, possibly non-fundamental, traits. Usually applied.	Discipline-oriented special theories.
No particular content. Too general to produce a usable methodology. Establishes guidelines.	Applicable to all (at least finite) systems. Contains general and possibly specific methodology.		Applicable to many systems in several disciplines.	Lowest level of generalization. Much specific content.
	Area of General-System-Oriented System Theories		Generalized System Theories	Specific or Special System Theories

Figure 7.1. Spectrum of system theories.

does not take into account the fact that the underlying phenomenon may be viewed as a system from perhaps several general definitional points of view.

Type II. It may be that several type I theories from different fields have overlapping content in certain respects. Certain classes of systems may be more than simply analogous in aspects; they may actually be isomorphic. It is then possible to generalize those aspects (common traits) and to produce common content. For example, the isomorphism between mechanical and electrical circuits leads to a theory of generalized circuits which is applicable to thermal and acoustic circuits as well. Certain aspects of language theory and machine theory combine to produce a theory which allows certain topics to be viewed from either the language or the machine point of view. Regular languages and finite-state automata, as well as context-free languages and pushdown automata, are examples of this. In the future more of these relations will be uncovered and will give rise to system theories. Nevertheless, the level of generalization is relatively low, and we would prefer to call these theories *generalized systems theories* rather than general systems theories. Hence a type I special theory of wiggets, if generalized, would be referred to as the theory of generalized wiggets.

Type III. A type III system theory will be discerned as satisfying three properties:
   1. It is applicable to at least *all* bounded (finite) systems.
   2. It reflects fundamental systemic traits common to *all* systems.
   3. It contains general methodical principles.
If a system theory satisfies these conditions, it will be called a type III general systems theory (or simply *general systems theory* when the context indicates type III). We note here that the type IV system theories are also general systems theories.

From this standpoint there may be many different general systems theories capturing different fundamental traits of systems. Such a development of supplementary theories would allow of periodic integration, thus yielding access to general systems theories of broader applicability and greater content. The integration of all current type III theories at any point in time would no doubt be recognized as "the" current working general systems theory. This of course presupposes the integration effort. It is exactly this type of theory, a fine balance of generality with useful content and a methodology, for which George J. Klir has been striving in his approach to general systems theory.

There are many theories in existence at a high level of generalization which do not satisfy the three requirements but nevertheless yield results of a general systemic nature. These constitute type $\overline{\text{II}}$. Most of these results

would be included in corresponding general systems theories satisfying the requirements.

Type IV. The type IV system theories are of the highest order of generalization. We are essentially working here with the *mathematical theory of systems* and the derivation of fundamental traits of systems from a formal definition of the concept of system. As such theories are more general than type III,
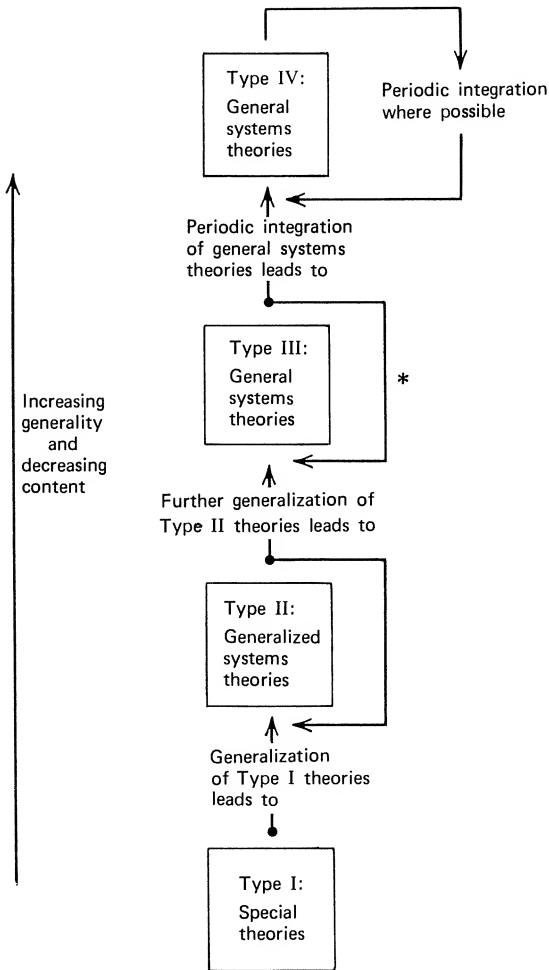
*Periodic integration may increase content of theory as well as broaden applicability.

Figure 7.2. Relationships between system theories.

they are of necessity not as rich in content. We see the role of these theories as outlining the borders of general systems theory, specifying general guiding principles, and producing results which may be used in type III theories. It follows of course that results obtained in type III theories must be consistent with those of type IV and may in turn contribute to further development of the latter. Periodic integration where possible should also occur at this level.

The relationships between the four basic types of system theories are summarized in Figure 7.2.

7.2. A PARTICULAR VIEW OF A TYPE III GENERAL SYSTEMS THEORY

In Klir's approach to general systems theory, it is recognized that a system theory can be developed for a given definition of the word "system" and that the definitions for systems in general are motivated by the problems that are to be solved for the systems. The theory may also take into account a class of systems, each of which is uniquely defined in some manner. If the class of systems, or equivalently the collection of system definitions, is chosen in such a way as to allow of general applicability to all disciplines, the systems theory will also be of general applicability. The theory itself will study the class of abstract systems (general systems) characterized by the definitions. However, the definitions should be chosen by considering only the general traits proper to all systems (called fundamental or basic traits). Of course a crucial question is, Exactly how does this general systems theory relate to a specific problem in a specific discipline? Briefly the answer is as follows.

A formulation of a problem is actually a point of view taken by an investigator concerning the problem. If he wishes to utilize a certain general systems theory, he must define a system on the underlying phenomenon (i.e., take a point of view) which coincides with a system definition encompassed by the particular general systems theory. It is important to note that the *point of view* is one which looks at the original problem in its systemic aspects. The system which is eventually defined on it takes into account only those fundamental traits of systems which are present in the collection of system definitions in the general systems theory being used; hence the results of any investigation by use of the theory will of necessity yield information of a systemic nature only. Particular knowledge involving concepts indigenous to the basic discipline involved must be derived initially by use of experiment or a type I theory.

Once the system is defined on the phenomenon, mappings or homomorphisms are established between the system defined on the investigator's

Figure 7.3. Relation of general systems theory to other disciplines.
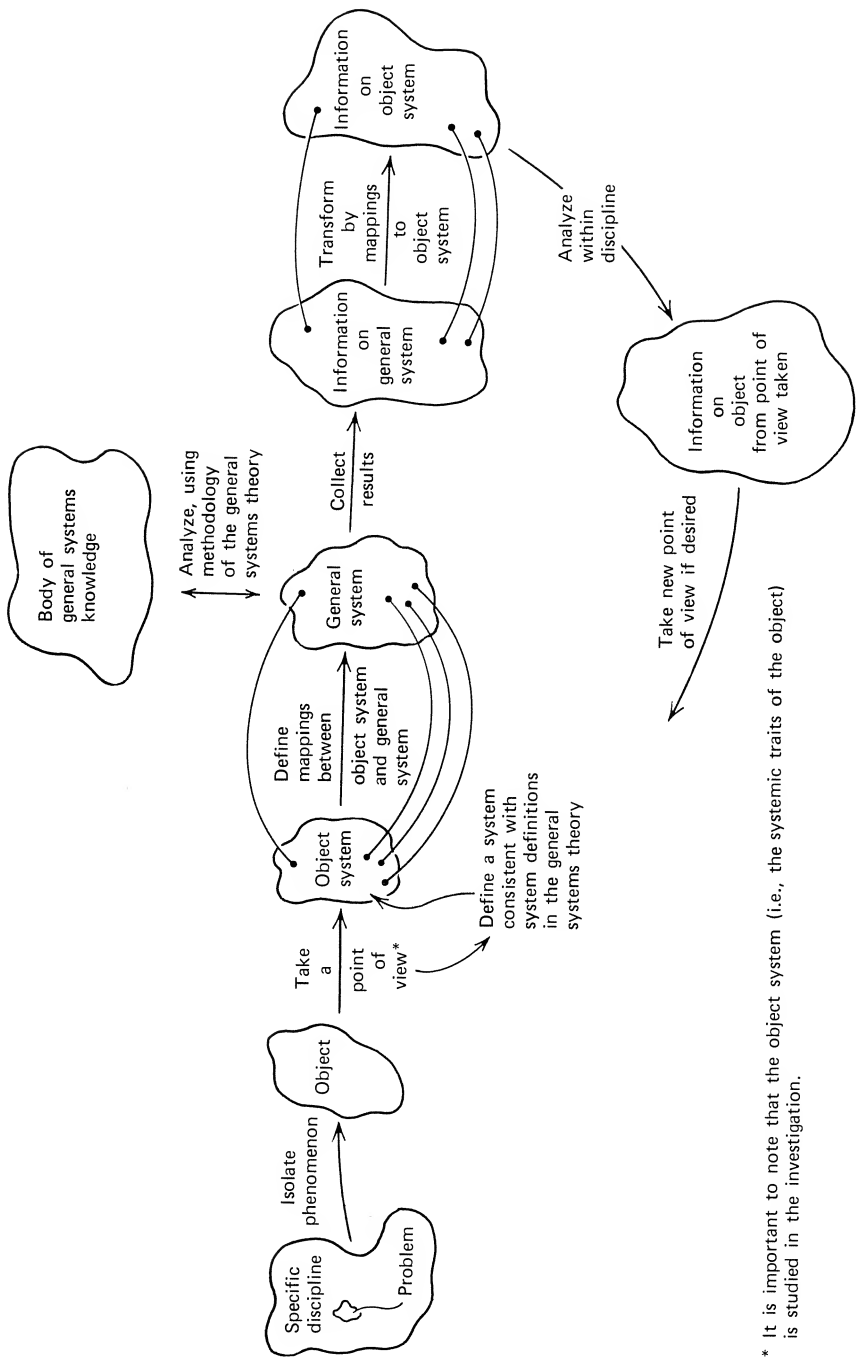
* It is important to note that the object system (i.e., the systemic traits of the object) is studied in the investigation.

problem and the general system based on the same definition. The general system (together with the mappings) is referred to as a model of the investigator's system, which again is a point of view taken concerning the original object of study. There is no reason why the same general system could not be used for a model of many systems as long as they coincide in definition and suitable mappings can be defined. We summarize this procedure in Figure 7.3.

In order to define a system on an object with respect to the general systems theory used, we must have available the class of system definitions proper to that theory. Klir undertook an extraordinarily critical study of the use of the system concept across numerous fields, including known expositions of general systems theory. He lists no less than two dozen distinct definitions of system found in the literature and presents an extensive bibliography and guide. A cross section of systems was analyzed, and a careful compilation of essential traits of systems which are not lost during the process of generalization was made. System definitions are subsequently based on these traits. A description of the fundamental concepts of systems based on Klir's study follows.

7.3. FUNDAMENTAL CONCEPTS OF SYSTEMS

Given a phenomenon (*object*) under investigation, we recognize that we cannot know the object completely in either its full simplicity or its full complexity. This is a basic epistemological premise. In an empirical science what is not under consideration may be referred to as the *environment of the object*. We would proceed to observe or measure *values* of certain *quantities* which are associated with certain attributes of the object. The values may be of a numeric or nonnumeric nature. Measuring and observing presupposes that a space and time reference frame, that is, a *space-time specification*, has been considered for each quantity of interest under consideration. The accuracy and frequency with which we record the chosen quantities is referred to as the *space-time resolution level* or simply the resolution level. For some quantities the space or the time specification or both may be irrelevant.

Assuming that the quantities have been chosen and a resolution level has been assigned to each, we measure the values of the quantities, starting at the reference time instant $t = 0$. The result is a matrix, the variation in time of all quantities, called the *activity* of the system (Figure 7.4). By observing the activity of the system we attempt to determine the relations existing between the observed quantities which are satisfied within a specified time interval. These are called *time-invariant relations*. We seek the properties that determine them and the manner in which these relations are composed of simpler relations.
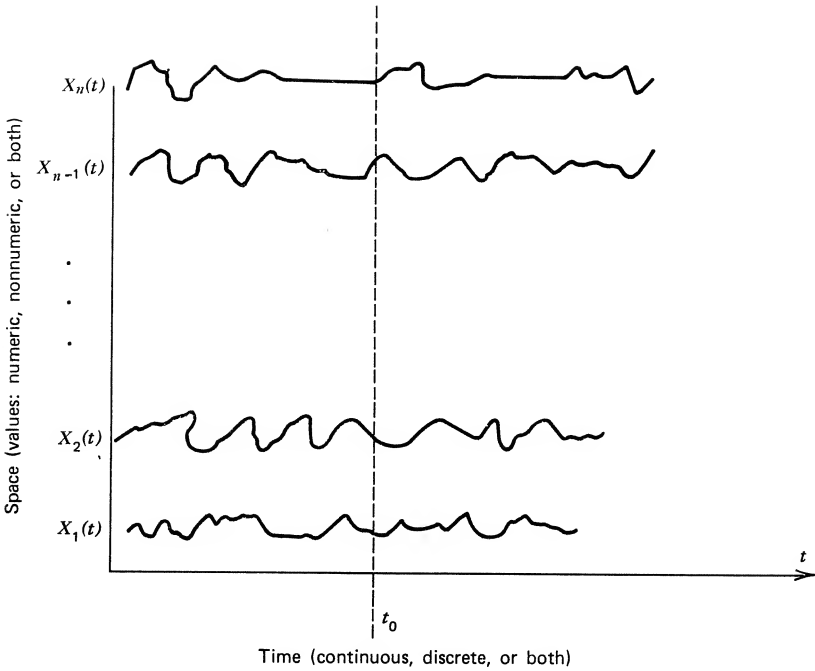
Figure 7.4. Activity.

It is entirely possible that an activity is not the direct result of observations or measurements. For example, if the quantities chosen are statistical in nature, the activity may reflect some preprocessing of data over a class of systems. Of course, the values may also be nonnumeric; no restriction is placed on the quantities other than their meaningful nature. The object itself may be a system or a class of systems defined on some other object.

The process outlined above is termed *defining a system on the object from a distinct point of view*. The fundamental traits of systems studied by experimental branches of science are the set of quantities, the resolution level, the time-invariant relations between quantities, and the properties that determine the relations. In point of fact, these are fundamental traits of every system, independently of the originating discipline which defines it. These are characteristic traits of systems in general and therefore of what we refer to as general systems. Of course, in different disciplines the problems associated with systems may be different. For example, the relations between quantities

may already be known or specified, the problem being to find a way of implementing the relation.

Once the system is defined on an object in the experimental or engineering branches of science, the system becomes the item of investigative interest. It is no longer the object but rather the object restricted to a point of view which is investigated. In abstract fields the system is the object. It is defined by enumerating the variables, the set of admissable values, and certain abstract properties, peculiar to the discipline, which determine the relations between the variables.

Let us now assume that systems are characterized by the four fundamental traits outlined. In order to facilitate the formulation of various system definitions, it will be necessary to introduce a collection of concepts and definitions related to the fundamental traits. We will attempt to do this in as straightforward a manner as possible. For a fuller discussion consult [1].

7.3.1. Behavior of a System

With Figure 7.4, we introduce the notion of the activity of the system. In order to analyze the activity and to determine whether time-invariant relations are present among the variables, we need a schema for sampling values of selected quantities at specified instants of time in accordance with the stated resolution level. Therefore the following notation is introduced.

$T = \{t \mid t$ is a considered time instance and $t \in [0, t_{max}]\}$.

$X = \{x_1, x_2, \ldots, x_n\}$ denotes the set of symbols identifying the observed quantities (also called external quantities).

$x_i(t)$ is the value of the external quantity x_i at time t.

X_i is the set of all possible values of x_i.

$L = \{X_1, X_2, \ldots, X_n, T\}$ expresses the resolution level.

p_j is a quantity called a *principal quantity* and is defined for given integers i, j, by

$$p_j(t) = x_i(t + \beta) \quad \text{for } 1 \leqq j \leqq m, 1 \leqq i \leqq n, \text{ and } \beta \text{ a real constant, } t, (t + \beta)$$
$$\in T, x_i(t + \beta) \in X_i, p_j(t) \in X_i.$$

$p_j(t)$ is the value of p_j at time t, and the index j can be associated with the pair (i, β). If $\beta = 0$ or $\beta < 0$ or $\beta > 0$, then the value $p_j(t)$ represents the instantaneous, a past, or a future value of x_i at time t, respectively.

P_j is the set of all possible values of $p_j(t)$. If j is associated with (i, β) for some β, then $P_j = X_i$.

The purpose of the principal quantities is to participate in any pattern of sampling of the activity. Their values can be construed to be sampling values.

The *behavior* of a system is a particular time-invariant relation specified for a set of quantities and a given resolution level, and based on samples of

a certain pattern. There are as many behaviors for a system as there are time-invariant relations among the quantities. In terms of the principal quantities p_j, the behavior is represented by a subset of the Cartesian product $P_1 \times P_2 \times P_3 \times \cdots \times P_m$. If the system is observed over a sufficient time interval, it may be possible, if desired, to associate each point of the behavior (relation) with its probability of occurrence.

We distinguish three basic types of behavior:

Permanent (real) *behavior.* The absolute relation satisfied over the entire time interval; the real property of the system.

Relatively permanent (known) *behavior.* The relative relation which is satisfied anywhere within a particular activity. This relation is consistent with all known data.

Temporary (local) *behavior.* A relation that is satisfied during a distinct section of a particular activity.

The relatively permanent behavior is represented by a set of all temporary behavior within a particular activity, and the permanent behavior is represented by the set of all temporary behavior. It is impossible to distinguish the permanent and relatively permanent behavior by empirical verification. To do this may require analysis of the properties that produce the behavior of the system. It must be kept in mind that the sampling pattern of quantities may be changed by the investigator, and hence there are many viewpoints that can be taken in the investigation and explanation of a system.

7.3.2. Organization of a System

The *organization* of the system is the collection of all properties producing the behavior of the system.

The *structure* of the system is that part of the organization which remains permanent, fixed, or constant and forms the basis for the permanent or relatively permanent behavior. The portion which forms the basis for the permanent behavior is called the *real structure*, and the portion forming the basis for the relatively permanent behavior is called the *hypothetic structure*.

The assumption of a decomposition of the behavior of the system into simpler behaviors implies that the system is composed of simpler systems, called *elements* (subsystems), each of which is characterized by its own behavior (time-invariant relation). Every element then is defined by a distinct set of quantities, a given resolution level, and a time-invariant relation (based on a certain sampling pattern of its activity) between certain principal quantities. Since the element is not the subject of study, its organization is irrelevant to the current investigation, although it may be relevant under a newly defined object system.

The *universe of discourse* of a system is the collection of all elements of the system. The composition of the element behaviors determines the behavior of the system.

The *coupling* of two elements is the set of all common external quantities. In the same fashion that was used to differentiate the contribution of structure to permanent and relatively permanent behavior by use of the adjectives "real" and "hypothetic," we define the contribution of the couplings to behavior. *Real couplings* are couplings valid over the entire time interval of any activity of the system. *Hypothetic couplings* are couplings valid anywhere within a particular activity of the system. The *structure of universe of discourse and couplings* (*UC-structure*) is the set of all elements and their couplings (or, equivalently, the set of all element behaviors and their compositions).

The *state* of the system is the set of instantaneous values of all quantities of the system (external as well as internal*). The *internal state* of the system is the set of all instantaneous values of all internal quantities of the system. A *transition* is a change from one state of the system to another.

The *program* of the system is that portion of the organization which is variable and which at any time t is an instantaneous state of the system, a set of some other states of the system, and a set of transitions from the instantaneous state to the states under consideration in time. We distinguish three types of programs.

Complete program. An instantaneous state, together with the set of all other states of the system, and the set of all transitions from the instantaneous state to all states of the system in time.

Subprogram. An instantaneous state, together with a nonempty subset of the set of all other states of the system, and a nonempty subset of the set of all transitions from the instantaneous state to all states under consideration in time.

Instantaneous program. An instantaneous state, together with the transitions from this state.

The complete set of states and the complete set of transitions between these states constitute a fixed, permanent portion of the complete program and therefore should be included in the structure of the organization. This is defined to be the *state-transition structure* (*ST-structure*), and we distinguish the real and hypothetic portions in the same manner as was done for the UC-structure.

Figure 7.5 expresses an overview of the definitions introduced and their

* Internal quantities of a system are quantities which are not observable from outside the system.

Organization
of the
system

Basis for Permanent
Behavior

Basis for Temporary
Behavior

The set of
transitions
from C to
elements of D
in time E

Universe
of
discourse

Permanent
behaviors
of
elements

A set of
states of
the system D

R
e
a
l

S
t
r
u
c
t
u
r
e

H
y
p
o
t
h
e
t
i
c

U
C

S
t
r
,
u
c
t
u
r
e

C
o
u
p
l
i
n
g
s

C
o
m
p
l
e
t
e

S
t
a
t
e

s
e
t

o
f

t
r
a
n
s
i
t
i
o
n
s

A

C
o
m
p
l
e
t
e

S
t
a
t
e
s

s
e
t

o
f

B

Instantaneous
state at time t C

E
n
v
i
r
o
n
m
e
n
t

P
r
o
g
r
a
m

E
n
v
i
r
o
n
m
e
n
t

Universe
of
discourse

Permanent
or
relatively
permanent
behaviors
of
elements

Set of all
transitions
from C F

Complete program A, B, C

Subprogram C, D, E

Instantaneous program C, F

Basis for relatively
permanent behavior

Figure 7.5. The Klir paradigm of systems.

217

manner of interrelating within the organization of the system. The environment is indicated on the diagram, and since we will shortly be introducing a symbol for it, several comments will now be helpful.

The behavior of the system generally is influenced by something in the environment. There may be traits in the behavior which follow not from the organization of the system but rather from the organization of the environment, which may itself be considered as an object of study. If the relevant elements of the environment are ever isolated, they should be incorporated in a new definition of the system in order that the behavior be less environment dependent.

Independent quantities are quantities which are independent of the system, cause events to occur in the system, but are produced by the environment. Quantities which are dependent on the system are called *dependent quantities*. Now in the general case the classification of external quantities into those which are independent and those which are dependent is not a priori given. All previous system theories assumed the classification as given. It follows, when possible, from other information about the system either given or observed. If the classification is known, the system is called *controlled* and we say that a *control of the system* is known. Otherwise we say that the system is *neutral*. Klir directs himself to this problem in [1].

In regard to the organization of the system, additional notation is now introduced for the definitions in Figure 7.5.

$S = \{s_1, s_2, \ldots, s_k\}$ is the complete set of states.

$R(S, S) \subseteq S \times S$ is the complete set of transitions between the states. The transition from state i to state j may be viewed probabilistically in the manner of the Markov process.

If $(s_i, s_j) \in R(S, S)$, then $\mathscr{P}(s_j \mid s_i)$ is the conditional probability of entering state j, given that we are in state i.

$\{a_1, a_2, \ldots, a_r\}$ is the set of symbols denoting the elements of the universe of discourse.

a_0 is the symbol for the environment of the system.

$A = \{a_0, a_1, \ldots, a_r\}$ contains the elements of the universe of discourse as well as those of the environment.

A_i is the set of principal quantities defined on a_i, $i = 0, \ldots, r$.

$A_i = \{p_j \mid j \leftrightarrow (k, \beta)$, where x_k is the quantity defined on $a_i\}$, $j \leftrightarrow (k, \beta) \Leftrightarrow p_J = x_k(t + \beta)$ for β a fixed constant depending on j.

b_i is the permanent behavior of the element a_i.

$B = \{b_1, b_2, \ldots, b_r\}$ is the set of all permanent behaviors of elements of the universe of discourse.

c_{ij} is the coupling of the pair of elements (a_i, a_j) and is given by $c_{ij} = A_i \cap A_j$, $i \neq j$, $c_{ij} = c_{ji}$.

$C = \{c_{ij} \mid c_{ij}$ is the coupling of (a_i, a_j), $a_i, a_j \in A$, $i \neq j\}$ and will be called the *characteristics*.

7.4. DEFINITIONS OF SYSTEMS

Five basic definitions may now be formulated, based on the four fundamental systemic traits isolated in Section 7.3 (i.e., quantities observed at a resolution level, activity of the quantities in time, time-invariant relations between them, and properties that determine these traits), and the concepts related to them which were introduced.

Usually, when solving problems, some traits of a system are given (*primary traits*) from which we are to determine other traits (*secondary traits*). We would not want to base a definition of system on traits which make it impossible to determine whether or not the secondary traits are consistent. Nor would we wish to base our definition on unknown or nonpermanent traits. Also, the traits chosen should not be redundant.

A careful search through all the traits imposing the above conditions yields the following candidates: the external quantities and the resolution level, a given activity, the permanent behavior, the real UC-structure, and finally the real ST-structure. Five definitions, each based on a separate trait, are now given. For each a verbal definition is followed by a mathematical definition, the two indicated as (a) and (b), respectively.

Definition 7.1. Set of external quantities and the resolution level

(a) A system **S** is a given set of quantities regarded at a given resolution level.

(b) A system **S** is a 3-tuple (X, t, L), where $X = \{x_1, x_2, \ldots, x_n\}$ is the set of external quantities, t is time, and $L = \{X_1, X_2, \ldots, X_n, T\}$ is the resolution level.

Definition 7.2. Activity

(a) A system **S** is a set of variations in time of the quantities under consideration.

(b) A system **S** is a 1-tuple (M), where M is a set of n-tuples $\{(x_1(t), x_2(t), \ldots, x_n(t)) \mid t \in T, x_i(t) \in X_i$ for all $i = 1, 2, \ldots, n\}$.

Definition 7.3. Permanent behavior

(a) A system **S** is a given time-invariant relation among instantaneous and/or past and/or future values of external quantities. The relation may admit of a probabilistic interpretation but is not required to do so.

(b) A system $\mathbf{S}$ is a 1-tuple $(R(P_1, P_2, \ldots, P_m))$, where R is a relation defined on $\bigtimes_{j=1}^{m} P_j$ and $P_j = X_i$ if $j \leftrightarrow (i, \beta)$ for some β, or $\mathbf{S}$ is 2-tuple $(R(P_1, P_2, \ldots, P_m), \mathcal{P}(R))$, where R is as defined previously and $\mathcal{P}(R)$ is a probability measure, defined on R, such that $\mathcal{P}(r)$ is the probability of the occurrence of r, $r \in R$.

Definition 7.4. Real UC-structure

(a) A system $\mathbf{S}$ is a given set of elements, their permanent behaviors, and a set of couplings between the elements and between the elements and the environment.

(b) A system $\mathbf{S}$ is a 2-tuple (B, C), where $B = \{b_1, b_2, \ldots, b_r\}$ and $C = \{c_{ij} \mid c_{ij} = A_i \cap A_j, i \neq j\}$.

Definition 7.5. Real ST-structure

(a) A system $\mathbf{S}$ is a set of states and a set of transitions between the states. The occurrence of transitions from one state to another may admit of a probabilistic interpretation but is not required to do so.

(b) A system $\mathbf{S}$ is a 2-tuple $(S, R(S, S))$, where S is a set of states and R a relation defined on $S \times S$, or $\mathbf{S}$ is a 3-tuple $(S, R(S, S), \mathcal{P}(R))$, where S and R are as previously defined and $\mathcal{P}(R)$ is a probability measure, defined on R, such that if $(s_i, s_j) \in R$ then $\mathcal{P}(s_j \mid s_i)$ is the conditional probability of transition from state s_i to state s_j.

A minimal definition of system would have to be one of the basic definitions. However, more information with respect to other fundamental traits may be available. Any combination of the five definitions would again be a valid system definition, but not a basic one. This process alone would yield thirty-one possible definitions of system. It may also be the case that partial information is available concerning a permanent trait. As long as at least one of the basic definitions is incorporated into the definition, the partial knowledge may also be included. For example, the set of states may be known but not the transitions. In the final analysis, whatever is added to the basic definitions must reflect some aspect of permanence.

If the control of a system is known, the classification of external quantities as input or output quantities is a permanent piece of information for this system. Each of the five basic definitions can be modified slightly to incorporate it. This would yield five basic definitions for the class of controlled systems. For example, for Definition 7.4, UC-structure, the modification would proceed as follows.

Let the sets A and B of elements and behaviors, respectively, be as defined previously for the UC-structure. The following new definitions will be needed:

I_i is the set of input quantities for the element a_i.

O_i is the set of output quantities for the element a_i.

A *directed coupling* of two elements is the set of all quantities that belong to the output quantities of the first element and to the input quantities of the second.

d_{ij} is the directed coupling of the pair of elements (a_i, a_j) and is given by $d_{ij} = O_i \cap I_j$. We note the following:

$d_{ij} \neq d_{ji}$, and d_{ij} is defined for $i = j$.*

$D = \{d_{ij} \mid d_{ij}$ is the directed coupling of $(a_i, a_j), a_i, a_j \in A\}$

and is called the *directed characteristics*.

A definition can now be given which will characterize a controlled system by its real UC-structure.

Definition 7.4′. UC-structure

(a) A controlled system **S** is a given set of controlled elements, their permanent behaviors, and a set of directed couplings between elements, including the environment.

(b) A controlled system **S** is a 2-tuple (B, D), where $B = \{b_1, b_2, b_3, \ldots, b_r\}$ and $D = \{d_{ij} \mid d_{ij} = O_i \cap I_j\}$.

In a similar manner the other four basic definitions for controlled systems could be formulated.

In retrospect, what is now available to us is a large class of possible systems definitions with the five basic definitions serving as the foundation for the class of all system definitions. If a given system definition is used in the process of solving a problem, and information is derived using the general systems methodology, then it may be incorporated in a redefinition of the system. This follows since the only thing learned from general systems theory formally is systemic. Of course the new information (traits) must be of a permanent nature. Information of a nonpermanent (time-varying) nature must be relegated to the area of secondary traits since we allow only permanent traits to characterize a system. In fact, if a system is characterized positively by its permanent traits, then it is defined by one of at most thirty-one definitions. Figure 7.6 shows this and also the relationship of various classes of system definitions.

Any definition of a system which is valid must be derived by taking a combination of the five basic definitions in the ground plane and modifying them by the traits contained on exactly one line through the vertical axis. The basis, for example, of the class of T_i-systems definitions is derived by

---

* The notion of feedback is implicitly defined by the definition of directed coupling.
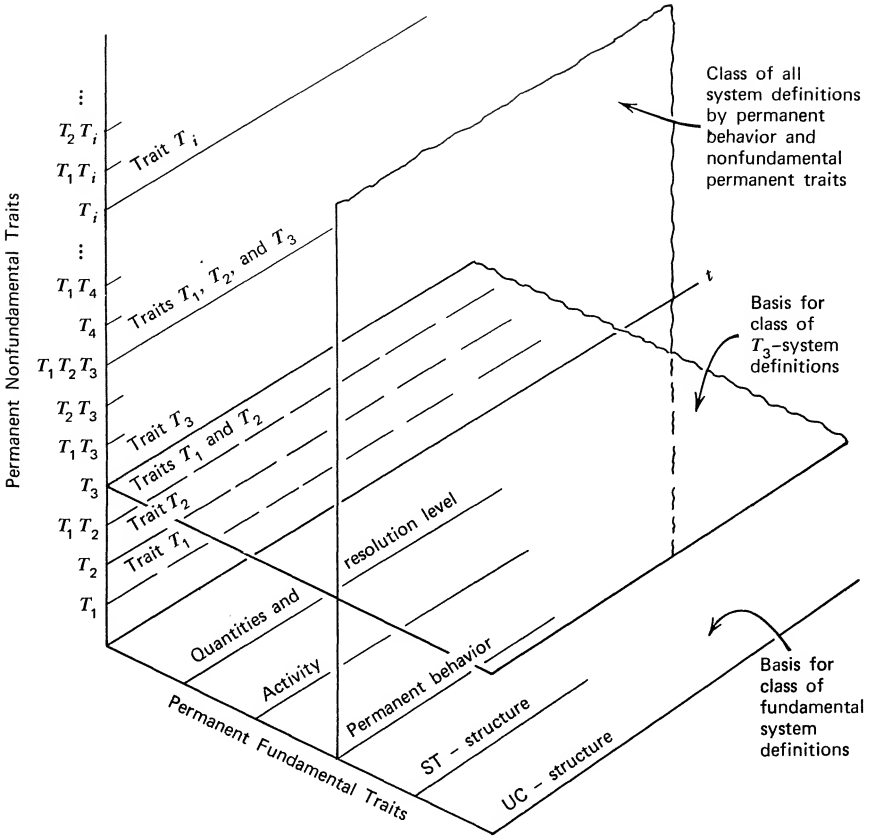
Figure 7.6. Classes of system definitions.

translating the ground plane up to the point T_i and modifying each definition (fundamental) by traits T_i. There are now thirty-one possible definitions of a T_i-system.

As one can see from observing the vertical axis in Figure 7.6, a good deal of work may be done in the general area of classifying systems. Particular classifications will depend on whether the classification is to be oriented to type I, II, or III system theories. Individual disciplines may contribute quite heavily to system classification and may point out certain classes of systems, heretofore not recognized, as of interest to type III theories. Some of the

classes of general interest which have been characterized are the physical, abstract, real, conceptual, continuous, discrete, pulse, hybrid, unique, repeated, controlled, neutral, deterministic, stochastic, memoryless, sequential, simple, complex, teleological, physically realizable, closed, relatively closed, and bounded. The concept of an open system appears in the literature and within the framework of Klir's theory represents or is represented by the *object* upon which a system is being defined.

The basic definitions require boundedness of a system as an essential ingredient. This does not prevent us from considering a system which has an infinite number of external quantities and/or an infinite structure (i.e., an unbounded system). It simply requires that the system be defined using a definition which does not refer to the trait possessing the infinite quality. In other words we recognize the trait not as a primary trait of the system but rather as a secondary one. In some cases it may not be possible to define a system within the current framework. A simple example would be a situation in which all five definitions referenced an infinite attribute of the unbounded "system."

A class of system definitions can be associated with any horizontal plane for which the corresponding set of permanent nonfundamental traits is consistent. There may be points $T_{i_1} T_{i_2} T_{i_3} \ldots T_{i_n}$ on the vertical axis which will be undefined because the traits are inconsistent. The class associated with a horizontal plane (the basis of basic T_i-definitions) consists of a finite number of definitions reflecting the nonuniversal nature of the T_i traits.

A class of system definitions can also be associated with a vertical plane. It will correspond to the collection of definitions based on one fundamental trait and an arbitrary permanent nonfundamental trait or traits. It will consist of a potentially infinite number of definitions (assuming the existence of a potentially infinite number of nonfundamental permanent traits). The basis for this class will be the subset of the plane corresponding to points on the vertical axis characterized as a single trait point. When taking combinations of these basis definitions, only those which are consistent with one another may be used. Clearly any combination of vertical planes yields another class of definitions. The potential infinity of a class of system definitions reflects the universal nature of the permanent fundamental traits. It appears that there is no shortage of possible system definitions. We note that these definitions are permanent and hence hold for all time t.

A similar circumstance exists when traits which vary in time are considered. Since they are not permanent, we cannot base a definition of system on them and where possible we would characterize them as secondary traits of the system. The relationships which exist between time-varying "systems" and infinite "systems" are currently under investigation and may prove to be

of an essential nature. In general, then, time-varying and infinite systems can be treat ed when the traits admit of a secondary classification consistent with the basic definitions.

.

7.5. CLASS OF GENERAL SYSTEMS AND THEIR USE

The class of general systems may now be identified by using the five basic definitions with the modification that defined variables, without any particular interpretation and in dimensionless form, be used everywhere in the definition in place of quantities. We see immediately that for any system defined there is a corresponding general system which could be defined with respect to the modifications just mentioned.

Since general systems theory is to be applicable to all systems, the essential preoccupation is with the class of systems definable by the five basic definitions and the development of a methodology for working with these systems. However, methodology may also be developed for certain subclasses of the class of general systems (in the sense of Figure 7.6). The controlled systems constitute a large portion of the systems actually studied by the various disciplines, and hence we would expect the general systems theory to address itself to the problem of developing a methodology for the subclass of controlled general systems. In the process of elaborating a theory for the subclass, additional concepts holding for all systems in that subclass can be introduced. These may be viewed as being "fundamental" traits for the subclass. The further we carry on this subclass differentiation, the closer that portion of the general systems theory comes to being a type II theory.

The motivation behind general systems theory includes of course the exploitation of analogies known to exist between pairs of systems. The feeling of similarity which sometimes haunts us when observing two apparently unrelated phenomena is usually an intuitive awareness of a systemic trait or traits common to both. It is clear that the fundamental traits shared by all systems (i.e., behavior, ST-structure, UC-structure, etc.) will in most instances account for a large portion of what is intuited to be "similar" in two objects although, as the type II theories indicate, there may be other components of the analogy.

Unless we can formalize the feeling of similiarity (a weak analogy) to yield a strong isomorphism between specific entities, any hope of extracting information by way of the analogy is futile. In order to carry out the formalization we proceed as follows:

1. Establish (define) a system S_1 on the object of interest, incorporating that trait (or traits) which appear to be the basis for the analogy.

2. Establish a general system S_2 on the basis of the same definition or another definition which also takes the trait or traits into account.

3. Define a unique transformation T (set of mappings or homomorphisms) between some components of S_1 and S_2, the type of transformation T depending on the trait or traits under investigation. To use S_2 as a model for S_1 with respect to the trait or traits investigated (or simply as a model of "trait" for S_1), T must be such that under the assignment the isolated trait or traits of interest become equal in both systems.

To summarize, we start off with an intuitive similiarity and isolate its basis. In step 2 an abstract system is established which also exhibits the isolated trait or traits, and we may now refer our feeling of similarity to it or rather view the analogy now holding between S_1 and S_2. Actually what has been done, no more or no less, is to say that a particular aspect or aspects of the original object appear to be systemic in nature (i.e., shared by a class of objects). We then formulate a system definition and proceed to query the general systems theory, which after all is a theory of widely held traits, for further information concerning the trait of interest.

The formulation of the isomorphism corresponding to the analogy is represented in step 3. The general system S_2 and the set of mappings T are referred to as a model of S_1. We refer to S_2 as the model understanding that a suitable T is available. The meaning of a general system is then an abstract model of an already existing system which reflects (to the degree we wish it to reflect) the basic systemic traits of the original system. Hence general systems theory allows us to study traits of systems independently of any accidental meaning but not in a manner which is independent of its own essential meaning (i.e., with content). In this respect it is in the best mathematical tradition.

For examples of the system concepts introduced so far and for the establishing of models the reader is referred to [1], where he will find an interesting collection of applications, including quantities and resolution level for weather, and for the institution of marriage in American society; activity of a traffic light system; activity of a musical composition and the melody generated by a given relation; permanent behavior of a retirement system; UC-structure for a heating system, a potted rosebush, a birth control system, and a cybernetic model of the instinct of self preservation; ST-structure of the U.S. legal system or of bacterial metabolism. Models of behavior, UC-structure, and ST-structure are given, as well as general paradigms of several classes of systems (sequential, simple, and complex probabilistic).

Models of behavior, ST-structure, and UC-structure are of primary interest since these are the models which appear for the class of general systems per se. They are by no means the only models, however, which general systems theory is capable of considering. As indicated previously, subclasses of general systems are characterized by additional traits, and these will of course be reflected in the models used in investigating these classes. Each of these models still must be characterized as being essentially a model either of behavior or of ST-structure or of UC-structure. Figures 7.7 and 7.8 give examples of some of the concepts.

→ s_1	s_2	s_3	s_4
s_1	0	1	0	0
s_2	.1	0	0	.9
s_3	.2	.3	.5	0
s_4	0	0	1	0

W/Probabilities

→ s_1	s_2	s_3	s_4
s_1	0	1	0	0
s_2	1	0	0	1
s_3	1	1	1	0
s_4	0	0	1	0

Wo/Probabilities
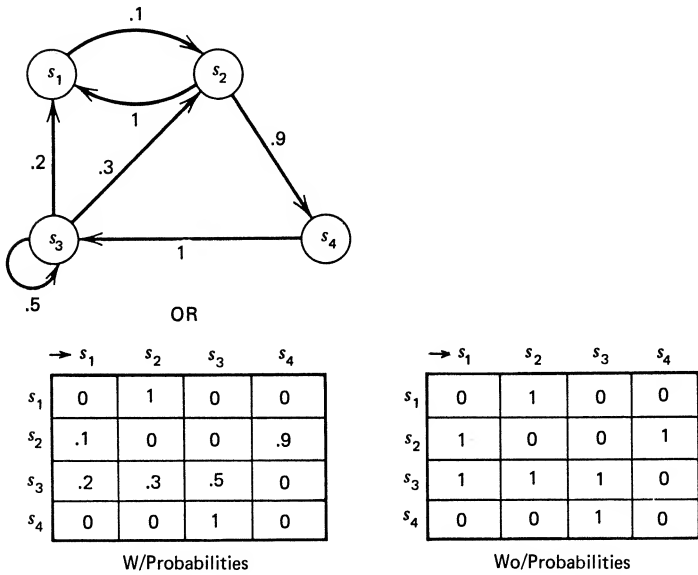
Circles denote states.
Directed lines denote transitions.
Probability of transition (if applicable) is denoted
by numbers associated with lines.

(a)

Figure 7.7 System representations. (a) Typical modes of representing ST-structure.
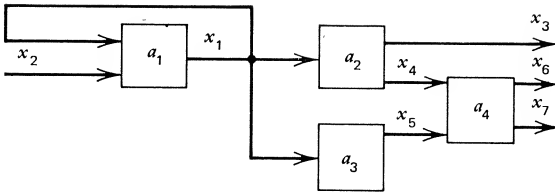
OR
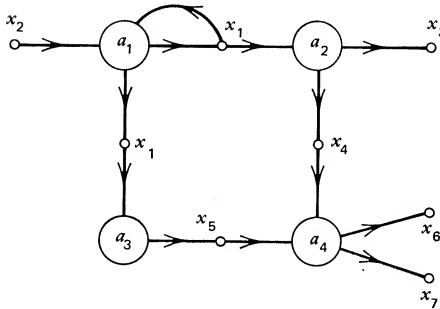
Listing of Behaviors of a_i

b_1 ...

b_2 ...

b_3 ...

b_4 ...

(b)

x	y	z
0	1	1
1	0	1
0	0	1
1	1	0
1	1	1
⋮	⋮	⋮

By table of values

$w = f(x, y, z)$

$y'' = x' + z'$

$x_n = x_{n-1} + z_{n-2}$

$$\vdots$$

$z_n = y_{n-2}$

Other symbolic
representations
(including language)

$x \longrightarrow * \longrightarrow y$
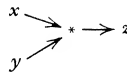
*Arbitrary well–
defined operation

(ex: concatenation
of symbols)

(c)

Figure 7.7 (*Continued*) (*b*) Typical modes of representing UC-structure. (*c*) Typical modes of representing behavior.

t	0	1	2	3	4	.	.	.	.
$x_1(t)$	0	1	0	3	2	.	.		
$x_2(t)$	1	0	0	4	0	.	.		
				.					
				.					
				.					
$x_n(t)$	1	1	0	0					

(1) $t = 0, 1, 2, 3$, encoding for $t = 0, \Delta t, 2\Delta t, 3\Delta t$.

(2) Integer values of x_i may be an encoding for disjoint discrete classes of value of variables.

OR

$$\left\{ x_1(t), x_2(t), \ldots, x_n(t) \right\} = \text{set of continuous functions representing activity.}$$

OR

t	0	1	2	3	4	5	6	7	.	.
x_1	T	H	I	S		I	S			
x_2	□	□	△	?	□	*	*	*		
x_3	α_0	α_1	α_2	.	.	.	.	α_7		
.										
.										
.										
x_n	.	.	.							

$\left.\begin{array}{r}\\\end{array}\right\}$ x_1 and x_2 take values from an alphabet. They could be integer encoded.

$\left.\begin{array}{r}\\\end{array}\right\}$ α_i are finite strings over a finite alphabet.

(d)

Figure 7.7 (*Continued*) (d) Typical modes of representing activity.

Figure 7.7 (*Continued*) (*e*) Typical sampling of activity by mask.
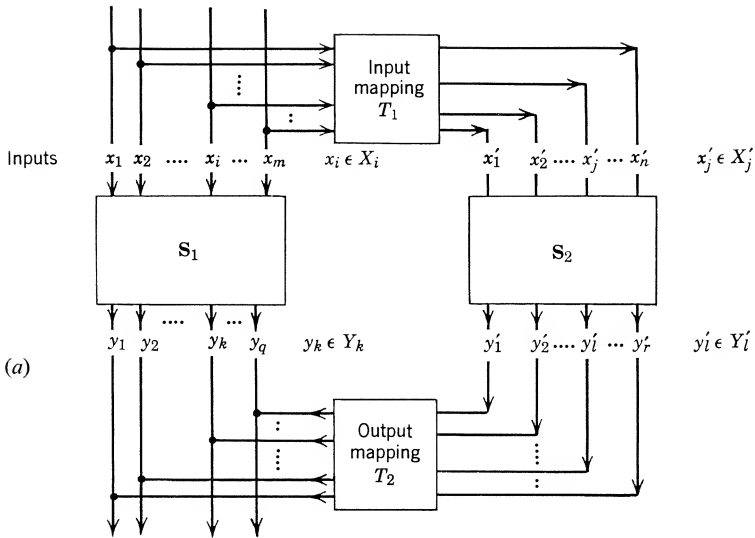
S_1 and S_2 are defined by permanent behavior,

S_2 is a model of behavior for S_1 if

> $T_1 : x'_j \rightarrow x_i \, (1 \le i \le m, \, 1 \le j \le n)$, is an assignment (mapping) between input
> $ X'_j \rightarrow X_i $ variables of S_1 and S_2 and between their values;
> $T_2 : y_k \rightarrow y'_l \, (1 \le k \le q, \, 1 \le l \le r)$ is an assignment (mapping) between output
> $ Y_k \rightarrow Y'_l $ variables of S_1 and S_2 and between their values;

and S_2 produces responses equal to those of S_1 for every equal stimulus (or a sequence of stimuli) when viewed through the mappings T_1 and T_2
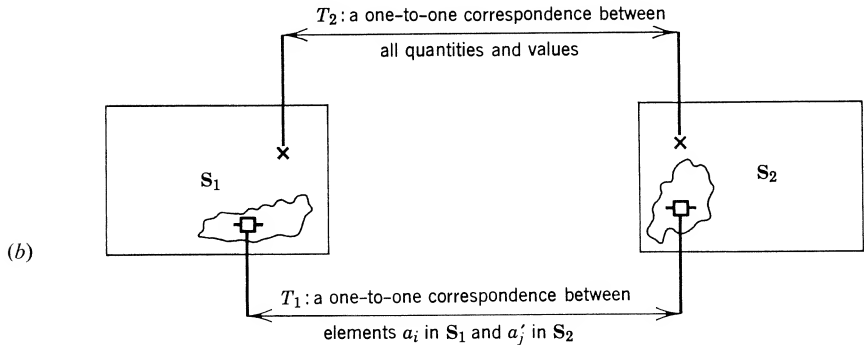
S_1 and S_2 are defined by UC-structure

S_2 is a model of UC-structure for S_1 if

> $T_1 : a_i \rightarrow a'_i$ and a'_i is a model of behavior at a_i under T_2 $(1 \le i \le r)$, and if
> $T_1 : a_i \rightarrow a'_i$ and $T_1 : a_j \rightarrow a'_j$, then $T_2 : c_{ij} \rightarrow c'_{ij}$ (preserves couplings),
> $ 1 \le i, j \le r$

Figure 7.8. System models. (*a*) Model of behavior. (*b*) Model of UC-structure.

7.6. SYSTEMS PROBLEMS AND METHODOLOGY

It is always easier to speak about problems than about the methodology which might be used for their solution, and general systems theory is no exception in this regard. The problems can be classified roughly into the three major areas of analysis, synthesis, and investigation of the unknown (a black-box problem). Within the framework in which we are working, a systems theory of type III at the most basic level of definition, *analysis* is the process of determining the behavior or ST-structure of a system, given its UC-structure; *synthesis*, the process of finding a UC-structure realizing a given ST-structure or permanent behavior or activity; *investigation of black box*, the process of acquiring knowledge of the system, given at least its definition as quantities at a given resolution level and possibly but not necessarily additional information. If the problem concerns a system known to be in a specific subclass of the class of general systems, the processes outlined may be subjected to certain additional constraints germane to the type of system. In synthesis, for example, minimization or maximization of some specified function may be required or a given set of elements must be considered for the UC-structure.

Klir has indicated methodology in several instances. He has developed additional concepts to delineate the class of discrete general systems and has introduced masking techniques for the activity matrix to yield the behaviors and corresponding ST-structures. In the area of controlled systems, the general paradigms of sequential systems and of simple and complex probabilistic systems have been given. The basic procedure of *system analysis* is outlined, but extensive detailed elaboration is currently under development. For a given UC-structure, methodology is available for the elimination of auxiliary variables for deterministic controlled systems. The determination of control for neutral systems and the detection of their behavior by way of casual relations or control of the elements lead immediately to ST-structure derivation.

For *system synthesis* it is difficult to speak of procedure in general, since so many constraints may be imposed on the UC-structure derived. Fundamental steps are available. Most procedures for synthesis have been formulated for engineering systems, and work in this area is needed for other classes of systems. For discrete systems, the classification and criterion for completeness of elements, the determination of ST-structure followed by UC-structure, the synthesis of combinational (memoryless) systems, and basic procedures for sequential systems have been evolved.

In the area of the *black box* or basic investigative problem, fundamental procedures leading to the solution of the problem have been established

from both a deterministic and a statistical point of view. A searching method for significant relations is being developed.

Even when methodical principles are relatively straightforward, there is always the problem of computational complexity when systems of moderate size are involved. Consider a system with several elements whose corresponding behaviors are described, let us say, by nonlinear differential equations (or difference equations). Such equations are difficult at best and in general are incapable of explicit solution. In such cases the digital computer is almost a necessity, and it should be used in general to aid in the analysis, synthesis, and investigation of systems. Since we will include the digital computer as a part of the methodology of general systems theory, a discussion of the role of continuous systems will be helpful at this point.

7.7. CONTINUITY AND THE DIGITAL COMPUTER

It can never be verified that the physical universe is of a continuous nature. Whenever it is observed it is at a discrete resolution level. For all intent and purpose, then, the world is a discrete object and this also we cannot verify in fact. However, we observe no instances of continuity, and discreteness is experienced in varied instances. From the point of view of general systems theory, continuous systems and discrete systems differ essentially in their resolution levels; there is, however, a complete analogy between the concepts defined for both types of systems. Some—the mask, for example—are pragmatically meaningless for continuous systems.

Since physical measurements require a discrete resolution level, physical systems are essentially discrete systems. Many times, however, they are viewed as continuous systems because the mathematics is available to handle them in this form. Mathematics has built up a superb collection of theories and a long record of achievement in applying these theories to the physical world, and many would be reluctant to give up this approach to a world view. There is no doubt that it accounts for a certain economy of thought and allows us to move from one point of information concerning a system to another with zero error or at worst as arbitrarily small an error as we choose.

Suppose that the elements of a physical system had behavior or secondary traits expressed by differential equations. Assume that by using the techniques of continuous mathematics we could express these behaviors in function form and proceed to compose the behavior of the system. Let us say that our result is a single continuous function. No error has been incurred, or at least no error which cannot be made arbitrarily small. If it is now desired to sample a value of the behavior at a particular time instance, we generally expand the

function in a series, truncate it, and use a discrete time instant. This is observation or rather measurement at a discrete resolution level.

If the physical system were assumed to be discrete, the behavior or secondary traits of the elements might be finite difference equations. In many cases the methods of solving these are analogous to, and no more difficult than, the solution of the corresponding differential equations, with two exceptions; the technique of change of independent variable is not available, and very few finite series can be summed to closed form. Both of these techniques are used in the solution of differential equations, especially those of nonlinear type.

If the digital computer is to be used for the solution of problems concerning systems, it would appear to be advantageous to define our systems from the discrete point of view. Systems work is not all computation, however, and reasoning theoretically about a system within the framework of discrete mathematics is not appealing from many obvious aspects. One cannot simply discard continuous mathematics until discrete techniques at least as powerful become available. Nevertheless, when faced with the computational aspects of the system, and in some instances even the theoretical aspects, it becomes necessary to contend with the digital computer and its resultant introduction of a discrete resolution level.

The system could be defined as discrete and envisioned as continuous where economy of thought is advantageous, in full realization that the operations we are thinking of will be carried out discretely on the computer, and the results taken at a discrete resolution level. What we are saying here is that we think integral and implement the calculus of summation.

An alternative approach is to define the system as continuous and to require continuous operations to be carried out. The final results of course will be available to us at a discrete resolution level. If we desire to convert this to the continuous form for further theoretical argument, it may be possible to utilize methodology available for transforming between discrete and continuous systems [2]. In point of fact, whether the operations were actually carried out continuously (as one assumes when using an analog computer) or discretely (i.e., on a digital computer) is of no consequence as long as the results are the same and the system is unaffected at the resolution level that we choose for observing our "continuous" results. The fact that ensuring this is a problem in numerical analysis constitutes the basis for the digital simulation of continuous processes.

A digital computer may be made to appear to the user as if it were an analog computer. The simple block diagram type of program used in programming analog computers is retained, and many of the complexities and disadvantages of the analog computer are lost. In general, speed is sacrificed. The use of a digital analog simulator is simple to learn and provides us with

a powerful tool for studies involving ST-structure, UC-structure, and the permanent behavior of the system. Such a simulator offers a definite advantage to the scientist or researcher with a minimal background in mathematics and is of particular interest to the social scientist and biologist. It is ideally suited to the handling of nonlinear problems, and this fact alone has attracted applied mathematicians to its use. An example of such a simulator is LEANS,* written in Fortran and available from the Lehigh University Computer Center; CSMP* is similar and can be obtained from IBM.

Figure 7.9 lists the available elements and their behaviors. The application
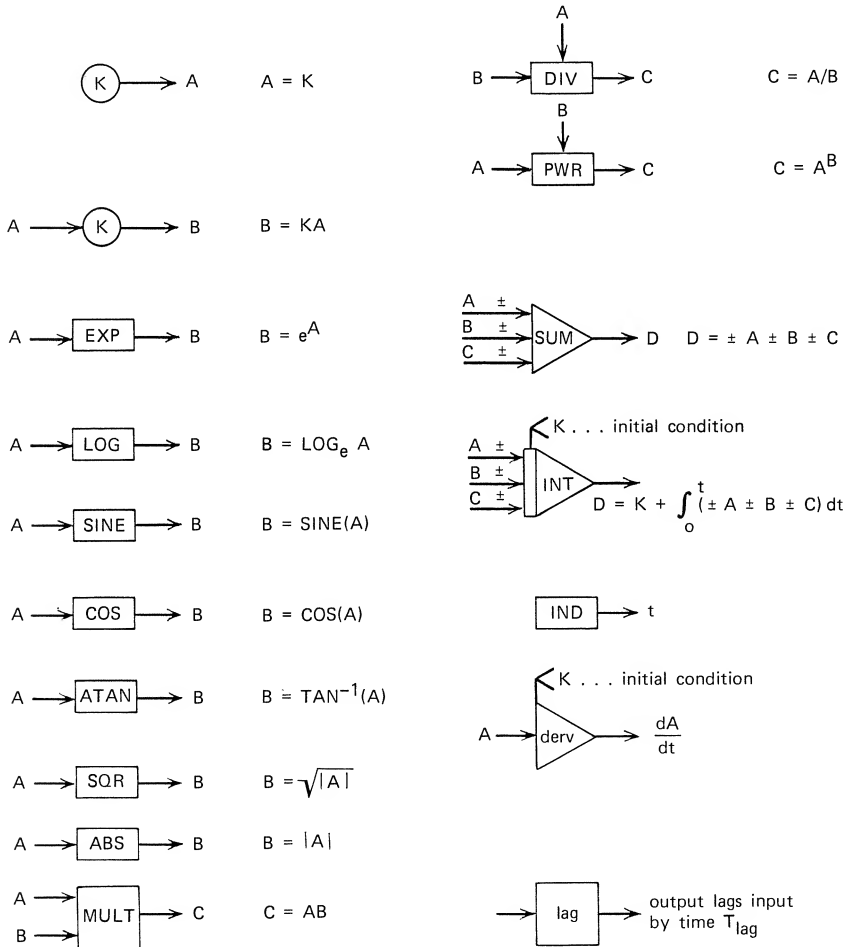
Figure 7.9. Elements and behaviors.

* LEANS (LEhigh ANalog Simulator); CSMP (Continuous System Modeling Program).
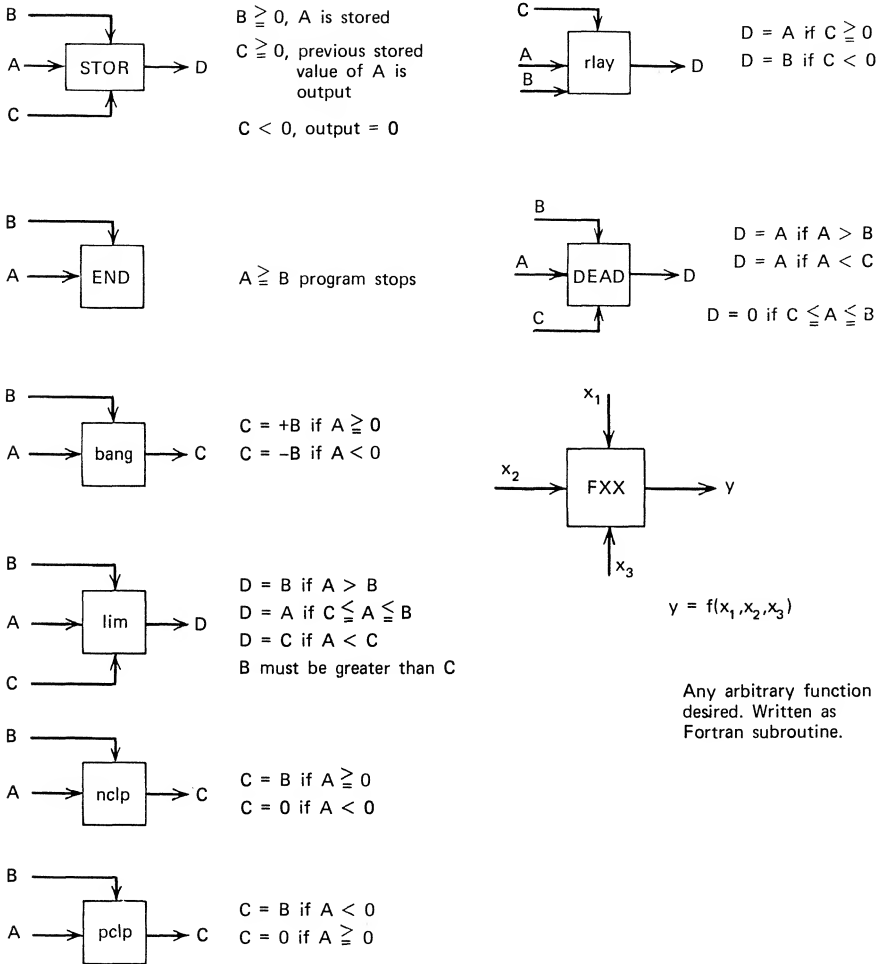
Figure 7.9. (*Continued*)

of user-programmed functions allows the freedom of designing an element with any specified behavior. For this purpose (but not otherwise) a knowledge of Fortran is required, or programming assistance must be available. Figure 7.10 shows simple examples of UC-structures simulated. Further information is available in [3]. A collection of Fortran subroutines which allow discrete simulation to be carried out, although not in the element-behavior style of the analog, is GASP II* [9].

* GASP II (General Activity Simulation Program).
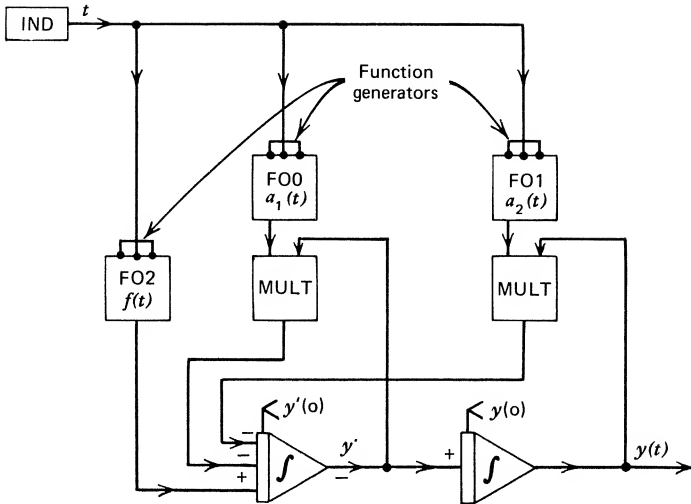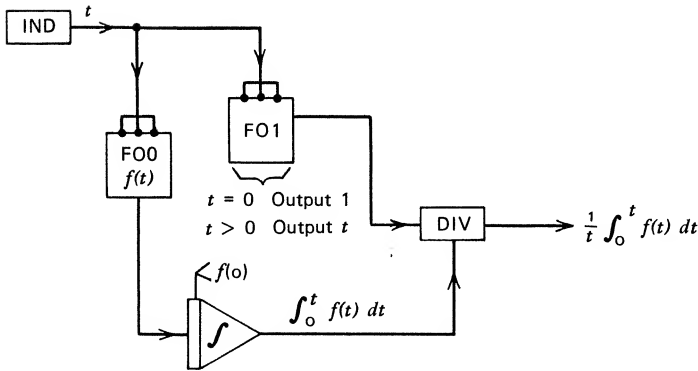
Figure 7.10. (a) UC-structure realizing the behavior $y'' + a_1(t)y' + a_2(t)y = f(t)$. (b) UC-structure realizing the behavior "The system output is the continuous estimate of the mean of the input."

7.8. A WORKING ENVIRONMENT FOR METHODOLOGY AND RESEARCH

Graph theory [4] is another mathematical theory which may be used in ST-structure (diagraph) studies. There are various implementations of computer programs to aid graph-theoretic work. A simulation of a sequential-parallel processing computer with a planar array on which a graph theory language may be implemented has been developed, based on a modification of Unger's original machine [5, 6]. Development of programs for the simulated device will be oriented toward ST-structure applications. Other languages in which similar manipulations may be carried out are also available (PL/I, Snobol, etc.). It may be that, as the ST-structures of very complicated systems are studied over a period of time, some definite patterns of structure will emerge. The theory of ST-structures could then be viewed as a specialized area of graph theory or as a theory of programming.

It appears to this author that *simulation* will play an increasingly important role in carrying out general system theoretic studies and the necessary systems modeling involved. Much of the algorithmic procedure developed for general systems theory will also eventually have to be encoded for use on a digital computer. A large body of mathematical methods is available for use with various classes of systems which could contribute heavily to the methodology of general systems. This will have to be culled, organized, classified, and integrated into the framework of the methodology—an undertaking that in itself is a major problem.

Until we can see clearly a reasonable synthesis of the procedures now available in various system-oriented disciplines (i.e., the abstraction or extraction of those suitable for general systems work), we should not accelerate the development of new " general systems " techniques. In the same manner as the definitions of systems were formulated (proceeding inductively and generalizing), a portion of the methodology may also be established. This would certainly be consistent with Klir's philosophy of approach.

In Section 7.7 we saw that certain useful tools are available as computer programs. There is an increasing tendency to make such collections of programs available to the research community at large. A recent example [7] presents a set of eight programs to facilitate the study of the class of linear control systems. It must be repeated that in our case the primary emphasis belongs on the methodology applicable to *all* systems. Secondary emphasis would be placed on the subclasses. Even if there is no current active development of techniques for a subclass of general systems, we should be aware of the developments outside the framework of general systems. There should be some method for recording potentially general methods.

What is now proposed is the development of a *computer-based, interactive system, for general systems-oriented work*. Some of the more essential features would include:

1. A *basic query system* to allow the user to characterize his system (definition), his discipline, and his general problem.

2. The maintenance of *information records of past users of general systems theory*, their disciplines, classes of systems, and types of problem (analysis, synthesis, black-box investigations). Not all records would be kept—only those of a highly representative type. This information could be used in a variety of ways to contribute to the development of the theory.

3. A *listing of the methodical tools applicable to all systems*, with simple examples of the use of each given on request.

4. A *listing of the methodical tools which appear to be germane to the user's problem* and to the specific subclass of systems characterized by the user in item 1.

5. The capability to apply methodology in an *interactive mode* under user control.

6. A one-year *record of advances in methodology* in all system-oriented fields.

Since such diverse capabilities as simulation languages, both "continuous" and discrete, and graph-theoretic and nonnumeric or string and symbol-oriented languages will be needed by various users, it appears that PL/I may be best suited for writing such a system. Actual construction will begin within the next two years. The system will incorporate such devices as graphic display units, plotting equipment, and light pen instruments. This system will be oriented toward the principle of the Klir paradigm of systems exclusively. This will allow an immediate and clear definition of the specification of the system and should significantly cut development time.

7.9. CURRENT ADVANCES

A new definition is now introduced which completes in a natural way the set of five basic definitions presented by Klir. It consists of taking the disjunction of the five definitions as a definition in its own right with the additional provision that the disjuncts hold or do not hold as a function of time. This will allow us to characterize systems by fundamental traits which may vary in time. Previously, these traits would have been relegated to the status of secondary traits of the system since a primary trait (defining the system) is required to be permanent. The alternative in some instances would have been to use the time varying trait to define a sequence of systems. In many

problems (e.g., evolutionary systems studies) the sequence of systems is precisely the system under study, but within the definition structure of general systems theory this system is not formalized. The following definition acknowledges that time variance of fundamental traits is again a fundamental trait of systems. We note that time invariance is a special case of time variance.

Let T be a well-defined set of values.

Let τ_i, $i \in \Lambda$ an indexing set, be an element of T.

Let s_j^i denote basic system definition j, $j \in \{1, 2, 3, 4, 5\}$, applied to an object with respect to τ_i.

Let $\mathscr{S}_i = \{s_j^i \mid$ basic system definition j holds for the object with respect to $\tau_i\}$.

Usually T will denote a set of discrete time instances, of intervals of discrete time instances, of continuous intervals, or a combination of the three. The sixth definition of system (definition by time-varying fundamental traits) follows.

Definition 7.6. Time-varying fundamental traits. A system **S** is said to be defined with respect to T if for any given value $\tau_i \in T$ there exists an $\mathscr{S}_i$ and if τ_i, $\tau_k \in T$ ($i, k \in \Lambda$); then there is a procedure **SP*** such that, for at least one $s_j^i \in \mathscr{S}_i$, **SP** derives from s_j^i an $s_{j'}^k \in \mathscr{S}_k$, $k > i$.

The permanent primary trait used in Definition 7.6 consists of a set of values T, a collection of system definitions, $\bigcup_{i \in \Lambda} \mathscr{S}_i$, holding with respect to τ_i, $i \in \Lambda$, and a process **SP** which allows us to move from one system definition to another in time. There are actually two levels of definition involved here: a system defined at the value τ_i, and the primary system **S** defined by Definition 7.6, both of which can be studied by methods indicated earlier.

If $\mathscr{S}_i = \mathscr{S}$, a singleton set for all $\tau_i \in T$, then **S** is defined by some one of the five definitions given previously. In this sense the new definition subsumes the five basic definitions and hence is consistent with them. The permanent primary trait used in Definition 7.6 then holds also for Definitions 7.1–7.5. The new definition allows us to characterize certain nonfinite systems, namely, those which are finitely describable. The application of the procedure **SP** describes the manner in which the fundamental traits vary in time. Systems defined by Definition 7.6 will be called K-systems.

In terms of the classes of system definitions generated, we may compare Figure 7.11 with Figure 7.6. A system is now defined by a trajectory in fundamental nonfundamental-time space. Just as the previous definitions were modifiable by nonfundamental permanent traits, we may modify them now within the context of Definition 7.6 by nonfundamental traits which vary in time. A system is now specifiable as a subset of $F \times NF \times T$, that is, as a

---

* System Procedure.
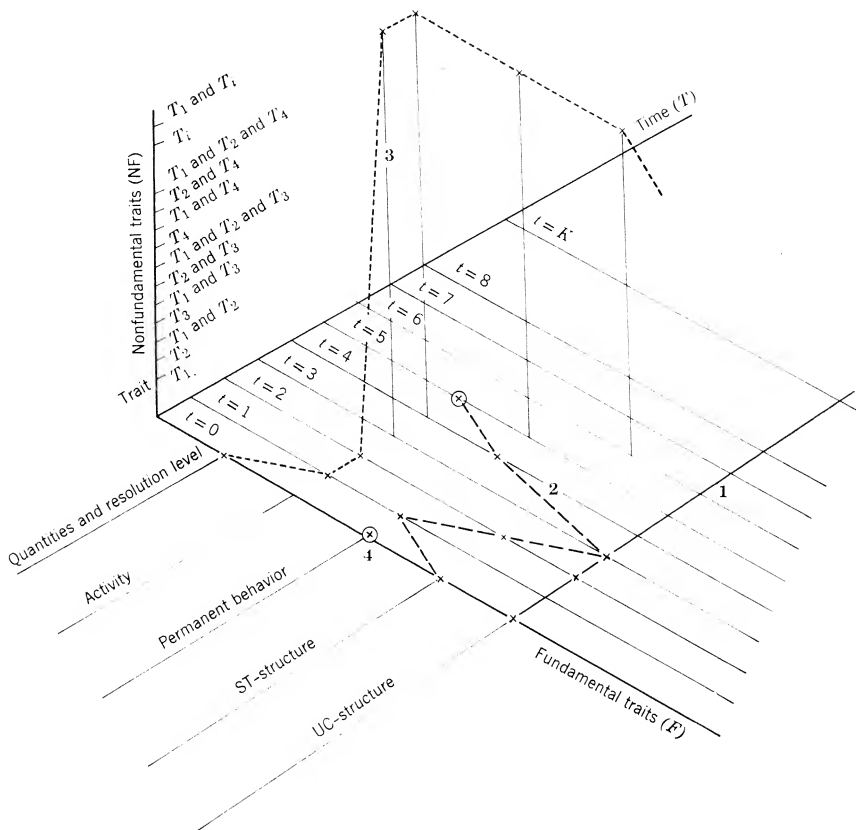
Figure 7.11. K-system trajectories.

set of triples (a, b, c), where a is the coordinate of a fundamental definition, b of a nonfundamental definition, and c of time. Since we are dealing essentially with integer 3-space, the possibility of bringing questions concerning the solvability and unsolvability of general system theoretic problems within the domain of recursive function theory is opened.

We note further that the infinite general systems will be infinite subsets of $F \times NF \times T$. In particular they will be nontrivial in the sense that the fundamental definitions used will change arbitrarily often (i.e., the systems cannot be defined by Definitions 7.1–7.5). There will be infinite subsets of $F \times NF \times T$ which will not qualify as definitions of system either because they are not finitely describable or because they are inconsistent.

The trivial infinite systems are those for which a definition from 7.1 to 7.5

may have been used and those which exhibit a *transient definitional behavior* and proceed to a *steady state*, that is, such that for all $t \geq t_0$ the system is characterizable by some of Definitions 7.1–7.5, in *time-invariant* form. Both of these types of infinite systems will be reclassified as finite general systems in the following way. We note that they do not depend on time in an essential manner.

In the first case we delete such systems from the set of infinite systems and reclassify them as single-point trajectories (see trajectory **4**, Figure 7.11). This simply removes from consideration as infinite general systems those systems whose fundamental traits are considered to be time invariant. In Figure 7.6, a straight line in the plane parallel to the time axis is equivalent to a point on the F axis in Figure 7.11.

For the second case, only that portion of the time axis during which the system is in a "transient" definition stage is considered. Once the system becomes characterized by a time-invariant fundamental definition, the system trajectory is considered to be terminated. An example of this would be trajectory **2**. Here is a system which for all $t \geq 5$ is definable by use of a time-invariant form of definition by activity. Circles at the ends of trajectories indicate that at that point in time the system definition becomes time invariant. Both of these system types characterize the finite subsets of $F \times NF \times T$.

The infinite general systems are now the systems which are defined by fundamental traits that are *strictly* time varying. We note that the trajectory of a system may not be known a priori but rather may be derived in time from the initial definition and use of the system procedure, **SP**. This will be further explicated by an example.

Trajectory **3** represents a system which is infinite in the nontrivial sense and is shown uncompleted. There are two possibilities: either it achieves a *class steady state* in the sense of trajectory **1** for all $t \geq t_0$ for some t_0, or it changes its F coordinate arbitrarily often. We do not consider the situation of a "steady state" (as in **2**), with arbitrary changes occurring in the nonfundamental traits, to characterize an infinite general system. Since the study of such a system is an excursion into the area of nonfundamental trait varitions, it can be considered to be a study in secondary traits. It can be handled concurrently with the study of fundamental traits if the infinite general system view is adopted, since the primary fundamental traits change regardless of what the nonfundamental traits do. Further work in the characterization of general systems by analysis of trajectories in $F \times NF \times T$ may lead to useful results.

Several additional comments are in order. The definitions of the system at $t = 1$ and $t = 2$ in trajectory **3** may be identical or may be different. Similiarly for **1** there may be arbitrarily long but finite repetitions of the same definition by UC-structure. There is no restriction on the interval of time (t_{k-1}, t_k);

therefore, for example, if time is measured in hours with respect to T, the period of time between successive system definition changes may be used independently of T or concurrently with it to observe the system defined at t_{k-1}. This is a consequence of the fact that Definition 7.6 embodies two levels of definition and can lead to a definition of heirarchial general system.

We proceed next to an example of a system defined by use of Definition 7.6. The theory of *tessellation automata* will be used. It has a varied background and includes as special cases a number of theories in which infinite arrays of uniformly interconnected identical finite-state machines form the primary structure [8].

Example 7.1. Consider a potentially infinite two-dimensional array of points as in Figure 7.12a. At each integer lattice point, (i, j), a finite-state automaton

(a)

(b)
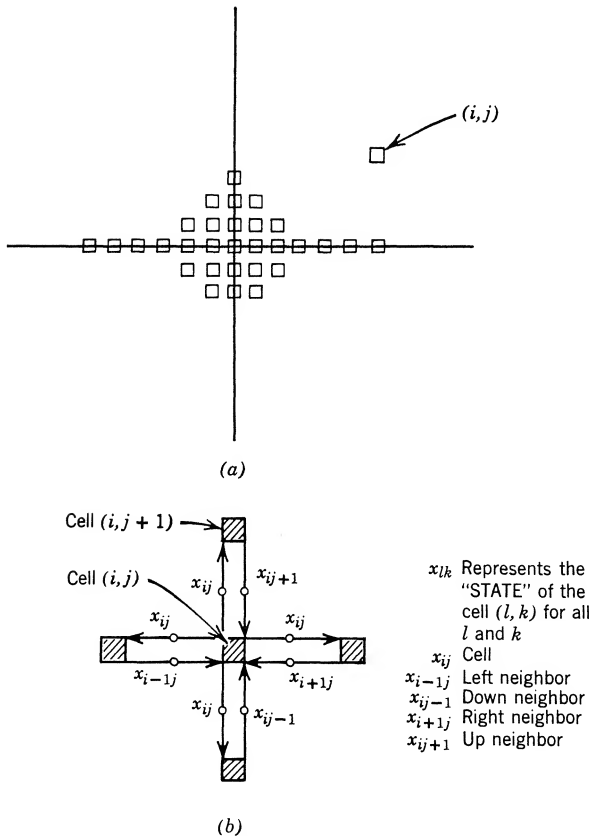
Figure 7.12. (a) Tessellation array. (b) Neighborhood structure.

(*cell*) is placed which is capable of assuming a state value 0 (inactive state) or a state value 1 (active state). A *neighborhood structure* is chosen for each cell which will hold uniformly throughout the array. For this example, the neighborhood for any cell (i, j) in the array will be the cell itself and its four nearest neighbors, that is, the cells to the left and below, to the right and above. The neighborhood of a cell is shown in Figure 7.12*b*. The double-line coupling is reduced to single-line coupling in later figures for purposes of clarity.

Each cell can change its state only at discrete time instances as a function of the states of the cells in its neighborhood. Although the function may be changed from time step to time step, it is identical for each cell at any given time instance. A next-state function can be defined for a cell on its neighborhood structure in the following way.

Let $x_{i-1,j}$, $x_{i,j-1}$, $x_{i,j}$, $x_{i,j+1}$, $x_{i+1,j}$ denote the left and down neighbors, the cell itself, and the upper and right neighbors, respectively. For ease of notation we will use x_1, x_2, x_3, x_4, and x_5 in the obvious order. There are five values to be considered hence 2^5 possible states of the neighborhood and 2^{2^5} ways of defining a function. It is required that a next-state function g satisfy $g(x_1, x_2, x_3, x_4, x_5) = 0$ if $x_1 = x_2 = x_3 = x_4 = x_5 = 0$. Under this constraint, if an inactive cell has an inactive neighborhood, it will remain inactive when g is applied to it.

Imposing the constraint above leaves us with 2^{31} possible next-state functions which may be defined. As an example, define $g(x_1, x_2, x_3, x_4, x_5) = x_3 + x_5$, where $+$ is defined by modulo 2. This particular function looks at the value of a cell and its right neighbor, adds them, and assigns the resulting value as the new state for the cell denoted by x_3. This is indicated in Figure 7.13, where $\square$ denotes the inactive cell state 0 and $\blacksquare$ indicates the active cell state 1. The arrow points to the initial center cell.

A cell is defined as *observable* if it either is in an active state or contains within its neighborhood a cell in the active state. If the full neighborhood of an observable cell is not in the collection of all observable cells, then the missing cells are taken to have state value 0 for purposes of applying the next-state function. This situation is depicted in Figure 7.13 by the use of the symbol 0 in place of the missing neighborhood cells. This may be taken to be an environmental coupling. We note in Figure 7.13 that the function g applied to (i) brings new observable cells into the system.

Let us suppose that the patterns which develop when the cell at (0, 0) is initialized at state value 1 and the array of observable cells is subjected to a sequence of next-state functions are of interest. All cells other than (0, 0) are assumed to be at state value 0. At any instant of time there will be only a finite number of observable cells, although the number is potentially infinite. If we define a system by UC-structure, the number of elements (cells) and
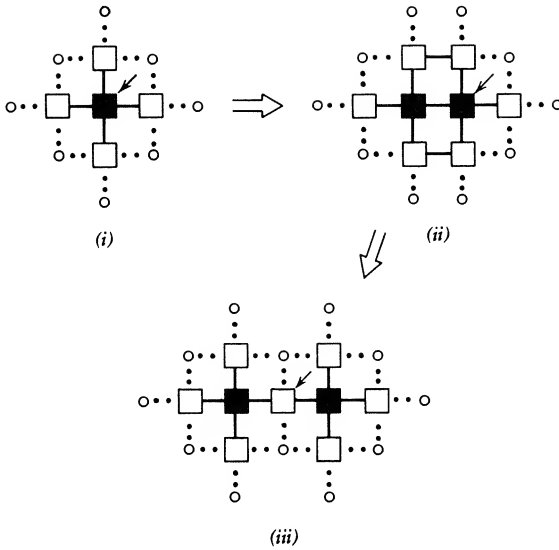
Figure 7.13. $g(x_1, x_2, x_3, x_4, x_5) = x_3 + x_5$ applied twice to (i).

their behavior (next-state function) will in general vary in time. If the sequence of next-state functions applied is a constant sequence (same next-state function), then the behavior of the elements will not vary in time, although the number of elements may. The couplings may be taken as the neighborhood structure and the environmental couplings as the coupling to missing neighborhood cells, as previously outlined. If one were to vary the neighborhood structure in time, then this aspect of the UC-structure definition (couplings) would be time varying. This is not done, however, in this example. The coupling of Figure 7.12b is used.

A system will now be defined on the tessellation array, using Definition 7.6 as follows:

Let $T = \{0, 1, 2, \dots\}$, that is, discrete time.

At any time instant $t = i$, $\mathscr{S}_i = \{s_5^i\}$, s_5^i, the definition by UC-structure holding at time $t = i$, is applied by taking the set of elements to be the set of observable cells at time $t = i$. The behavior of the elements is identical, being the ith next-state function in a given sequence of next-state functions. The couplings between the elements and between the elements and the environment is given for all time t by the neighborhood structure, which is assumed to hold uniformly throughout the array of observable cells. The cell $(0, 0)$ is assumed to be initialized with state value 1 at $t = 0$, all other elements referenced in s_5^0 having state value 0. The neighborhood structure being used

(Figure 7.12*b*) and the initialization of the observable array to that of Figure 7.13(i) at $t = 0$ are arbitrary and are used for purposes of example.

The system procedure, **SP** is as follows:

Here $s_5{}^0$ is given as the set of observable elements: $\{(- 1, 0), (0, - 1), (0, 0), (0, 1), (1, 0)\}$.

Let g_0 denote the next-state function holding at $t = 0$. To derive $s_5{}^1$, apply g_0 instantaneously and uniformly to every observable cell, yielding a new set of observable cells. This set is taken to be the set of elements for $s_5{}^1$. The behavior for these elements is g_1, the second next-state function in a given sequence of functions. This sequence of behaviors is either given or selected on a random basis (but in a well-defined manner) from the 2^{31} possible functions. These functions may be viewed as "instructions" to the array of observable "cells."

Suppose that the sequence of next-state functions is as follows:

$$g_0(x_1, x_2, x_3, x_4, x_5) = x_3 + x_4$$
$$g_1(x_1, x_2, x_3, x_4, x_5) = x_1 + x_3$$
$$g_2(x_1, x_2, x_3, x_4, x_5) = x_1 + x_2 + x_5$$
$$g_3(x_1, x_2, x_3, x_4, x_5) = x_3 + x_5 \, (= x_i + x_j, i = 3 \text{ and } j = 5)$$
$$g_4(x_1, x_2, x_3, x_4, x_5) = x_2 \cdot x_3$$
$$g_i = g_{i \bmod 5}$$

To add a probabilistic aspect to the problem, we may assume that indices i and j for function g_3 are to be chosen randomly with respect to some probability distribution from the set $\{1, 2, 3, 4, 5\}$ whenever the function is referenced. It is assumed that 3 and 5, respectively, were generated randomly for the first application of the function. Figure 7.14 indicates the various UC-structures obtained by carrying out the first five "instructions" to the "cells" using the procedure **SP**. Here $+$ is to be interpreted as modulo 2 addition, and $\cdot$ in g_4 stands for ordinary multiplication.

Comments. With no a priori knowledge of how the UC-structures (patterns) will develop in time we may consider three alternatives. Either the UC-structures will become arbitrarily large, or they will become bounded in size but oscillate in some manner, or they will achieve a steady state in the sense that the cardinality of the element set remains constant, including the possibility of eventual annihilation (null pattern) with all cells inactive. In the situation in which the number of elements remains constant, the UC-structure may still vary because of the possibility of time-varying behavior for the elements. If at the point where the elements become constant in number the behavior can be redefined so that it is time invariant, then the system has achieved "steady state" and the current definition is time invariant. It is then a finite general system.
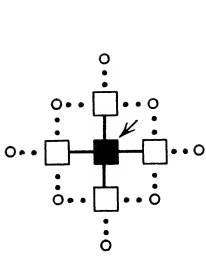
(a) UC-structure
$t = 0$
$g_0 = x_3 + x_4$
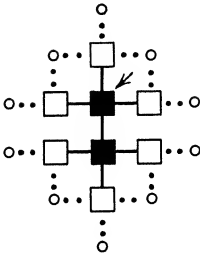
(b) UC-structure
$t = 1$
$g_1 = x_1 + x_3$

(c) UC-structure
$t = 2$
$g_2 = x_1 + x_2 + x_5$

(d) UC-structure
$t = 3$
$g_3 = x_3 + x_5$

(e) UC-structure
$t = 4$
$g_4 = x_2 \cdot x_3$
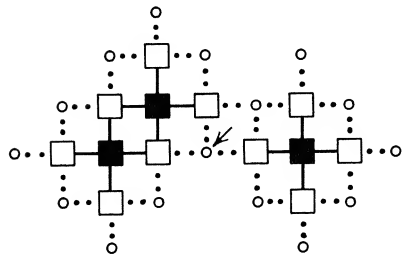
(f) UC-structure
$t = 5$
$g_5 = g_0$

Figure 7.14. Sequence of UC-structures generated by the first five steps of the system procedure **SP**, using Definition 7.6.

246

Another aspect of Definition 7.6 mentioned previously is the freedom to specify behavior for the cells independently of the behavior defined by the next-state function. At $t = k$, for example, one may specify behaviors $b(i, j)$ for every observable cell (i, j), which may hold for all t or until a new UC-structure is defined at $t = k + 1$. In addition to the 0 and 1 state values, the cells may be endowed with larger state alphabets and may take state values which consist of strings of symbols. A further possibility would be to generate the next-state function for the next discrete time instant in T on the basis of computations made in the intervening interval, using the behaviors $b(i, j)$ and the state values of the cells. This procedure would become incorporated in the procedure **SP**.

It is hoped that the example used makes evident the usefulness of Definition 7.6 in referencing general systems whose fundamental traits vary in time. The essential ingredient in the new definition is the requirement for a definitive, finitely describable procedure to indicate the manner in which one moves from one definition of the system to another.

A central problem in general systems theory may now be posed with respect to Definition 7.6: Given a system trajectory, determine the permanent behavior ST-structure, and UC-structure holding at any given time instant and compatible with the trajctory at that instant of time. This is indicated in Figure 7.15.

It appears to this author that the following areas of investigation may prove fruitful in the further development of this approach to general systems theory.
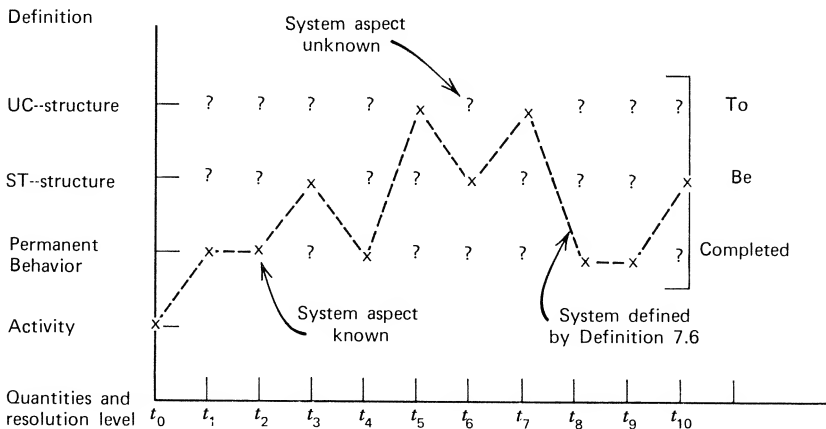
Figure 7.15. Completion problem for system trajectories.

Some have been mentioned previously but are included again for the sake of completeness.

1. The applied role of universal algebra, category theory, topology, and recursive function theory in this approach to general systems theory must be investigated.

2. Development of a computer-based working environment for general systems theory is in order. The role of artificial intelligence and heuristic programs in reducing the complexity of problems should be seriously considered, as well as the development of a high-level general systems theory programming language. The latter should be extensible, simple to use, and totally user oriented.

3. In considering the problem of general systems education, equal attention must be given to the rapid transmission of advanced mathematical techniques, both discrete and continuous, to nonmathematical specialists. This will require highly innovative educational approaches, and considerable research, utilizing both digital and analog computers. Some work is being done on this problem at Fairleigh Dickinson University.

4. Procedures of synthesis for general systems need to be developed, and the problem of interpreting these synthesis procedures in nonengineering branches of science must be investigated.

5. A complete survey and documentation of methodical tools and principles common to all fields of endeavor and suitable for general systems theoretic work is in order. This will serve as an essential basis for the further development of methodology in general systems theory.

7.10. SUMMARY

The approach to general systems theory as formulated by George J. Klir has been outlined. The foundations of the theory consist of five systemic traits shared by all systems; on the basis of these traits five fundamental definitions of systems are developed. Additional traits may be added to yield a wide variety of definitions characterizing special subclasses of general systems.

System problems are distinguished in terms of the five basic traits, and appropriate methodology is indicated. The basic traits common to all systems have been isolated by a procedure of induction and generalization. It is suggested that the development of the methodology of general systems theory should proceed in the same manner wherever possible. A plan to construct a computer environment in which work of a general systems-theoretic nature may be carried out is outlined. The needs of individuals without extensive mathematical background will be a strong factor in its design.

A sixth fundamental definition of general system, which completes in a natural way the set of five definitions formulated by Klir, is introduced. It

allows the characterization of systems by time-varying fundamental traits and recognizes the time variance of fundamental traits as basic to all systems. A system is defined on a tessellation automaton to exhibit the potential use of the new definition.

PROBLEMS

7.1. Starting with two system theories of type I, find an aspect common to both and attempt to generalize it in clearly defined terms at the type II system theory level. Can it be generalized to apply to all systems? Is it one of the six fundamental traits covered in this approach to general systems theory?

7.2. Utilizing the system bibliography, find an example of each of the system theory types indicated in Figure 7.1. Present the reasons why you have categorized each as being of a given type.

7.3. Choose two objects with which you are familiar. Use one of the six definitions to define a system on each of these objects. Explicitly show the correspondence and interpretation of the trait used in the definition with respect to the objects. Can both systems be defined by the same type of definition? Can one be used as a model for the other?

7.4. Using the notions introduced by Figure 7.6,
  (a) determine the number of distinct definition classes containing a potentially infinite number of definitions utilizing arbitrary nonfundamental traits;
  (b) choose an object with which you are familiar. Determine a significant nonfundamental permanent trait. Construct the basis (five definitions) for general systems with this trait.

7.5. (a) Give an example of an object not definable as a system by using Definitions 7.1–7.5.
  (b) Is the object definable by Definition 7.6?
  (c) Is there an object on which a system cannot be defined by using Definition 7.6?

7.6. (a) Choose two objects which you intuitively feel to be similar in some respect. Isolate the basis for your intuition. Can you use this to define both systems by the same *type* of definition?
  (b) Find two objects which are definable by the *same* definition (i.e., are isomorphic).

7.7. Define a *K*-system, using Definition 7.6, that has a trajectory consisting of more than one point and is not in a *class steady state*.

7.8. Construct an example of an infinite general system.

7.9. Using the elements and behaviors shown in Figure 7.10, construct a UC-structure realizing the following behavior: $x''(t) = x'(t) + y(t)$, $y'(t) = f(t)$, $x(0) = x'(0) = 0$, $f(0) = 300$.

Hint: Start the procedure as if you had $x''(t)$ and complete the link-up when the diagram allows $x'(t)$ and $y(t)$ to be "tapped." The completed diagram can then be simulated on LEANS.

7.10. Using the solution to Problem 7.9 and the UC-structure of Figure 7.10*b*, construct a UC-structure realizing the continuous estimate of the mean of the behavior of Problem 7.9.

7.11. Suppose that the following periodical activity is generated by a system:

$$
\begin{array}{lccccccccccccc}
t = 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\
x(t) = 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0
\end{array}
$$

a period

Find a deterministic behavior of the system. Hint: Consider principal variables (quantities) $x(t)$, $x(t-1)$, $x(t-2)$, $x(t-3)$.

7.12. (a) Is the behavior "$x = 0$ with probability 0.5 and $x = 1$ with probability 0.5" consistent with the system defined in Problem 7.11 by its activity?

(b) Is the behavior "$x(t) - x(t-1) = 0$ with probability 0.5, $x(t) - x(t-1) = 1$ with probability 0.25, and $x(t) - x(t-1) = -1$ with probability 0.25" consistent with the system defined in Problem 7.11 by its activity?

7.13. Describe each of the following functions of time, where time t may assume all positive real values including zero, by a time-invariant relation: (a) $y = t^2$; (b) $y = \sin t$; (c) $y = e^{-t}$; (d) $y = \log_e t$.

REFERENCES

1. **Klir, G. J.,** *An Approach to General Systems Theory.* New York: Van Nostrand Reinhold, 1969.

2. **Freeman, H.,** *Discrete-Time Systems.* New York: Wiley, 1965.

3. **Morris, S. M.,** *The Lehigh Analog Simulator.* Bethlehem, Pa.: Lehigh University Bookstore, 1965.

4. **Berge, C.,** *The Theory of Graphs and Its Applications.* New York: Wiley, 1966.

5. **Minafri, R. H.,** *The SPAC II Simulator.* Master's Thesis. Fairleigh Dickinson University, Teaneck, N.J., 1970.

6. **Unger, S. H.,** "A Computer Oriented Toward Spatial Problems." *Proceedings of the IRE,* Vol. 46, No. 10, pp. 1744–1750, October 1958.

7. **Melsa, J.,** *Computer Programs for Computational Assistance.* New York: McGraw-Hill, 1970.

8. **Yamada, H.,** and **S. Amoroso,** "Tessellation Automata." *Information and Control,* Vol. 14, No. 3, pp. 299–317, March 1969.

9. **Pritsker, A.,** and **P. J. Kiviat,** *Simulation with Gasp II.* Englewood Cliffs, N.J.: Prentice-Hall, 1969.

8. A Mathematical Theory of General Systems

MIHAJLO D. MESAROVIC

Systems Research Center, Case Western Reserve University, Cleveland, Ohio

EDITOR'S COMMENTS

It is unthinkable to read this chapter without a certain level of maturity in mathematics, primarily in set theory and mathematical logic. The book

Stoll, R. R., *Set Theory and Logic.* San Francisco: W. H. Freeman, 1961,

seems to be a sufficient prerequisite. Alternatively, the following pair of books may be chosen:

Halmos, P. R., *Naive Set Theory.* Princeton, N.J.: Van Nostrand, 1960,
Quine, W. V. O., *Methods of Logic.* New York: Holt, Rinehart & Winston, 1950.

Readers who like many solved problems may prefer to learn set theory from the book

Lipschutz, S., *Set Theory and Related Topics.* New York: Schaum, 1964.

Mesarovic uses standard notation of set theory and mathematical logic. This notation is used in, for example, the books recommended above. Therefore it was not deemed necessary to provide this chapter with a glossary of symbols.

8.1. OBJECTIVE AND SCOPE

The objective of this chapter is to present an approach to general systems theory which, although initiated ten years ago is still in the developmental stage. A selected bibliography on this development consists of [1]–[11]. The characteristics of this approach are as follows:

1. It is a mathematical theory; basic concepts are introduced axiomatically, and the properties and behavior of the system are investigated in a precise manner.

2. It is concerned as much with the decision-making or control representation of systems as with the input-output representation. For example, the study of hierarchical, multilevel, decision-making systems was from the very beginning of major concern.

This approach differs from other approaches to general systems in many respects. The theory is completely general, and mathematical structures are used in such a way that precision is introduced without losing any generality. This will become apparent in Section 8.3, where the basic concepts are discussed. It is important to realize that nothing is gained by avoiding the use of precise language, that is, mathematics, in making statements about the system of concern. We take exception, therefore, to considering the general systems theory as a scientific philosophy [12], but rather regard it as a scientific enterprise, without denying, however, the impact of such a scientific development on philosophy in general and epistemology in particular. Furthermore, once a commitment to mathematical method is made, logical inferences can be drawn about the system's behavior. Actually, investigation of logical consequences of systems having given properties is of central concern for the general systems theory, which cannot be limited to the descriptive classification of systems.

The decision-making or goal-seeking view of system behavior is of paramount importance. General systems theory is not a generalized circuit theory [13], a position that we believe has introduced much confusion and has contributed to rejection of systems theory and the systems approach in fields where the goal-seeking behavior is central, such as psychology and biology. Actually the theory presented in this chapter can just as well be termed general cybernetics, that is, a general theory of governing and governed systems. The term "general systems theory" was adopted at the initiation of the theory as reflecting a broader concern, but in retrospect it appears that the choice might not have been the happiest one, since the term had already been used in a different context.

8.2. FORMALIZATION APPROACH TO DEVELOPMENT OF GENERAL SYSTEMS THEORY

In order to clarify the meaning of general systems theory it is necessary to be explicit about what the systems theory is. We consider systems theory as dealing with the explanations of observed phenomena or conceptual constructs in terms of information-processing and decision-making concepts. The field is best delineated by contrast with other scientific disciplines, such as physics, chemistry, or biology. Each of these fields has a set of its own primary concepts (e.g., energy, force, quantum, etc., in physics). In systems theory, observations are explained in terms of how the information is transmitted and the goals are being pursued, or, still differently, how these processes are organized without an explicit reference to the nature of the mechanisms (physical, biological, social, etc.) involved. The subject of study in systems theory is not a physical object or a chemical phenomenon but a "system," a relationship of observed features in the context of information processing and goal seeking.

General systems theory deals with the most fundamental concepts and properties of systems. Since various branches of systems theory (control theory, automata theory, theory of adaptation, etc.) have preceded the development of the mathematical theory of general systems, an important question is how to arrive at the concepts and problems which are general and fundamental. Two lines of development have been investigated. One operates *via abstraction*: starting from well-known specific theories, one develops constructs which will embrace the common features of the initial structures; differential equations systems and automata are used widely as the starting points for these attempts [14, 15, 17, 18]. The principal deficiency of this approach is that the newly developed concepts are not general enough (e.g., the concept of a system is given in [16] by a page-long definition), so that one is bogged down by technical problems of minor conceptual importance [19]. Perhaps equally limiting is the fact that the resulting concept of a system could not have been used for the study of large systems consisting of a number of subsystems because of the complexity of the concept [17]. What is clearly needed is a *simple, elegant, general*, and *mathematically precise* concept of a (general) system which provides a point of departure for more detailed and more complex notions and problems. We have introduced such a concept in the development of the general systems theory *via formalization*. The approach, essentially, is as follows: One first defines the concept verbally as intuitively understood in the fields of application and then defines the concept axiomatically, using *minimal mathematical structure*.

Starting from this basis, one can study large-scale systems as interconnections of subsystems, or as systems which have more specialized properties, which may be investigated by adding more axioms and studying the logical consequences of the new assumptions.

8.3. BASIC CONCEPTS

The starting point is provided by the notion of a (*general*) *system* as a relation on abstract sets

$$S \subset \bigtimes \{V_i : i \in I\}, \tag{8.1}$$

where $\bigtimes$ denotes a Cartesian product, while I is the index set. When I is a finite set the more common notation will be

$$S \subset V_1 \times \cdots \times V_n.$$

The components of the relation, V_i, are referred to as (the system's) *objects*. An object stands for a feature or the characteristics in terms of which the system is described; the set V_i is the totality of alternative ways in which the respective feature is observed or experienced. The system, then, is the totality of proper combinations of the appearances of the system's objects.

The following remarks will help to clarify some of the reasons for adopting the concept of a system as a relation. A system is defined in terms of observed features or, more precisely, in terms of the relationship between these features rather than what they actually are (physical, biological, social, or other phenomena). This is in accord with the nature of the systems field and its concern with the organization and the interrelationships of components into an (overall) system, rather than with the specific mechanisms within a given phenomenological framework.

The notion of a system as given in Eq. 8.1 is perfectly general. On one hand, if the system is described by more specific mathematical constructs, say a set of equations, it is obvious that these indeed serve to define or specify a relation as given in Eq. 8.1. Different systems, of course, have different methods of specification, but they all are but relations as given in Eq. 8.1. On the other hand, in the case of the most incomplete information, when the system can be described only in terms of a set of verbal statements, they still, by their linguistic function as statements, define a relation as in Eq. 8.1. Indeed, every statement contains two basic linguistic categories; nouns and functors— nouns denoting objects, functors denoting the relationship between them. For any proper set of verbal statements there exists a (mathematical) relation (technically referred to as a model for these statements). The adjective "proper" refers here, of course, to the conditions for the axioms of a set

theory. In short, then, a system is always a relation, as given in Eq. 8.1, and various types of systems are more precisely defined by the appropriate methods —linguistic statements, mathematical constructs, computer programs, and so on.

A system is defined as a set (of a particular kind, i.e., a relation). It stands for the collection of all appearances of the object of study rather than for the object of study as such. This is necessitated by the use of mathematics as the language for the theory in which a "mechanism," a function or a relation is defined as a set, that is, by means of all proper combination of components. Such a characterization of a system ought not to create any difficulty since the set-relation, with additional specifications, contains all information about the actual "mechanism" that we can legitimately use in further development of the theory starting from a given framework, as defined by the objects in terms of which the system is defined.

The specification of a given system is often expressed in terms of equations defined on appropriate variables. To every variable corresponds a systems object that represents the range of the respective variable. In stating that a system is defined by a set of equations on a set of variables, one says that the system is a relation on the respective systems objects specified by the variables, each one with a corresponding object as a range, and such that for any combination of elements from the objects, that is, the values for the variables, the given set of equations is satisfied.

As we proceed along the path of formalization, that is, from more general to better structured and more specific, the next question is, What are the methods of systems specification, of defining a given relation as being distinct from others defined on the same objects? There are two basic approaches here: the *input-output approach* (referred to also as terminal, causal, stimuli-response, and the like) and the *goal-seeking approach* (referred to also as decision-making, problem-solving, teleological, etc.).

In the input-output approach the objects are grouped into two categories, *inputs (stimuli)*, $X = \bigtimes\{V_i : i \in I_x\}$, and *outputs (responses)*, $Y = \bigtimes\{V_i : i \in I_y\}$, where $\{I_x, I_y\}$ is a partition of I. The system is then a relation on inputs and outputs:

$$S \subset X \times Y.$$

Next, a constructive procedure, a "mechanism," defined by means of more specific mathematical structure in the sets X and Y, is provided, which to a given input associates an output (or outputs).

In the goal-seeking approach the inputs and outputs are again recognized, but instead of providing a "mechanism" relating stimuli with responses the behavior of the system is described in terms of a goal-seeking process, that is, as if the system responds to any given stimulus so that a given goal (or

objective) is pursued. Let us consider a general example. Assume that in addition to X and Y there are given the following objects: decision object M and value object V. Furthermore, let there be given two functions: the outcome (process) function

$$P: X \times M \to Y,$$

and the performance function

$$G: M \times Y \to V.$$

Let V be such that every subset has a minimum. A system $S \subset X \times Y$ can then be defined by the following statement:

For any $x \in X$ and $y \in Y$, $(x, y) \in S$ if and only if there exists $m_x \in M$ such that for all $m \in M$

$$G(m_x, P(x, m_x)) \leq G(m, P(x, m))$$

and

$$y = P(x, m_x).$$

In other words the system $S \subset X \times Y$ is defined by the following statement:

For any input $x \in X$ the system's response (output), $y \in Y$, is such that an appropriate performance function is minimized while some constraints as specified by an outcome function are observed. The goal of the system appears, then, to be minimization of G.

For proper understanding of the goal-seeking approach the following remarks will be helpful. We have defined the goal-seeking activity of the system in the example as minimization. This, of course, is just a special case. Many other approaches (satisfaction, general problem solving, etc.) can be used to define the goal-seeking activity.

The goal-seeking procedure, such as minimization in the given example, is introduced solely in order to provide a specification of the given system, that is, a relation on inputs and outputs. In general, the only objects actually observed are X and Y; M and V are additional objects assumed for convenience of an appropriate and efficient specification of the system. There is little point (within the systems theory as such) to arguing whether the system is actually pursuing such a goal or not. All it matters is that the systems functioning can be described and is most appropriately described in such a manner. In general there may be more than one way to describe the system as goal-seeking. Also, there might be a case in which the systems functioning can be described only as goal seeking, while an input-output transformation specification is not given. That does not mean that the system fails to satisfy some kind of causality requirements and has some intrinsically different

character from, say, a system described by a set of differential equations. It means only that within the family of constructive procedures which we are currently using to describe input-output transformation there is none which corresponds to the observed input-output relationship. After all, there is no reason to believe that all input-output transformations ought to be describable by the transformation procedures currently used. The availability of a goal-seeking description, then, can be simply considered as a convenience and indeed a necessity for an efficient specification of systems.

The goal-seeking activity of a system, in general, is of a much more complicated nature, involving a multiplicity of goals. This complexity, however, does not change the basic character of the goal-seeking systems description, in contradistinction to the input-output description discussed above. Further discussion of this subject can be found in [9] and [21].

In principle, any system can be described as an input-output (terminal) system or a goal-seeking system. This observation can provide a point of reconciliation for many conflicting approaches, such as Skinner's behaviorism and Chomsky's structuralism.

8.4. SOME RESULTS IN THE ABSTRACT DYNAMIC SYSTEMS THEORY

In order to give a constructive specification for an input-output system it is necessary to introduce more mathematical structure into the system objects. There are two basic ways to do this: (1) define elements of the objects as functions, $x: T_x \to A_x$ and $y: T_y \to B_y$, and define the constructive specification of a system by means of an induction process on the domains of these functions; and (2) define system objects as algebras and introduce constructive specification in terms of algebraic operations in S or, rather, X and Y. The first approach leads to the time and dynamic systems; the second, to the algebraic systems. We shall follow the first route here as being intuitively somewhat more appealing.

Proceeding along the formalization path, we have the following two notions:

(i) *A (general) time system* is a general system such that $X = A^T$, $Y = B^T$, that is,

$$S \subset A^T \times B^T,$$

where A and B are (abstract) sets referred to as *alphabets*, while T is a linearly ordered set referred to as a *time set*. It will also be necessary to assume that T is a complete lattice.

(ii) *A* (*abstract*) *dynamical system* is a time system for which there are given a (abstract) set Z and a pair of functions

$$\rho: Z \times X \times T \to Y \times T,$$

$$\phi: Z \times X \times T \times T \to Z,$$

such that

$$(\exists z)[\rho(z, x, t) = (y, t)] \leftrightarrow (x, y) \in S,$$

$$\rho(\phi(z, x, t, t'), x, t') = (y, t').$$

Here Z is referred to as the *state space*, ρ as the *state representation* (or systems response function), and ϕ as the *state transition*.

An alternative way to define an abstract dynamic system, which is slightly less elegant but more intuitively appealing, is to define ρ and ϕ on the restrictions of the time set. Let T_t, T^t, and $T_t^{t'}$ be the segments of T defined as follows:

$$T_t = \{t^*: t^* \geq t\}; \qquad T^t = \{t^*: t > t^*\};$$

$$T_t^{t'} = T_{t'} \cap T^{t'} = \{t^*: t < t^* \leq t'\}.$$

We can now define restrictions of the inputs and outputs, for example, $x_t = x|T_t$, $x^t = x|T^t$. For every t there is defined, then, a state-representation function

$$\rho_t: Z \times X_t \to Y_t$$

such that

$$(\exists z)[\rho(z, x_t) = y_t] \leftrightarrow (x, y) \in S \ \& \ x_t = x|T_t \ \& \ y_t = y|T_t$$

For the state transition there is given, then, a two-parameter family of functions

$$\phi_{tt'}: Z \times X_{tt'} \to Z,$$

such that

$$\rho_{t'}(\phi_{tt'}(z, x_{tt'}), x_{t'}) = \rho_t(z, x_t)|T_{t'}$$

The interpretation of ρ and ϕ for the time evolution of the system is now apparent: Given the state at a specific time, z, and the remaining input, x_t, the state-representation function gives the rest of the output. The state-transition function, on the other hand, gives the new state at any time t' if the state z at an earlier time, t, is given.

We shall now present briefly some of the results in the general theory of time and dynamical systems. The following format will be used. The question of interest will be stated verbally, and the answer given in the form of a theorem. References for the proofs and more detailed discussion will be indicated.

1. *Existence of the state space.* The question of interest is, What properties ought a system (i.e., the observations) to have so that the state space can be introduced, that is, the state-space approach can be applied? The answer is simply provided by the following:

Theorem 8.1 [4]. For every time system there exist a state space and associated state-representation and state-transition functions, that is, every time system can be represented as a dynamical system.

The theorem answers the question, What are the minimal axioms needed for the introduction of a state space or, rather, for viewing the system as a dynamical system? Minimal axioms are given in the notion of a time system; that is, the elements of the objects ought to be functions defined on a linearly ordered set which is a complete lattice.

2. *Natural (minimal) state space.* For a given system there might be any number of alternative state spaces, and the question is whether some of them have special character implied by the system's observations, that is, the input-output pairs. Let Z_t be the set of states the system can be in at time t; Z_t will be referred to as the set of *natural states* (at time t) iff there exists a function $\lambda_t : X^t \times Y^t \to Z_t$ such that

$$\lambda_t(x^t, y^t) = z_t \leftrightarrow (\exists z)[z \in Z_0 \ \& \ \phi(z, x, t_0, t) = z_t]$$

and furthermore

$$(\forall x)[\rho(z_t, x, t) = \rho(z_t', x, t)] \to (z_t = z_t').$$

A state space is natural if for any t it contains only the natural states if they exist; that is, if the system can be in $z \in Z$ at t and there exist natural states at t, z is such a state. We shall say that the system is *pastdetermined* at t iff

$$(x^t, y^t) = (x'^t, y'^t) \to y(t) = y'(t).$$

We have then the following:

Theorem 8.2 [11]. A system has a natural state space iff there exists $t \in T$ such that the system is pastdetermined for all $t' \geq t$.

Basic characteristics of the natural states are that they are determined solely by the past history of the input-output pairs and do not depend on the initial states. While every time system can have a dynamic system characterization, not every time system has natural state space. The existence and the type of the natural state space for a system represent an important characteristic of the system.

It has also been proved that the natural state space is minimal in the appropriate sense; for example, it gives the state space in terms of the minimal number of component sets (minimal dimension) in appropriate concrete realization (see [9]).

3. *Causality in dynamical systems.* Causality in a system is determined by the recognition of the input and output objects, the dependence among the output objects, and the nonanticipation in the time evolution. It is the third aspect to which we are referring here. We shall say that a given dynamical system is *causal* iff for any $z \in Z$ and $t \in T$

$$(y, t) = \rho(z, x, t) \ \& \ (y', t) = \rho(z, x', t) \ \& \ x^t = x'^t \to y(t) = y'(t).$$

In other words, starting from a given state $z \in Z$, the outputs can differ only after the inputs become different. The corresponding state space will be referred to as causal. We have then the following:

Theorem 8.3 [4, 10]. For any time system there exists a causal state space, that is, the system can be represented as a causal dynamical system.

The conclusion here is similar to that in item 1: All that is needed to be able to define a system as a causal dynamical system is that the elements of the input and output objects be functions defined on the sets which are linearly ordered and complete lattices.

4. *Realization for linear systems.* Let the system objects have the following structure. There is given a binary operation in X, $\circ: X \times X \to X$, and a set of mappings $A = \{\alpha: X \to X\}$ and similarly $*: Y \times Y \to Y$ and $B = \{\beta: Y \to Y\}$ and furthermore $K: A \to B$. A functional system $S: X \to Y$ is then *linear* iff for all $x, x' \in X$ and $\alpha \in A$

  (i) $S(x \circ x') = S(x) * S(x')$,
  (ii) $S(\alpha(x)) = K(\alpha)(S(x))$.

The linearity of a system as a relation is defined analogously. If the system objects are monoids, it has been proved that the state can be decomposed into the null-state and null-input parts, that is,

$$\phi(z, x, t, t') = \phi(z_0, x, t, t') \cdot \phi(z, x_0, t, t'), \tag{8.2}$$

where z_0 and x_0 are the null state and the null input, respectively, while $\cdot$ is the binary operator in Z. The state transition for a linear system is not, in general, a bilinear function, that is, linear in z for every x and linear in x for every z, although the null-input and the null-state parts are linear functions. However, not any two linear functions defined on appropriate objects can be composed into the state transition for a linear system. A pair of functions, $M_{z_0}: X \times T \times T \to Z$ and $N_{x_0}: Z \times T \times T \to Z$, will be referred to as a realization for a linear system iff they yield (in the manner given in Eq. 8.2) a state transition for such a system. We have then the following:

Theorem 8.4 [9]. Let Z be a group and let ϕ_{x_0} represent a null-input part of a state transition, $\phi_{x_0}: Z \times \{x_0\} \times T \times T \to Z$. There exists a linear realization for the system iff there exists a function linear in X, $N: \{z_0\} \times X \times T \times T \to Z$, such that

$$N(z_0, x, t_0, t) = \phi_{x_0}(z_x^{-1}, x, t_0, t),$$

where

$$z_x = N(z_0, x, t, t_0) \quad \text{and} \quad z_x \cdot z_x^{-1} = z_0.$$

A direct application of this theorem for differential equations systems gives the condition for a matrix to be a weighing pattern (impulse response) function of a linear system.

5. *Controllability.* Controllability is defined in reference to the objective of control. Let $S: M \times U \to Y$ be the system and $G: M \times Y \to V$ the performance function. Also, M is the control object, while U can be the set of initial states and/or disturbances. Then S is *controllable* in $V' \subset V$ over $U' \subset U$ iff

$$(\forall v)(\forall u)(\exists m)\{v \in V' \ \& \ u \in U' \to G(m, S(m, u)) = v\}$$

If $U' = U$ and $V' = V$ or if these sets are understood from the context, we simply say that S is controllable (relative to G). The function G depends on the objective of control. In the classical guidance problem both U and V are the state space, and controllability is defined as the ability to bring the system from one state to another. In functional reproducibility V is the output object and the controllability means that any output from V can be reproduced. In other applications V is the set of real numbers giving the cost of operation. At any rate V in general has several components, that is, $V = V_1^* \times \cdots \times V_k^*$. We have now the following:

Theorem 8.5 [6]. A system cannot be controllable if there exists a number $r < k$ and a function $F: V^r \to V$, where $V^r = V_1^* \times \cdots \times V_r^*$, such that the diagram
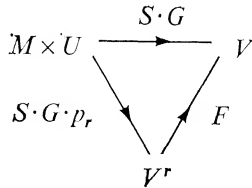
is commutative, where p_r is the projection map, $p_r: V \to V^r$.

Theorem 8.6 [6]. Let S and G be linear and V a module. Then S cannot be controllable iff the conditions from Theorem 8.5 are satisfied.

Applications of these theorems give various controllability conditions, such as those for classical point-wise state-space controllability and for functional reproducibility. It should be noticed that Theorem 8.5 holds for nonlinear systems too and that the linearity assumption makes the conditions necessary and sufficient.

6. *Stability.* We shall consider here only the stability of the null-input part of the state-transition function. Let Z be a topological space (although the closure space would suffice) with the topology θ. A subset $Z' \subset Z$ is *stable*, for a system S, iff, for any $\alpha \in \theta$, $Z' \subset \alpha$, there exists $\beta \in \theta$, $Z' \subset \beta$, such that for any $z \in \beta$

$$\phi_{x_0}(z, x_0, t_0, t) \in \alpha$$

for all $t \in T$, where ϕ_{x_0} is the null-input state transition. In other words, the set Z' is stable iff for any open neighborhood α of Z' there exists another open neighborhood β such that the state of the system will remain in α whenever the initial state is in β. Apparently the stability depends on the topology which the state space has, that is, on the concept of neighborhood implied by this topology. Every system can be viewed as stable by appropriate selection of a topology. Of course, in general, the topology cannot be selected at will but is given by the problem at hand.

Let F be a function on Z into a linearly ordered and lower-bounded set V, $F: Z \to V$. Then F is a *Lyapunov function* iff:

   (i) $(\forall z)(\forall z')[z' = \phi(z, x_0, t, t') \to F(z) \geq F(z')]$;
   (ii) $(\forall v)(\exists \alpha)[v \in V \ \& \ Z' \subset \alpha \ \& \ z \in \alpha \to F(z) \leq v]$;
   (iii) $(\forall \alpha)(\exists v)[v \in V \ \& \ Z' \subset \alpha \ \& \ z \in Z \sim \alpha \to F(z) \geq v]$.
We have now the following:

Theorem 8.7 [9]. A set $Z' \subset Z$ is stable for a system S whenever there exists a Lyapunov function.

To make the existence of a Lyapunov function both necessary and sufficient additional structure ought to be introduced in Z, probably the uniformity, that is, θ has to be a uniform topology. However, this question has not been fully explored as yet. A related study can be found in [20].

8.5. GENERAL SYSTEMS THEORY AND METAMATHEMATICAL PROBLEMS

To demonstrate the unifying power of the proposed general systems approach we shall show how the questions of consistency and completeness in the logical systems, in the sense of Gödel, can be considered in the proposed formalism. Starting from this result many questions of computability and linguistics can be treated similarly.

Let $S \subset X \times Y$ be a system and $\rho: Z \times X \to Y$ its state representation, that is,

$$(\exists z)[\rho(z, x) = y] \text{ iff } (x, y) \in S.$$

Apparently, for every $z \in Z$, there is given a function $S_z: X \to Y$ such that, for any input $x \in X$, $S_z(x) = \rho(z, x)$ whenever defined. The system S is the union of such functions: $S = \bigcup \{S_z: z \in Z\}$. To any $z \in Z$ corresponds a subset of inputs X_z, the domain of S_z, $X_z = \mathscr{D}(S_z)$. A set $X' \subset X$ will be termed *acceptable* by ρ if there exists $z \in Z$ such that $X' = X_z$.

Let $W \subset Y$ be a set of outputs which are considered as undesirable, and V be the set of the system's outputs, that is, $V = \{y: (\exists z)(\exists x)(y = \rho(z, x))\}$ is a state representation. The system is consistent (relative to W) iff $V \cap W = \varnothing$ and complete iff $V \cup W = Y$. We shall now give the characterization of a class of systems which cannot be both consistent and complete.

To investigate the conditions for the consistency and completeness of S_z, let g be an injection $g: Z \to X$; g will be termed *Gödel (restricted) mapping*. In reference to a given Gödel mapping we can now define a norm (or *diagonalization*) for a state $z \in Z$, $y_z = \rho(z, g(z))$. Let U be an arbitrary subset of Y. There is defined, then, a set of states Z_U^d whose norm is in U, that is,

$$z \in Z_U^d \leftrightarrow \rho(z, g(z)) \in U$$

Let X_U^d be the image of Z_U^d under g, $X_U^d = g(Z_U^d)$. We can now give the following theorem, stating a sufficient condition for the inconsistency or incompleteness of S:

Theorem 8.8 [5]. Let S be a given system, ρ its state representation, and W a set of undesirable (*forbidden*) outputs. If $X_W{}^d$ is an acceptable set of inputs, the system is either inconsistent or incomplete.

The general systems consistency and completeness theory can be directly applied to obtain various results in metamathematics. The procedure consists of two steps: (1) construct a system in the sense given in Section 8.4 that corresponds to the given axiomatic mathematical structure and establish the counterparts of the notions of acceptable inputs, norm, consistency, and so on; and (2) prove the desired theorem as an application of Theorem 8.1. Application of Theorem 8.8 to derive classical theorems of Gödel, Rosser, and Tarski is reported in [5].

8.6. GENERAL SYSTEMS THEORY OF HIERARCHICAL SYSTEMS

Finally we shall show how the proposed formalism can be used in the large-scale systems area or, more specifically, in the study of multilevel, hierarchical systems [22]. At the same time we shall show the usefulness of the goal-seeking approach since some of the subsystems will be represented as decision-making units.

A block diagram of the system under consideration is given in Figure 8.1. The basic subsystems are the following. A controlled process, $P: M \to Y$, is decomposed into n interacting subprocesses, $P_i: M_i \times U_i \to Y_i$, where $M = M_1 \times \cdots \times M_n$ are controls, $Y = Y_1 \times \cdots \times Y_n$ are outputs, and $U = U_1 \times \cdots \times U_n$ are the interactions which can be expressed in terms of controls by the interaction functions, $K_i: M \to U_i$. There are n first-level control subsystems defined by the respective subprocess models (assumed to be exact), $P_i: M_i \times U_i \to Y_i$, and a local performance function, $G_i: M_i \times Y_i \times \{\beta\} \to V$, where β is the coordination term while V is the value set. Finally, there is a second-level subsystem, termed a coordinator, whose task is to select the coordination term β so that the first-level units, while being concerned with their local control problems, will actually satisfy the overall control objective for the total system defined by the total process, $P: M \to Y$, and an overall performance function, $G: M \times Y \to V$. Assume that:

1. The overall objective is to minimize G with the constraint given by P. Denote by $\hat{m}$ a corresponding optimal control.

2. The local control objective is to minimize G_i, for any given coordination term β, by selecting optimal values for both m_i and u_i, the optimal pair being denoted by $(\hat{m}_i{}^\beta, \hat{u}_i{}^\beta)$.

P   Overall process
$P_1, \ldots, P_n$ Local subprocesses
$D_1, \ldots, D_n$ First-level decision units
D_0 Coordinator
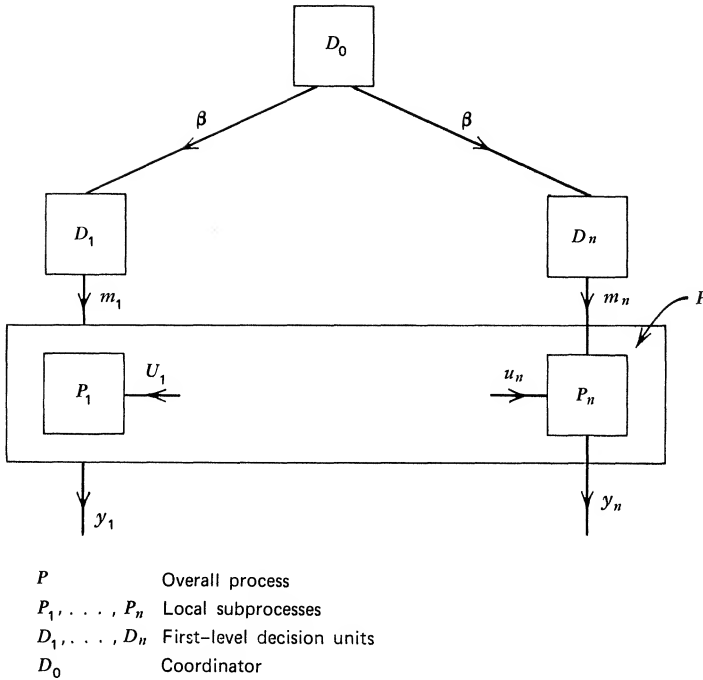
Figure 8.1. Basic block diagram of hierarchical systems.

3. The coordination strategy which the second-level unit uses to select the coordination term β is based on the following reasoning: Although each of the first-level units selects an optimal pair $(\hat{m}_i^\beta, \hat{u}_i^\beta)$ only the control $\hat{m}_i^\beta$ can be applied to the process P. Denote by u_i^β the (actual) interaction which occurs when the locally optimal controls $\hat{m}^\beta = (\hat{m}_1^\beta, \ldots, \hat{m}_n^\beta)$ are applied, that is, $u_i^\beta = K_i(\hat{m}^\beta)$. In general, for an arbitrary β, the locally optimal and the actual interactions are different, $u_i^\beta \neq \hat{u}_i^\beta$ or $K_i(\hat{m}^\beta) \neq \hat{u}_i^\beta$. A coordination strategy, termed the interaction balance, then requires that the second-level unit select $\hat{\beta}$ so that the actual and locally optimal interactions are equal, that is, $\hat{u}_i^\beta = u_i^\beta$ or $\hat{u}_i^\beta = K_i(\hat{m}^\beta)$.

The question then is, Under which conditions is coordination by interaction balance possible? Since the subprocesses and the respective models are already defined, the conditions ought to be given in terms of the relationship between the overall and the local performances, G and $G_1, \ldots, G_n$, which up to now are arbitrary. If the balance $\hat{u}_i^\beta = u_i^\beta$, when achieved, necessarily implies that the overall optimum is achieved, we shall say that the *interaction*

balance principle is applicable. If in addition there exists a coordination for which such a balance is reached, we shall say that the *system is coordinable by the interaction balance principle.*

We have now the following results:

Theorem 8.9 [22]. Balance of interactions, that is, the condition $\hat{u}^\beta = u^\beta$, implies the attainment of the overall optimum whenever:

(i) There exists a function, $\psi: V^n \to V$, such that for all $m \in M$

$$\psi(G_1(m_1, P_1(m_1, u_1), \beta), \ldots, G_n(m_n, P_n(m_n, u_n), \beta)) = G(m, P(m))$$

whenever $u = K(m)$.

(ii) ψ is an order-preserving (monotone) function.

Theorem 8.10 [22]. Let conditions (i) and (ii) of Theorem 8.9 hold. The system is coordinable, that is, the optimal coordination $\hat{\beta}$ which results in the balance exists, only if

$$\max_{B} \min_{M \times U} \psi = \min_{M} G \qquad (8.3)$$

where B is the set of all coordination terms, $B = \{\beta\}$.

Theorem 8.11 [22]. Let conditions (i) and (ii) hold; furthermore ψ is strictly monotone. Condition of Eq. 8.3 is then necessary and sufficient for the existence of an optimal coordination $\hat{\beta}$.

There are many other theorems on general systems and more specific levels about the coordination processes in multilevel systems. Of interest here, however, is only the fact that by using mathematically precise and general formalism we were able to make concrete and definite, useful and general statements about the hierarchical systems. This is in sharp contrast with the verbose approach to general systems theory whereby only hints and suggestions are offered as conclusions, and with the approaches based on more classical foundations, notably differential (and difference) equations and automata, whereby the conclusions, although precise, are limited to consideration of systems which do not have explicitly recognized subsystems, that is, to "small-scale" systems. The usefulness of these results is based on their generality. For example, the preceding theorems are true for an extremely broad class of systems, including dynamic, nonlinear, and even nonnumerical, algorithmic types. It appears that interaction balance coordination strategy is fundamental indeed. Many additional general systems theory results about multilevel systems can be found in [22], where other coordination principles are also considered.

8.7. CONCLUSION

In conclusion, it is of interest to ask a question about the usefulness of general systems theory, or the purpose of its further development. The *raison d'être* of general systems theory seems to rest on five functions that it ought to perform.

1. *Unification of various branches of systems theory.* There are a number of situations wherein the general systems theory, by providing a unifying basis for various branches of systems theory, will help in the process of mathematical model building. For example, in a given experimental situation it is not known a priori which of the mathematical models available is most suitable, and the general systems theory provides a framework to relate and evaluate various alternatives. More generally, the model-building process can be viewed as consisting of the following steps: verbal description of the observations and assumptions; general systems model; specific, detailed mathematical model suitable for the analysis or simulation studies. Presently, detailed mathematical models (e.g., differential equations or linear programming) are applied directly on the observed data in order that certain well-developed techniques can be used. In such an approach one has to make many assumptions not present in the observed data. By adding the general systems model, even if only as an intermediate step in the model-building process, one is in possession of a precise, although not detailed, mathematical description which is as close as possible to the phenomena actually observed.

2. *Application of the systems models in new areas.* In many areas of application the use of systems models was hampered by the conditions that had to be satisfied in order to use a mathematical model with sufficient accuracy, for example, continuity and differentiability conditions for the differential equations models. General systems models require very few assumptions and can be applied in many areas, such as social and political systems, where more restrictive models are simply too poor as approximations. Actually, even the observations described solely by verbal statements can be formalized in the general systems framework. This capability should make possible a more definite application of the systems approach in many areas of social and humanistic studies.

3. *Provision of a precise language for multidisciplinary problems and interdisciplinary communications.* Use of the knowledge developed in one field within another field and the study of problems involving several disciplines are greatly hampered by the diversity of concepts used to describe phenomena which have similar (or even the same) formal structures. What is really needed is development of a precise language which can be used to describe the invariant, structural aspect of the observations, divorced from the specific

phenomenological interpretations. General systems theory can be used to define many of these concepts and can serve as a language for scientific investigations in such areas as the social sciences and humanities, as well as for the study of large systems involving a variety of diverse factors (e.g., social, political, ecological, and technological).

4. *Means for describing large-scale systems.* The concept of a large-scale system cannot be attributed solely to the type of system under consideration; it depends also on the approach taken in describing the phenomena involved. For example, what might be considered as a small system for a social scientist can be a very large system for a psychologist. It is the number of details taken into account and the way in which they are described that is crucial. For example, a national economy is certainly a large system, but if it is described by a set of, say, five differential equations, some if not all of them being linear, it hardly deserves to qualify as a large-scale system. For a system to be viewed as large scale, it is necessary that the "largeness" be reflected in the model which describes the system. A prime example is a system which consists of a (large) number of subsystems whose existence is explicitly recognized, as in the multilevel hierarchical systems considered in Section 8.6. For the study of such large-scale systems it is very important to have as simple a description of the subsystems as possible so that one can be concerned with the problems of the overall system, rather than becoming bogged down with the complexity of subsystems behavior. General systems theory provides such a description, as demonstrated in Section 8.6 for the multilevel systems.

5. *Synthesis and organization of human knowledge in rather diverse but still related areas.* This is of importance in attacking and solving many practical problems, for example, in ecology. Another endeavor in need of such a synthesis is the field of *education.* General systems theory provides a framework for such a broad synthesis and may well become a cornerstone of any scientific or science-oriented education of the future, much as physics, chemistry, and mathematics serve this function today.

What properties should a general systems theory have to satisfy all the needs indicated? It ought to be *precise* (mathematical), *nonrestrictive* (general), and *elegantly* (simply) formulated. It is with these qualities in mind that the theory described in this chapter was proposed, was developed up to this point, and, hopefully, will continue to evolve.

REFERENCES

1. **Eckman, D. P.,** and **M. D. Mesarovic,** "On Some Basic Concepts of General Systems Theory." In *Proceedings of the Third International Congress on Cybernetics,* pp. 104–118. Namur, Belgium, 1961.

2. **Mesarovic, M. D.,** "Toward the Development of a General Systems Theory." *Neue Technik*, Journal of the Swiss Federation of Automatic Control, Vol. 5, No. 7, pp. 370–377, July 1963.

3. **Mesarovic, M. D.,** "Foundations for a General Systems Theory." In *Views on General Systems Theory*, M. D. Mesarovic (Editor), pp. 1–24. New York: Wiley, 1964.

4. **Mesarovic, M. D.,** "Auxiliary Functions and Constructive Specification of General Systems." *Mathematical Systems Theory*, Vol. 2, No. 3, pp. 203–222, September 1968.

5. **Mesarovic, M. D.,** "On Some Metamathematical Results as Properties of General Systems." *Mathematical Systems Theory*, Vol. 2, No. 4, pp. 357–361, December 1968.

6. **Mesarovic, M. D.,** "Controllability of General Systems." *Mathematical Systems Theory*, Vol. 5, No. 3, 1971.

7. **Mesarovic, M. D.,** "Mathematical Theory of General Systems and Some Economic Problems." In *Mathematical Systems Theory and Economics*, H. Kuhn and G. Szego (Editors), pp. 93–116. New York: Springer-Verlag, 1969.

8. **Mesarovic, M. D.,** "Systems Concepts." In *Scientific Thought*. Paris: UNESCO, 1970.

9. **Mesarovic, M. D.,** *Foundations for a Mathematical Theory of General Systems* (to appear).

10. **Windeknecht, T. G.,** "Mathematical Systems Theory: Causality." *Mathematical Systems Theory*, Vol. 1, No. 4, pp. 279–288, 1967.

11. **Macko, D.,** "Natural States and Past-Determinacy in General Time Systems." *Information Sciences*, Vol. 3, No. 1, pp. 1–16, January 1971.

12. **Bertalanffy, L. von,** *General System Theory*. New York: George Braziller, 1968.

13. **Zadeh, L.,** "From Circuit Theory to Systems Theory." *Proceedings of the IRE*, Vol. 50, No. 5, pp. 856–865, May 1962.

14. **Zadeh, L.,** "The Concept of State in System Theory." In *Views on General Systems Theory*, M. D. Mesarovic (Editor), pp. 39–50. New York: Wiley, 1964.

15. **Arbib, M.,** "Automata Theory and Control Theory—a Rapprochement." *Automatica*, Vol. 3, pp. 161–189, 1966.

16. **Arbib, M.,** *Theories of Abstract Automata*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.

17. **Wymore, A. W.,** *A Mathematical Theory of Systems Engineering: The Elements*. New York: Wiley, 1967.

18. **Kalman, R. E., P. L. Falb,** and **M. A. Arbib,** *Topics in Mathematical System Theory*. New York: McGraw-Hill, 1969.

19. **Balakrishnan, A. V.,** "State Space Theory of Linear Time-Verying Systems." In *System Theory*, L. A. Zadeh and E. Polak (Editors), pp. 95–125. New York: McGraw-Hill, 1969.

20. **Bushaw, D.,** "A Stability Criterion for General Systems." *Mathematical Systems Theory*, Vol. 1, No. 1, pp. 79–88, 1967.

21. **Mesarovic, M. D.,** "Systems Theory and Biology—View of a Theoretician." In *Systems Theory and Biology*, M. D. Mesarovic (Editor), pp. 59–87. New York: Springer-Verlag, 1968.

22. **Mesarovic, M. D., D. Macko,** and **Y. Takahara,** *Theory of Hierarchical, Multi-level Systems*. New York: Academic Press, 1970.

9. A Wattled Theory of Systems

A. WAYNE WYMORE

Department of Systems Engineering, The University of Arizona, Tucson

EDITOR'S COMMENTS

This chapter requires the reader to be of a certain maturity in mathematics. The prerequisites are approximately the same as those for Chapter 8, but knowledge of basic calculus, differential equations, and elementary abstract algebra may prove beneficial. For these areas, the following books are recommended:

Thomas, G. B., *Calculus and Analytic Geometry*, 4th ed. Reading, Mass.: Addison-Wesley, 1968.
Ziebur, A. D., *Topics in Differential Equations*. Belmont, Calif.: Dickenson Publishing Co., 1970.
Fraleigh, J. B., *A First Course in Abstract Algebra*. Reading, Mass.: Addison-Wesley, 1967.
Ayres, F., *Modern Algebra*. New York: Schaum, 1965.

Wymore employs a special mathematical notation, which is not used in the common mathematical literature and may be another source of difficulty for the reader. Therefore, we advice the use of the Glossary of Symbols at the end of this chapter (p. 298). Also, for a better understanding of the basic concepts introduced in Section 9.2, we strongly urge a study of the Wymore book [33].

9.1. INTRODUCTION

A wattle is a woven work comprised of many intertwined branches. The trend in general systems research that will be discussed in this chapter is wattled in the sense that it is a thatch composed of divers branches of mathematics, science, and engineering for the purpose of providing a roof under which the problems of *systems engineering* can be attacked.

The wattling occurs, however, in metamathematical motivations and evaluations. The approach to system-theoretic research to be discussed herein, is, in essence, mathematical. It relies for its validity on perfect mathematical rigor or, at least, on what passes for perfect mathematical rigor in these particular times. Thus, in the spirit that all mathematics is tautological, the distinctive flavor of the theory and its ultimate development are both completely determined by the basic definitions. These will be emphasized in this discussion. The approach is based on a set of postulates in which the concept of system is defined in terms of set- and function-theoretic concepts.

The principal criterion for development of this theory has been the degree of its usefulness in applications to problems in systems engineering. Most developments have been motivated by the necessity to create intellectual tools useful in modeling real phenomena. Much of the progress of this approach has been due to the desire to create a framework within which particular mathematical, scientific, and engineering techniques might be wattled.

This theory of systems has been developed to subsume both the theory of *discrete automata* [4] and the theory of *continuous systems* defined by differential equations [39].

A completely rigorous framework has been missing within which to state system concepts precisely, concepts, in particular, which have had vague meanings in the folklore of systems engineering: "time-varying," "adaptive," "artificial intelligence," and so forth [9].

Another motivation has been to provide a framework whereby extant system tools might be used consistently in concert. Tools of particular interest in this context are *the mathematical models of operations research* [30], *information/communication theory* [26, 16], and *linear control theory* [8]. A completely general but entirely rigorous framework within which each of these might find a place has been lacking.

The technique of *digital computer simulation* has been widely used in systems analysis and design, and appropriately so because simulations have proved to be extremely useful [19, 7]. Nonetheless, simulations are difficult to work with because the language in which they are couched is more appropriate to digital computers than it is to the description of

system-theoretic properties. What is needed here is a simulation language based on system-theoretic considerations, not on computer requirements.

Ultimately, the aim of mathematical system theory, from this point of view, is to lay a methodological foundation for the engineering of vast, complicated systems, systems necessary for the solution of social problems, exotic systems whose shapes and forms are not known.

An abstract theory of systems is hardly necessary if an engineer's problem is to design a better mousetrap. If the problem is to design a system to control the pestilential infestation of a set of human domiciles, however, then the engineer must have some methodological guidelines.

From this point of view, the basic manipulations of systems that are important are the putting together of *simple* system components to achieve *complex* system behavior and the manipulation inverse to that one: that of taking a very *complicated* system and breaking it down into *simpler* components interconnected by input-output relationships.

Thus several historical intellectual branches are herein interwoven to produce a wattled theory of systems. There is a thread from mathematical logic, represented by the papers of Turing [28] and of McCulloch and Pitts [18]. These generated the extant theory of discrete automata, whose greatest achievement so far, from the point of view of the basic manipulations of systems described above, is that of Krohn and Rhodes [15].

Another branch contributing to this wattled theory of systems represents the work of the topological algebraists: for example, that of Pontryagin in *topological groups* [24] and that of Hille in the theory of *topological semigroups* [11]. Such works as these gave rise, on the one hand, to the field of *topological dynamics* [3] and, on the other hand, to the study of topological algebraic structures such as *Banach algebras* [17] and the *qualitative theory of differential equations* [20].

The theory of differential equations, however, has determined yet another branch leading to mathematical system theory, an engineering branch, which resulted finally in what is now known as *modern control theory* [1]. This theory must also be subsumed by mathematical system theory.

Then there is a more or less pseudo-mathematical branch leading to mathematical system theory represented by the writings in the area of what is known as *general systems theory* [2]. General systems theory evolved more concretely into the area of *cybernetics* [32] and more concretely still into what is called *computer science* [31]. A theory of systems which aspires to lay a foundation for systems engineering must subsume these basic areas also.

The final thread leading into mathematical system theory is represented by *information theory* and its offshoots. Information theory [26] is such a beautiful theory that its limited application to any but a very narrow type of system is surprising. But the whole theory of *stochastic processes* [6] is also

included in this general branch of mathematical system theory, and these, of course, have had extensive applications.

The attempt to intertwine and subsume all of these branches is the reason for calling this, in particular, a *wattled theory of systems.* It is an intertwining of many branches which, nonetheless, may yet form a suitable roof under which a great deal of useful and pertinent systems engineering may be accomplished.

9.2. DESCRIPTION OF THE APPROACH

The first object of the approach to a wattled theory of systems is to answer the question, *What is a system?* An intuitive notion of a system is that of *a black box with inputs and outputs.*

In the *theory of automata* the input is usually described as an alphabet. Some member of this alphabet of inputs presents itself for processing by the black-box system at every integral value of time.

Systems described by *differential equations* usually are thought of as having a forcing function providing input to the system. If a theory of systems is to subsume both these types of systems, the idea of the input alphabet must be modified somewhat.

The time scale of a discrete system is some subset of the set of all integers. The time scale of a system described by differential equations is usually thought of as some subset of the real numbers which is a continuum. One way to include both of these intuitions into a theory of systems is to assume, for heuristic purposes, that there is some " real " time scale. That is, the time scale that is ascribed to a given system is the time scale of a set of experiments which will be performed on the system. A discrete system, then, would correspond to a system in an experimental situation where real time is going on but where the experimental data are sampled at discrete intervals of time. Thus, in this context, even for discrete systems, it is not sufficient to talk only about an input alphabet; it is now necessary to talk about an *input function*, a function which is defined everywhere over real time.

The actual inputs themselves can still be left quite arbitrary.

A set of input functions appropriate for systems models must also have some characteristics demanded by this intuitive notion of the existential condition of systems. One of these notions is the translation of the input function. If f is an input function for a system, and if an experimenter on the system is to be allowed to specify an arbitrary origin for the time scale of the experiment, at, say, the point r on the " real " time scale, then the translation of f by the amount r must also be a legitimate input function for the system. The translation of f by the amount r, evaluated at any time, is the same as

the value of the input function f evaluated at the time $r + t$. Thus, associated with every input function f, is the set of all its translates. This gives the system-theoretic approach the ability to set the origin of the time scale at any instant of the " real " time scale.

Furthermore, if the concept of experimenting on systems is to be retained, another manipulation on input functions must be allowed. If an experiment is to begin at time 0, and at time 0 an arbitrary input function is to be generated for the system, then that total input history composed of the history of input to the system up to time 0, and from time 0 on into the future, the input function generated by the experimenter, must be a legitimate input function for the system. This requires that all functions that can be generated by this process result in legitimate input functions. Let f and g be legitimate input functions to the system, and let the segmentation of f and g be defined as follows: the segmentation of f and g is an input function of the system, defined at time t as being $f(t)$ if t is negative and as $g(t)$ if t is not negative.

Following Pontryagin, a set of functions which has these characteristics will be called an *admissible set of input functions*. These notions are formalized in the following definition.

Definition 9.1. Admissible sets of input functions. Let P be a set not empty. Let $f, g \in FUNCTIONS(REALS, P)$, and let $r \in REALS$, then

$translation(f, r) = \{(t, f(r + t)): t \in REALS\}$ and
$segmentation(f, g) = \{(t, f(t)): t \in NEGATIVEREALS\}$
$\qquad\qquad \cup \{(t, g(t)): t \in NONNEGATIVEREALS\}.$

A set F is an admissible set of input functions with values in P if and only if:

$\qquad F \subset FUNCTIONS(REALS, P),$
$\qquad F \neq \varnothing$, if $f, g \in F$, and $r \in REALS$, then

$\qquad translation\ (f, r) \in F \quad$ and $\quad segmentation\ (f, g) \in F.$

The set of all admissible sets of input functions with values in P is denoted $ADMISSIBLES(P)$.

If G is an arbitrary subset of $FUNCTIONS(REALS, P)$, then the smallest admissible set of input functions which contains G is denoted $ADMISSIBLE\text{-}SET(G)$ and is defined as follows:

$$ADMISSIBLESET(G) = \bigcap \{F: F \in ADMISSIBLES(P), F \supset G\}.$$

Discussion. In accordance with Turing and the school of discrete automata theorists, the basic concept of what goes on inside the black box will be represented by a set of states and the change from one state to another.

Thus, to describe a system, it is necessary to describe the inputs, the admissible set of input functions, the set of states, the time scale, the set of permissible

transformations of the set of states (i.e., the totality of the behavior available to the system), and, finally, the way in which this behavior is causally or (better) functionally coordinated to the input functions and the time scale.

These notions are all brought together under the heading of an assemblage; this definition is given next.

Definition 9.2. Assemblages. An *assemblage* is a 6-tuple $Z = (S, P, F, M, T, \sigma)$, where:

> S is a set not empty;
> P is a set not empty;
> $F \in ADMISSIBLES(P)$;
> $M \subset FUNCTIONS(S, S)$, such that $identity(S) \in M$;
> $T \subset REALS$ and $O \in T$;
> $\sigma \in FUNCTIONS(F \times T, \text{onto}, M)$.

If Z is an assemblage and $Z = (S, P, F, M, T, \sigma)$, then the states of Z, the inputs of Z, the set of input ports of Z, the set of input functions of Z, the set of state transitions of Z, the time scale of Z, and the state-transition function of Z are denoted, respectively: $STATES(Z)$, $INPUTS(Z)$, $INPUT$-$PORTS(Z)$, $INPUTFUNCTIONS(Z)$, $BEHAVIOR(Z)$, $TIMESCALE(Z)$, and $motion(Z)$, and are defined, respectively, as follows:

$STATES(Z) = S$;
$INPUTS(Z) = P$;
$INPUTPORTS(Z)$
$= I$ if $P = \times I$ for some set I not empty of sets not empty and $\#(I) \geq 2$
  (if I is a set not empty of sets not empty, then the Cartesian product of the sets in I is denoted $\times I$ and is defined as follows:

$$\times I = \{x: x \in FUNCTIONS(I, \bigcup I), x(A) \in A \text{ for every } A \in I\}),$$

$= \{P\}$ otherwise;
$INPUTFUNCTIONS(Z) = F$;
$BEHAVIOR(Z) = M$;
$TIMESCALE(Z) = T$;
$motion(Z) = \sigma$.

If $f \in F$, $t \in T$ and $x \in S$, then the state of the assemblage Z at time t, given the input function f and the initial state x, is $(\sigma(f, t))(x)$.

If $f \in F$ and $x \in S$, then the time trajectory of Z determined by f and x is denoted *timetrajectory*(Z, f, x) and is defined as follows:

$$timetrajectory(Z, f, x) = \{(t, (\sigma(f, t))(x)): t \in T\}.$$

An output function for Z is any $\zeta \in FUNCTIONS(S, Q)$, where Q is any set not empty.

Discussion. The concept of an assemblage is still too broad, however. It encompasses mathematical constructs which do not satisfy intuitive notions of what systemlike behavior should be. For example, it is possible to have an assemblage whose state at time t depends on values of the input function which will be received beyond t, that is, assemblages which are intrinsically anticipatory.

Three conditions are necessary to rule out such constructs and to set in the concrete of mathematical postulates intuitive notions of what constitutes systemlike behavior and intuitive notions of what constitutes the state of the system at time t.

Definition 9.3. Systems. A *system* is an assemblage $Z = (S, P, F, M, T, \sigma)$ such that

$\sigma(f, o) = identity(S)$ for every $f \in F$;

$\sigma(translation(f, s), t)\sigma(f, s) = \sigma(f, s + t)$ for every $f \in F$, s, $t \in T$ such that $s + t \in T$;

$\sigma(f, r) = \sigma(g, r)$ for every $f, g \in F$, and $r \in REALS$, if

$$restriction(f, T[o, r)) = restriction(g, T[o, r)) \quad \text{when } r \geq 0,$$

or if

$$restriction(f, T[r, o)) = restriction(g, T[r, o)) \quad \text{when } r < 0.$$

Discussion. It is here, at this point, clearly stated what, precisely, a system is There is no longer any question. A system is a 6-tuple such that and so forth, as indicated in Definitions 9.2 and 9.3.

The next question on the agenda to be answered is, *What does it mean to interconnect systems?*

The main purpose of the discussion will be to attempt to find a framework within which complicated interconnections of system components can be deduced from the behavior of the individual components. The first objective is to formalize the idea of a coupling recipe which essentially answers precisely the following questions: Which systems are involved in the intercoupling? Which of the input ports of each system are assigned to receive input from which other systems? What are the output functions from systems providing inputs to input ports of receiving systems?

The idea of a coupling recipe is made explicit in the following definition.

Definition 9.4. Coupling recipes. A *couple* is a triple $C = (\mathscr{Z}, a, o)$, where $\mathscr{Z}$ is a set not empty of systems, such that for every $Z, Z' \in \mathscr{Z}$;

$STATES(Z) \neq STATES(Z')$,

$TIMESCALE(Z) = TIMESCALE(Z')$,

$\{(projection(INPUTPORTS(Z) \cap INPUTPORTS(Z'))f:$

$f \in INPUTFUNCTIONS\}$

$= \{(projection(INPUTPORTS(Z) \cap INPUTPORTS(Z'))f:$
$f \in INPUT FUNCTIONS (Z')\};$

a is a function defined on $\mathscr{L}^2$ such that for every $Z, Z', Z'' \in \mathscr{L}, a(Z, Z') \subset$
$INPUTPORTS(Z)$ and $a(Z, Z') \cap a(Z, Z'') = \varnothing$, if $Z' \neq Z''$;

o is a function defined on $\mathscr{L}^2$, such that for every $Z, Z' \in \mathscr{L},$
$o(Z, Z') \in FUNCTIONS(STATES(Z), \bigtimes a(Z', Z))$ if $a(Z', Z) \neq \varnothing$;
$o(Z', Z) = \varnothing$ if $a(Z', Z) = \varnothing;$

$\bigcup \{INPUTPORTS(Z) \sim a(Z, Z'): Z' \in \mathscr{L}, Z \in \mathscr{L}\} \neq \varnothing.$

If $C = (\mathscr{L}, a, o)$ is a couple and $Z \in \mathscr{L}$, then the set of system components of the resultant determined by C, the input port assignments of C, the output function assignments of C, the set of input ports of Z designated as occupied by C, the set of input ports of Z left unoccupied by C, the set of states of the resultant determined by C, the total set of input ports managed by C, the total input determined by C, the set of total input functions determined by C, the set of unoccupied input ports of the resultant determined by C, and the time scale of the resultant determined by C are denoted, respectively: $COMPONENTS(C)$, $inputports(C)$, $outputs(C)$, $OCCUPIEDPORTS(Z, C)$, $UNOCCUPIEDPORTS(Z,C)$, $RESULTANTSTATESET(C)$, $TOTAL$-$INPUT PORTS(C)$, $TOTALINPUT(C)$, $TOTALINPUTFUNCTIONS(C)$, $TOTALUNOCCUPIEDPORTS(C)$, $RESULTANTINPUTS(C)$, and $RESULTANTTIMESCALE(C)$, and are defined, respectively, as follows:

$COMPONENTS(C) = \mathscr{L};$

$inputports(C) = a;$

$outputs(C) = o;$

$OCCUPIEDPORTS(Z, C) = \bigcup \{a(Z, Z'): Z' \in \mathscr{L}\};$

$UNOCCUPIEDPORTS(Z, C) = INPUTPORTS(Z)$
$\sim OCCUPIEDPORTS(Z, C);$

$RESULTANTSTATESET(C) = \bigtimes \{STATES(Z'): Z' \in \mathscr{L}\};$

$TOTALINPUTPORTS(C) = \bigcup \{INPUTPORTS(Z'): Z' \in \mathscr{L}\};$

$TOTALINPUT(C) = \bigtimes TOTALINPUTPORTS(C);$

$TOTALINPUTFUNCTIONS(C)$
$= \{f: f \in FUNCTIONS(REALS, TOTALINPUT(C)), (projection(INPUT$-$PORTS(Z')))f \in INPUTFUNCTIONS(Z')$ for every $Z' \in \mathscr{L}\};$

$TOTALUNOCCUPIEDPORTS(C)$
$= \bigcup \{UNOCCUPIEDPORTS(Z', C): Z' \in \mathscr{L}\};$

$RESULTANTINPUTS(C)$
$= \bigtimes TOTALUNOCCUPIEDPORTS(C);$

$RESULTANTINPUTFUNCTIONS(C)$
$= \{f: f \in FUNCTIONS(REALS, RESULTANTINPUTS(C))$, there exists $g \in TOTALINPUTFUNCTIONS(C)$ such that $f = (projection(TOTALUNOCCUPIEDPORTS(C)))g\};$

$RESULTANTTIMESCALE(C) = TIMESCALE(Z).$

Discussion. In Definition 9.4 a resultant determined by the coupling is mentioned but is not explicitly defined. Certainly, it is desirable that the resultant of coupling systems together according to some coupling recipe be a system. The state-transition function of such a system can be defined, however, only if the coupling recipe also determines a coupling function which satisfies certain equations. The development of these equations can be seen as follows.

As indicated in Definition 9.4, the input to the resultant will not be the same as the input to any of the individual component systems because some or all of the input ports of any individual component system could be designated as occupied by the coupling recipe. Nonetheless, if there is to be a way of deducing the state-transition function of the resultant system from the state-transition function of the component systems, there must be a mapping defined on the set of restricted inputs of the resultant together with the initial states of the resultant, with values in the totality of input functions of the component systems reflecting the conditions of coupling.

It is clear that the resultant of an arbitrary couple may not be a system in the sense of Definition 9.3, but the following definition gives the way in which the resultant can be constructed on the basis of the existence of a coupling function.

Definition 9.5. System coupling. Let $C = (\mathscr{Z}, a, o)$ be a couple. A function κ is a *coupling function* with respect to the couple C if and only if:

$\kappa \in FUNCTIONS(RESULTANTINPUTFUNCTIONS(C) \times RESULT\text{-}$
$ANTSTATESET(C), TOTALINPUTFUNCTIONS(C));$

for every $f \in RESULTANTINPUTFUNCTIONS(C)$, $x \in RESULTANT\text{-}$
$STATESET(C)$, $t \in RESULTANTTIMESCALE(C)$, and $V \in TOTAL\text{-}$
$INPUTPORTS(C)$:

$(\kappa(f, x))(t)(V)$
$= (f(t))(V)$ if $V \in TOTALUNOCCUPIEDPORTS(C)$,
$= ((o(Z, Z'))(((motion(Z))((projection(INPUTPORTS(Z)))\kappa(f, x), t))$
$\quad (x(STATES(Z))))))(V)$
$\quad$ if $V \notin TOTALUNOCCUPIEDPORTS(C)$, $V \in a(Z', Z)$ for some $Z, Z' \in \mathscr{Z}$;
$\quad$ if $g \in TOTALINPUTFUNCTIONS(C)$, $x \in RESULTANTSTATESET(C)$,
$s \in RESULTANTTIMESCALE(C)$, and $(g(t))(V)$
$= ((o(Z, Z'))(((motion(Z))((projection(INPUTPORTS(Z)))g, t))(x(STATES$
$\quad (Z)))))(V)$ for every $V \in a(Z', Z)$, for every $Z, Z' \in \mathscr{Z}$, and for every $t \in$
$\quad RESULTANTTIMESCALE(C)[o, s)$ if $s \geq 0$, or for every $t \in RESULT\text{-}$
$\quad ANTTIMESCALE(C)[s, 0)$ if $s < 0$, then $(\kappa(projection(TOTALUN\text{-}$
$\quad OCCUPIEDPORTS(C))g, x))(t) = g(t)$ for every
$\quad t \in RESULTANTTIMESCALE(C)[0, s)$ if $s \geq 0$, or
$\quad$ for every $t \in RESULTANTTIMESCALE(C)[s, 0)$ if $s < 0$;

if f, $g \in RESULTANTINPUTFUNCTIONS(C)$ and $s \in RESULTANT\text{-}$
$TIMESCALE(C)$, and if
$restriction(f, RESULTANTTIMESCALE(C)[0, s))$
$= restriction(g, RESULTANTTIMESCALE(C)[0, s))$ when $s \geq 0$, or if
$restriction(f, RESULTANTTIMESCALE(C)[0, s))$
$= restriction(g, RESULTANTTIMESCALE(C)[s, 0))$ when $s < 0$, then
$restriction(\kappa(f, x), RESULTANTTIMESCALE[0, s))$
$= restriction(\kappa(g, x), RESULTANTTIMESCALE[0, s))$ when $s \geq 0$, or
$restriction(\kappa(f, x), RESULTANTTIMESCALE[s, 0))$
$= restriction(\kappa(g, x), RESULTANTTIMESCALE[s, 0))$ when $s < 0$.

If C is a couple and κ is a coupling function with respect to C, then the resultant of C is an assemblage denoted $RESULTANT(C)$ and is defined as follows:

$STATES(RESULTANT(C)) = RESULTANTSTATESET(C);$
$INPUTS(RESULTANT(C)) = RESULTANTINPUTS(C);$
$INPUTFUNCTIONS(RESULTANT(C))$
$$= RESULTANTINPUTFUNCTIONS(C);$$
$BEHAVIOR(RESULTANT(C)) = RANGE(motion(RESULTANT(C)));$
$TIMESCALE(RESULTANT(C)) = RESULTANTTIMESCALE(C);$

if $f \in INPUTFUNCTIONS(RESULTANT(C))$, $t \in TIMESCALE(RE\text{-}$
$SULTANT(C))$, and $x \in STATES(RESULTANT(C))$, then

$(((motion(RESULTANT(C)))(f, t))(x))(STATES(Z))$
$= ((motion(Z))((projection(INPUTPORTS(Z)))\kappa(f, x), t))(x(STATES(Z)))$
    for every $Z \in \mathscr{Z}$.

A couple C is a system couple if and only if there exists a coupling function κ with respect to C.

Discussion. Subsequent developments in this wattled theory have shown that, when a coupling function exists, the resultant assemblage is well defined and is, indeed, a system. Coupling functions generally exist when the components all consist of discrete systems or all are differentiable systems.

An example will clarify most of the concepts of Definitions 9.4 and 9.5 better than extended discussion. Let

$C = (\mathscr{Z}, a, o)$, where
$\mathscr{Z} = \{Z_1, Z_2\}$, where
$Z_1 = (S_1, P_1, F_1, M_1, T_1, \sigma_1)$, where
$P_1 = \bigtimes \{P_{11}, P_{12}\}$, where
$P_{11} = REALS$,
$P_{12} = REALS$,
$F_1 = \{f: f \in FUNCTIONS(REALS, P_1), \int_0^t ((f(\tau))(P_{11}) + (f(\tau))(P_{12})\, d\tau \in$
$REALS$ for every $t \in REALS\}$,
$M_1 = RANGE(\sigma_1)$,

$T_1 = NONNEGATIVEREALS$, and, if $f \in F$, $t \in T$, $x \in S$,

$\sigma_1(f, t)(x) = x + \int_0^t ((f(\tau)(P_{11}) + (f(\tau))(P_{12})) \, d\tau$,

$Z_2 = (S_2, P_2, F_2, M_2, T_2, \sigma_2)$, where

$S_2 = REALS$,

$P_2 = REALS$,

$F_2 = \{f: f \in FUNCTIONS(REALS, P_2), \int_0^t f(\tau) \, d\tau \in REALS$ for every $t \in REALS\}$,

$M_2 = RANGE(\sigma_2)$,

$T_2 = NONNEGATIVEREALS$, and, if $f \in F_2$, $t \in T_2$, $x \in S_2$,

$\sigma_2(f, t)(x) = x + \int_0^t f(\tau) \, d\tau$;

$a = \{((Z_1, Z_1), \varnothing), ((Z_1, Z_2), \{P_{12}\}) ((Z_2, Z_1), \{P_2\}), ((Z_2, Z_2), \varnothing)\}$; and

$o = \{((Z_1, Z_1), \varnothing), ((Z_1, Z_2), identity(S_2)), ((Z_1, Z), identity(S_1)), ((Z_2, Z_2), \varnothing)\}$.

This coupling recipe is caricatured in Figure 9.1. Since the only output functions involved are the identity functions on the state set, it is the state of each system that is transmitted to each other system in the couple.
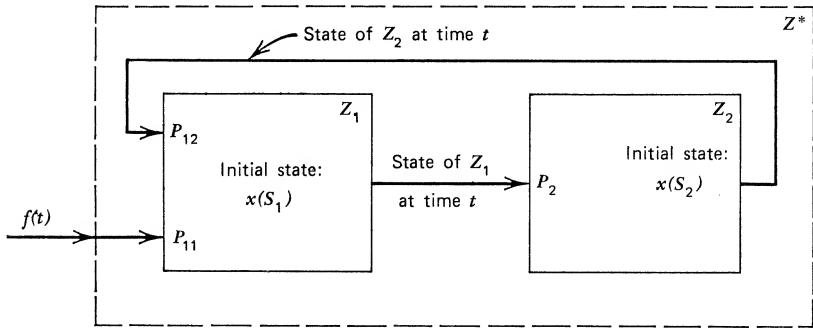
Figure 9.1. Caricature of the resultant of a system couple.

The coupling function κ, for each $f \in RESULTANTINPUTFUNC$-$TIONS(C)$ and $x \in RESULTANTSTATESET(C)$ [where, in this case, $RESULTANTINPUTFUNCTIONS(C) = \{f: f \in FUNCTIONS(REALS, P_{11}),$ there exists $g \in F_1$, $f = projection(\{P_{11}\})g\}$, and $RESULTANTSTATE$-$SET(C) = \bigtimes \{S_1, S_2\}]$, must determine inputs to each of the input ports P_{11}, P_{12}, and P_2. Hence, for each $t \in RESULTANTTIMESCALE(C) (= NON$-$NEGATIVEREALS$, in this case), $(\kappa(f, x))(t)$ is a vector with three dimensions whose components are indicated as $((\kappa(f, x))(t))(P_{11})$, $((\kappa(f, x))(t))(P_{12})$, and $((\kappa(f, x))(t))(P_2)$.

From Figure 9.1 or from Definition 9.5, it is clear that $(\kappa(f, x))(t)$ must satisfy the following set of equations:

if $V \in TOTALINPUTPORTS$ ($= \{P_{11}, P_{12}, P_2\}$, in this case), then
$$((k(f, x))(t)(V)$$
$$= f(t) \quad \text{if } V = P_{11},$$
$$= (\sigma_2((projection(\{P_2\}))\kappa(f, x), t)(x(S_2)) \quad \text{if } V = P_{12},$$
$$= (\sigma_1((projection(\{P_{11}, P_{12}\}))\kappa(f, x), t)(x(S_1))) \quad \text{if } V = P_2 ;$$
or, in view of the definitions of σ_1 and σ_2,
$$((\kappa(f, x))(t))(V)$$
$$= f(t) \quad \text{if } V = P_{11},$$
$$= x(S_2) + \int_0^t (\kappa(f, x))(\tau)(P_2) \, d\tau \quad \text{if } V = P_{12},$$
$$= x(S_1) + \int_0^t (f(\tau) + ((\kappa(f, x))(\tau))(P_{12})) \, d\tau \quad \text{if } V = P_2.$$

The solution of this set of equations for the two unknown functions, $(projection(\{P_{12}\}))\kappa(f, x)$ and $(projection(\{P_2\}))\kappa(f, x)$ (which can be found, in this case, by the standard methods of linear differential equations), can be written as follows:

$$((\kappa(f, x))(t))(P_{12})$$
$$= \sinh(t) \times x(S_1) + \cosh (t) \times x(S_2) + \int_0^t \sinh (t - \tau) \times f(\tau) \, d\tau,$$
$$((\kappa(f, x))(t))(P_2),$$
$$= \cosh(t) \times x(S_1) + \sinh (t) \times x(S_2) + \int_0^t \sinh (t - \tau) \times f(\tau) \, d\tau.$$

If $Z^* = RESULTANT(C)$, then $Z^* = (S^*, P^*, F^*, M^*, T^*, \sigma^*)$, where
$$S^* = \bigtimes \{S_1, S_2\},$$
$$P^* = P_{11},$$
$$F^* = \{f : f \in FUNCTIONS(REALS, P_{11}), \text{ there exists } g \in F \text{ such that } f = (projection(\{P_{11}\}))g\},$$
$$M^* = RANGE(\sigma^*),$$
$$T^* = NONNEGATIVEREALS,$$
if $f \in F^*$, $t \in T^*$, $x \in S^*$, then
$$(\sigma^*(f, t)(x))(S_1)$$
$$= (\sigma_1((projection(\{P_{11}, P_{12}\}))\kappa(f, x), t))(x(S_1)) \quad \text{(by Definition 9.5)},$$
$$= x(S_1) + \int_0^t (f(\tau) + \sinh (\tau) \times x(S_1) + \cosh (\tau) \times x(S_2) + \int_0^\tau \sinh (\tau - \tau')$$
$$\times f(\tau') \, d\tau') \, d\tau \quad \text{(by the definition of } \sigma_1 \text{ and of } ((\kappa(f, x))(\tau))(P_{12})),$$
$$= \cosh (t) \times x(S_1) + \sinh (t) \times x(S_2) + \int_0^t f(\tau) \, d\tau + \int_0^t \int_0^\tau \sinh (\tau - \tau')$$
$$\times f(\tau') \, d\tau' \, d\tau; \text{ and}$$
$$(\sigma^*(f, t)(x))(S_2)$$
$$= (\sigma_2((projection(\{P_2\}))\kappa(f, x), t))(x(S_2)) \quad \text{(by Definition 9.5)},$$
$$= x(S_2) + \int_0^t (\cosh (\tau) \times x(S_1) + \sinh (\tau) \times x(S_2) + \int_0^\tau \cosh (\tau - \tau')$$
$$\times f(\tau') \, d\tau') \, d\tau \quad \text{(by the definitions of } \sigma_2 \text{ and of } ((\kappa(f, x))(\tau))(P_2)),$$
$$= \sinh (t) \times x(S_1) + \cosh (t) \times x(S_2) + \int_0^t \int_0^\tau \cosh (\tau - \tau') \times f(\tau') \, d\tau' \, d\tau.$$

These are the basic concepts: the concept of the *assemblage*, the concept of *system*, the concept of a *couple*, and the concept of the *resultant of a couple*.

These definitions make explicit exactly what a system is and what it means to couple systems together to make more complex systems.

These concepts also make clear several other parts of the folklore of system theory, for example, the ideas of subsystem and component. The following definition makes these terms explicit.

Definition 9.6. Subsystems and components. Let Z and Z' be systems. Then Z is a *subsystem* of Z' if and only if

$STATES(Z) \subset STATES(Z')$;
$INPUTS(Z) \subset INPUTS(Z')$;
$INPUTFUNCTIONS(Z) \subset INPUTFUNCTIONS(Z')$,
$TIMESCALE(Z) \subset TIMESCALE(Z')$;

and, for every $f \in INPUTFUNCTIONS(Z)$ and $t \in TIMESCALE(Z)$, $(motion(Z))(f, t) = restriction((motion(Z'))(f, t), STATES(Z))$.

A system Z is a component of a system Z' if and only if there exists a system couple C such that $Z \in COMPONENTS(C)$ and Z' is isomorphic to $RESULTANT(C)$.

Discussion. The concepts of subsystem and component are *not* the same; it is clear that they are distinct and differing. In fact, they are dual concepts in a very real way. The theory can be sketched rapidly as follows.

A system can be *decomposed* into its isolated subsystems. There always exists a maximal decomposition of this sort (which may be trivial). Then the given system is said to be the disjunction of its isolated subsystems. If Z is the given system and $\mathscr{Z}$ is the maximal set of isolated subsystems, the above relationship is symbolized as $Z = OR(\mathscr{Z})$.

A system can be *resolved* into its independent components. There always exists a maximal resolution of this sort (which may be trivial). Then the given system is said to be the conjunction of its independent components. If Z is the given system and $\mathscr{Z}$ is the maximal set of independent components, the above relationship is symbolized as $Z = AND(\mathscr{Z})$. [The system $AND(\mathscr{Z})$ can also be characterized as the resultant of the trivial system couple $C = (\mathscr{Z}, \{((Z', Z''), \varnothing): Z', Z'' \in \mathscr{Z}\}, \{((Z', Z''), \varnothing): Z', Z'' \in \mathscr{Z}\}).$]

If $Z = (S, P, F, M, T, \sigma)$ is a system, Q is a set not empty, and G is a subset of $FUNCTIONS(S, Q)$, then the dual of Z with respect to G is an assemblage Z^*, defined as follows:

$STATES(Z^*) = G$,
$INPUTS(Z^*) = P$,
$INPUTFUNCTIONS(Z^*) = F$,
$BEHAVIOR(Z^*) = RANGE(motion(Z^*))$,
$TIMESCALE(Z^*) = -T$;

if $f \in INPUTFUNCTIONS(Z^*)$, $t \in TIMESCALE(Z^*)$, $\zeta \in STATES(Z^*)$,

and $x \in S$, then

$$(((motion(Z^*))(f, t))(\zeta))(x) = \zeta(\sigma(translation(f, t), -t)(x)).$$

The first duality theorem gives conditions sufficient to ensure that the dual of Z with respect to G is a system.

The second duality theorem states that Z is always isomorphic to a dual of the dual of Z.

The third duality theorem says that, if $\mathscr{Z}$ is a set of systems and if $DUALS(\mathscr{Z})$ is a set of duals of the systems in $\mathscr{Z}$, then the conjunction of the systems in $DUALS(\mathscr{Z})$ is a system dual to $OR(\mathscr{Z})$; or, expressed in suggestive symbolism: $DUAL(OR(\mathscr{Z})) = AND(DUALS(\mathscr{Z}))$.

This suggests a way to compute the maximal conjunctive resolution of a system Z:

$$Z = \begin{cases} DUAL(DUAL(Z)) & \text{(by the first and second duality theorems),} \quad \text{or} \\ DUAL(OR(\mathscr{Z})) & \text{(e.g., by inspection of the state-transition diagram of } DUAL(Z) \text{ to discover the set } \mathscr{Z} \text{ of isolated subsystems of } DUAL(Z)), \quad \text{or} \\ AND(DUALS(\mathscr{Z})) & \text{(by the third duality theorem).} \end{cases}$$

Undoubtedly it has been noted that the concept of system isomorphism was introduced in Definition 9.6.

One of the basic tools for the development of a wattled theory of systems such as this is the system homomorphism, and its specialization, the system isomorphism. On the basis of these two definitions and the definition of subsystem, what it means to simulate a system and to implement a system can be defined.

Definition 9.7. System homomorphisms. Let Z and Z' be systems; $Z = (S, P, F, M, T, \sigma)$ and $Z' = (S', P', F', M', T', \sigma')$. Then Z is the *homomorphic-(ρ, μ, θ) image* of Z' if and only if

$\rho \in REALS$ and $T = \rho \times T'$;

$\mu \in FUNCTIONS(F', \text{onto}, F)$ such that, for every $f, g \in F'$ and $r \in REALS$, μ (*segmentation*(f, g)) = *segmentation*$(\mu(f), \mu(g))$ and $\mu(translation(f, r)) = translation(\mu(f), \rho \times r)$; and

$\theta \in FUNCTIONS(S', \text{onto}, S)$ such that, for every $f \in F'$, $t \in T'$, $x \in S'$, $\theta(\sigma'(f, t)(x)) = \sigma(\mu(f), \rho \times t)(\theta(x))$.

A system Z is an *homomorphic image* of a system Z' if and only if there exist ρ, μ, and θ such that Z is the homomorphic-(ρ, μ, θ) image of Z'.

A system Z is *isomorphic*(ρ, μ, θ) to a system Z' if and only if Z is the homomorphic-(ρ, μ, θ) image of Z' and μ and θ are 1 to 1.

A system Z is *isomorphic* to a system Z' if and only if there exist ρ, μ, and θ such that Z is isomorphic-(ρ, μ, θ) to Z', or Z' is isomorphic-(ρ, μ, θ) to Z.

A system ·Z′ *simulates* a system Z if and only if there exists a subsystem Z″ of Z′ such that Z is an homomorphic image of Z″.

Let $\mathscr{Z}$ be a set not empty of systems. Then a system Z is *implementable* in $\mathscr{Z}$ if and only if there exists a system couple C such that

$$COMPONENTS(C) \subset \mathscr{Z}$$

and $RESULTANT(C)$ simulates Z.

Discussion. The fundamental theorems of homomorphisms will not be presented here: the conditions under which an assemblage, which is an homomorphic image of a system, is a system; construction, from the artifacts of Z′, of a system Z″ isomorphic to Z if Z is an homomorphic image of Z′ (analogous to similar theorems in group theory); and the fact that any system Z′, of which Z is an homomorphic image, satisfies the same input-output specifications as Z does. More will be said presently about input-output specifications and their satisfaction by systems.

There are two archetypical subclasses of systems: discrete systems and systems defined by differential equations. Definition 9.8 makes these classes of systems explicit. The definitions depend for validity on the proofs of preliminary theorems, which are not included here, asserting that the constructs given actually constitute systems.

Definition 9.8. Discrete systems and differentiable systems. Let A and B be sets not empty and let $b \in B$. Then the function constant on A and equal to b is denoted $constant(A, b)$ and is defined as follows: $constant(A, b) = \{(a, b): a \in A\}$.

A system $Z = (S, P, F, M, T, \sigma)$ is a *discrete system* if and only if

$F \supset \{constant(REALS, p): p \in P\}$;

$T = NONNEGATIVEINTEGERS$;

and, for every $f \in F$, $t \in T$, and $x \in S$:

$\sigma(f, t)(x)$

$= x$ if $t = 0$,

$= (\sigma(constant(REALS, f(t-1)), 1)\sigma(f, t-1))(x)$ if $t \neq 0$.

A system $Z = (S, P, F, M, T, \sigma)$ is a *differentiable system* if and only if there exists a topological vector space E such that

$S = VECTORS(E)$;

$F \supset \{constant(REALS, p): p \in P\}$;

$T = NONNEGATIVEREALS$;

and, for every $f \in F$, $t \in T$, and $x \in S$, $\sigma(f, t)(x)$ is the value, $y(t)$, of the function y at time t, where

$y(0) = x$, and for every $s \in T$,

$$\frac{dy}{d\tau}(s) = \frac{d(timetrajectory(Z, constant(REALS, f(s)), y(s)))}{d\tau} \quad (0)$$

(or, less precisely, but more suggestively,

$$\frac{dy}{d\tau}(s) = \frac{d(\sigma(constant(REALS, f(s)), \tau)(y(s))))}{d\tau}(0)),$$

where the implied limits exist in $VECTORS(E)$ with respect to $TOPOLOGY(E)$.

Discussion. From Definition 9.8, it can be seen that a discrete system is completely determined by specifying

$$\sigma(constant(REALS, p), 1)(x)$$

for every $p \in P$, and $x \in S$.

From Definition 9.8, it can be seen also that a differentiable system is completely determined by specifying

$$\frac{d(\sigma(constant(REALS, p), \tau)(x))}{d\tau}(0)$$

for every $p \in P$ and $x \in S$.

If Z is a differentiable system and h is a positive real number, then a discrete system Z_h, approximating Z, can be determined as follows:

$$\sigma_h(constant(REALS, p), 1)(x) = x + h \times \frac{d(\sigma(constant\ REALS, p), \tau)(x))}{d\tau}(0)$$

for every $p \in P$ and $x \in S$ (where the algebraic operations indicated by $+$ and $\times$ are understood to represent the vector sum and the scalar product, respectively, in the vector space E).

Conversely, if for every $h \in POSITIVEREALS$ a discrete system Z_h is defined with the same inputs and the same state space for each h, then a differentiable limiting system Z is determined by setting:

$$\frac{d(\sigma(constant(REALS,\ p), \tau)(x))}{d\tau}(0)$$

$$= limit\left(\left\{\left(h, \frac{\sigma_h(constant(REALS, p), 1)(x) - x}{h}\right):\right.\right.$$

$$\left.\left. h \in POSITIVEREALS\right\}, \geq\right),$$

provided the limit of the indicated net exists in $S = VECTORS(E)$ with respect to $TOPOLOGY(E)$ for every $p \in P$ and $x \in S$.

As a simple illustration of the application of the concept of homomorphisms consider the following assertion: many discrete systems are simulatable by discrete semigroup systems.

If G is a semigroup, then a discrete semigroup system, $Z = (S, P, F, M, T, \sigma)$, is determined as follows:

$S = P = ELEMENTS(G)$,

$F = FUNCTIONS(REALS, P)$,

if $f \in P$ and $x \in S$, then $\sigma(constant(REALS, p), 1)(x) = (operation(G))(p, x)$.

If $Z = (S, P, F, M, T, \sigma)$ is an arbitrary discrete system such that $F = FUNCTIONS(REALS, P)$, then there is a natural discrete semigroup system Z' associated with Z; Z' is determined by the semigroup G according to the construction given above, where $ELEMENTS(G) = M$ and $operation(G)$ is the composition of the mappings in M: $Z' = (S', P', F', T', \sigma')$, where

$S' = P' = M$,

$F' = FUNCTIONS(REALS, P')$,

if $\beta \in P'$, $\alpha \in S'$, and $x \in S$, then $(\sigma(constant(REALS, \beta), 1)(\alpha))(x)) = \beta(\alpha(x))$.

Now suppose that in M all the mappings $\sigma(c_p, 1)$ are distinct; that is, assume that, if p, $p' \in P$ and $p \neq p'$, then $\sigma(c_p, 1) \neq \sigma(c_{p'}, 1)$. Suppose, furthermore, that there exists $x_0 \in S$ such that every state, $y \in S$, is reachable from x_0; that is, for every $y \in S$, there exist $f_y \in F$ and $t_y \in T$ such that $\sigma(f_y, t_y)(x_0) = y$.

Then Z' simulates Z. In fact, consider the discrete subsystem Z'' of Z', defined as follows:

$S'' = S'$,

$P'' = \{\sigma(c_p, 1): p \in P\} \subset P'$,

$F'' = FUNCTIONS(REALS, P'')$,

if $f \in F''$, $t \in T''$, then $\sigma''(f, t) = \sigma'(f, t)$.

Then Z is the homomorphic$(1, \mu, \theta)$ image of Z'', where, for every $f \in F''$, $\alpha \in S''$,

$(\mu(f))(t) = \phi(f(t))$ for every $t \in REALS$, where, for every $p \in P$,

$\phi(\sigma(c_p, 1)) = p$, and

$\theta(\alpha) = \alpha(x_0)$.

The crucial relation that must be proved is that, if $f \in F''$, $t \in T''$, and $\alpha \in S''$, then $(\sigma''(f, t)(\alpha)) = \sigma(\mu(f), t)(\theta(\alpha))$. If $t = 0$, then

$\theta(\sigma''(f, 0)(\alpha))$

$= \theta(\alpha)$ (by the definition of σ''),

$= \alpha(x_0)$ (by the definition of θ),

$= \sigma(\mu(f), 0)(\alpha(x_0))$ (by the definition of σ),

$= \sigma(\mu(f), 0)(\theta(\alpha))$ (by the definition of θ).

If $\theta(\sigma''(f, t - 1)(\alpha)) = \sigma(\mu(f), t - 1)(\theta(\alpha))$, then

$\theta(\sigma''(f, t)(\alpha))$

$= (\sigma''(f, t)(\alpha))(x_0)$ (by the definition of θ),

$= (\sigma''(constant(REALS, f(t-1), 1)\sigma''(f, t-1)(\alpha))(x_0)$ (by the definition of σ''),

$= (f(t-1))((\sigma''(f, t-1)(\alpha))(x_0))$ (by the definition of $\sigma''(constant(REALS, f(t-1)), 1$ as the composition of mappings),

$= (f(t-1))(\theta(\sigma''(f, t-1)(\alpha)))$ (by the definition of θ),

$= (f(t-1))(\sigma(\mu(f), t-1)(\theta(\alpha)))$ (by the induction hypothesis),

$= \sigma(constant(REALS, \phi(f(t-1))), 1)\sigma(\mu(f), t-1)(\theta(\alpha))$ (by the definition of ϕ, for, if

$f(t-1) = \sigma(constant(REALS, p), 1)$ for some $p \in P$, then $\phi(f(t-1)) = p$ and

$f(t-1) = \sigma(constant(REALS, \phi(f(t-1))), 1))$,

$= \sigma(constant(REALS, (\mu(f))(t-1)), 1)\sigma(\mu(f), t-1)(\theta(\alpha))$ (by the definition of μ),

$= \sigma(\mu(f), t)(\theta(\sigma))$ (by the definition of σ).

So much for the basic concepts of this wattled theory of systems.

The basic principle involved in guiding the development of the theory is to stick very close to mathematical rigor because the intuition of most people with regard to complex system phenomena is not good. There is nothing, therefore, to guide the intuition except mathematical rigor. On the other hand, it is important that the system theory not be subservient to currently popular mathematical techniques. It is a principle here to develop the mathematics necessary to advance the system theory, not to apply any particular branch of mathematics that happens to be extant.

On the other hand, it seems that almost any mathematical tool that is available *can* be applied in this particular context. For example, the extensive development of the *theory of topological vector spaces* can be applied here by imposing, as indicated in Definition 9.8, the structure of a topological vector space on the state space. If the mappings in *BEHAVIOR*(Z) of the system Z are all linear, then some results concerning system resolution can be obtained by applying the *theory of the spectral resolution* to the mappings in *BEHAVIOR*(Z).

Stochastic processes can be studied in this context by imposing a probability distribution on *INPUTS*(Z) or on *INPUTFUNCTIONS*(Z), or even on *STATES*(Z) if the initial state of the system must be regarded as a random variable, or even on *TIMESCALE*(Z) if the time of the observation of the system is a random variable. It may be desirable to consider a joint probability distribution on *INPUTFUNCTIONS*(Z) $\times$ *TIMESCALE*(Z) $\times$ *STATES*(Z).

A basic system structure has been proposed in this chapter. No matter what system structure is adopted, however, problems remain, as they have always existed, with respect to a mathematical theory of systems. First of all,

it is necessary to be very precise concerning what a system is; perhaps, ultimately, there will be a theory in which the undefined term is "system." Meanwhile we can, at least, be precise about what a system is in terms of set-theoretic structures. The second problem for any theory of systems is to determine what it means to couple systems together to form more complex systems and to deduce the behavior of the resulting kludge. The problem inverse to that one is equally important: given a complicated kludge, to resolve it into simple components. The third basic problem of any system theory must be to create a class of models powerful enough to represent any engineering phenomenon of interest. "Engineering" here is taken in its broadest sense, to include the design and analysis of health delivery systems, wealth delivery systems, ecological management systems, and any other sort of system phenomena which our society will need to model in order to design precise solutions to social problems.

Hence one of the main methods for the development of this theory is to look at the needs of society and of engineering in its broadest sense and to attempt to wattle the solution to the problem.

9.3. STATE OF THE ART

Some aspects of the state of the art of this wattled theory of systems were indicated in Section 9.2.

The elements of the theory, as outlined in Section 9.2, have been published in book form [33]. In this book, the basic definitions are derived from intuitive understandings and from numerous examples. There it is shown formally how this wattled theory of systems subsumes the ideas of Turing machines, discrete systems generally, and differentiable systems. Much of the classical work in discrete systems is described there and generalized to the system concept as given above. It is in this book that the work on duals is developed extensively. It is shown there how a wide variety of engineering phenomena may be modeled by the system-theoretic constructs of this wattled theory.

Elsewhere [35] the relationship between continuous systems and discrete systems is extensively explored within the wattled theory of systems. Some of the implications of this work for education in systems engineering are explored in nonmathematical language in [34].

One of the most important developments of this wattled theory of systems engineering is oriented toward the development of a *methodology for systems engineering*. This methodology must include a definition of what constitutes input-output specifications for a system, as well as what is meant by a system satisfying a set of input-output specifications. An introduction to this method-

ological discussion is contained in [33], the book cited above. A great deal more work has been done subsequently, however, including a discussion of measures of effectiveness on the universe of systems which satisfy a given set of input-output specifications. In order that decisions may be made with respect to system design, the range of a measure of effectiveness must be partially ordered. This is one half of the methodological problem. The other half consists in the specification of the technology within which a system solution must be found. A specification of the technology then determines the universe of systems implementable within that technology. A formal definition of what it means for a system to be implementable in a set of systems is given in Definition 9.7.

The solution to a system design problem, of course, must exist in the intersection of these two universes of systems: the final system must satisfy the input-output specifications and be implementable in the given technology. A measure of effectiveness on the universe of systems satisfying the input-output specifications reflects the definition of what constitutes a good solution in terms of the system satisfying the input-output specification. The universe of systems implemented in the technology must also have a measure of effectiveness defined on it, reflecting reliability, maintainability, produceability, and so forth. The range of this measure of effectiveness must also be partially ordered. All of this theory is discussed extensively in a book soon to be published entitled, *A Notebook of Systems Engineering Methodology*. The mathematics employed in that book are somewhat less than perfectly rigorous.

The discipline of this wattled theory of systems has been used extensively in developing mathematical models of complex systems. Some of these can be found in [33], but the work of Norling in stream-flow management systems [21], that of Cross in open-pit mining systems [5], that of Goodkin in oceanic transport management systems [38], and that of Wadsworth in computer-assisted instructional systems [29] are particularly worthy of mention.

In 1970, Dr. Tuncer Oren designed a compiler whose source language is that of this wattled theory of systems [22]. He called this language GEST, an acronym for *GEneral System Theory implementor*. Statements into the compiler take this form: let Z be a system, where $STATES(Z)$ equals so and so, $INPUTS(Z)$ equals so and so, and so on. One can specify a system couple with the input port assignments and the output function definitions, and then ask for a particular time trajectory or output trajectories of the resultant system thus defined. The computer will deduce, from the specifications of the systems and the interconnections, the behavior of the resultant system and compute the outputs as desired. When the language and compiler are fully implemented, this wattled theory of systems will really be available for immediate application to problems of systems engineering.

9.4. PRESENT EFFORTS AND FUTURE PERSPECTIVES

The wattle that is being described here still has a great many holes in it. In fact, it could fairly be said that only the basic framework has been established and just a few of the rafters have been laid. A great deal of work remains to be done to fill in the gaps. To this end, there are seven basic research thrusts, which can be characterized by the following generic terms: algebraic, topological, probabilistic (stochastic), modeling, methodology, computer implementation, and bridges. Each of these research thrusts will be discussed briefly in turn.

The objective of the *algebraic approach* to research in this wattled theory of systems is mainly to subsume the work of the discrete automata theorists— to subsume from that area, in particular, the applications of the theory of semigroups to the computation of resolutions of systems into components. This is mainly a process of generalization, clarification and application.

The thrust in the *topological direction* is mainly to explain, to subsume, and to generalize what has been called linear control theory in engineering and to place in this same context the theories of stability and of optimization of systems. All three of these basic areas require extensive algebraic structures such as vector spaces, as well as topological constructs, the means for discussing limits and continuity. Furthermore, classical treatments of these areas require a calculus of vector-valued functions. Much of the work that has been done thus far in this direction has consisted of reformulating many of the basic concepts and much of the symbolism used in point-set topology and topological vector spaces. The classical symbolism is adequate when the principal objects of study are point-set topologies and topological vector spaces. But if these kinds of constructs are to be imposed on another mathematical structure, represented by the system-theoretic construct, the classical symbolism is inadequate. A great deal of work is being done to remedy this situation. This effort will yield dividends in another direction as well. It is clear that there are no natural topologies in nature, and hence the topologies to be imposed on a given model are also at the discretion of the modeler. A strategic choice of a topology may yield interesting and useful system properties.

The third direction in which research is progressing is toward the subsumption of *stochastic processes* into this wattled theory of systems. The basic system structure described in this paper can be used to discuss stochastic processes simply by imposing probability measures on the inputs, the input functions, the state space, or the time scale, or on all of these. Then probability statements of various kinds can be made about the system behavior.

It is possible to discuss the adequacy of a model from a statistical point of view or to choose a statistically optimum model.

The same criticism voiced concerning classical symbolism used in the discussion of point-set topology can be made about the basic symbolism in the theory of probability. The difficulty here arises not so much in imposing probability measures on the basic system-theoretic artifacts, as in using probability theory as a modeling tool. It is quite clear that in the definition or imposition of measures of effectiveness on various sets of systems, probability measures will play an important part, but these will be extremely arbitrary, based not only on empirical data but also on subjective appraisals, as well as on the desirability of outcome. Therefore, some way must be found to make it easy to define arbitrary probability measures on arbitrary spaces. The way suggested by the classical theory is simply too awkward: to define a sigma ring of subsets and then to impose on it a countably additive measure is just too much to ask, especially if such a construct is to be developed on the basis of intuition or of subjective beliefs.

The fourth direction for research in this wattled theory of systems is toward the development of *specific models* of interest in the systems engineering context. The principal class of models that is the target for development here consists of system-theoretic models of human behavior, and here human behavior is intended to encompass not only the internal psychological and physiological behavior of an individual human being, but also human interactions in small groups and human interactions in large groups. The point of the development of such models is to provide information useful in the design of systems which have such human aspects: aspects of the design of a man-machine interface, the design of interfaces between man and man in human organizations, and the design of interfaces between systems and the social environment in which they are embedded.

The fifth thrust for research in this theory of systems is toward the development of a precise and rigorous *design methodology*. The methodological framework described in Section 9.3 and more extensively in the forthcoming book, *A Notebook of Systems Engineering Methodology*, still has many gaps in it from the mathematical point of view as well as from the point of view of practical implementation of the framework. The ideas of input-output specifications and the satisfaction of input-output specifications by systems are not hard to define; thus it is possible to arrive at definitions which are satisfying both intuitively and practically. These definitions then raise questions of existence: What are necessary and sufficient conditions that there exists a system which satisfies a given set of input-output specifications? There are a few theorems extant in this direction, but none of them is very complete or satisfactory.

If, for a given set of input-output specifications, a system satisfying them

can be found and defined mathematically, then, in general, such a system will not be immediately implementable within a given real technology. Therefore this system has to be described in a great deal more detail. If a system satisfies a set of input-output specifications, then any system of which that given system is a homomorphic image will also satisfy the same input-output specifications. Hence one can proceed to define ever more detailed systems simply by choosing a system in the set of all systems of which a given one is a homomorphic image. Are there mathematical techniques by which this selection can be made? Where does the classical systems engineering technique of system function analysis fit into this framework? Where does the technique of reticulation of the input fit?

What are useful definitions for measures of effectiveness, and how can they be defined in practice? To find algorithms for the definition of the measures of effectiveness and for the partial ordering on their ranges is an extremely important practical problem for systems engineering.

Much more powerful techniques for finding resolutions of systems are needed. The achievement of Krohn and Rhodes is impressive, but it is a far cry from what is actually needed in systems engineering practice. One of the limitations of the Krohn-Rhodes theory is that only cascade and conjunctive resolutions have been considered. It is quite clear that feedback interconnections also need to be considered, since these will quite often yield great economies in system design and development. (If the Krohn-Rhodes lead were followed, this would necessitate the studying of a symmetric wreath product.) Furthermore, it is necessary to discover theorems dealing with the resolution of systems, or rather their implementation, in very restricted classes of technology. It is not possible to assume that one has at his command all finite-state machines or all discrete automata. It is necessary to assume, for example, that there is a very definite upper bound on the number of states and the number of inputs that can be used to implement a given system design, and theorems in this direction are badly needed.

The sixth thrust in research connected with this wattled theory of systems is concerned with the *design of systems involving computers* whose aim is to make available in practical form to practicing systems engineers and systems engineering organizations the fruits of the research in the other five areas. The work of Oren [22] is a step in the right direction, but further development, expansion, and implementation are urgently needed.

A great deal of research still remains to be done in the numerical analysis necessary for computation within the system-theoretic framework; that is, the development of discrete approximations to continuous systems needs much more work. Sometimes it is impractical to implement a given continuous model directly on the digital computer; on occasion it is necessary to approxi-

mate a continuous system with a discrete system, and this in itself is a real system-theoretic problem fraught with difficulties because of the primary problem of assigning a measure of effectiveness to the class of discrete systems reflecting intuitive or technical standards for a good approximation.

The seventh direction for research in this wattled theory of systems is perhaps not a direction for *research* at all. It is an ongoing effort to provide *bridges to other scientific and engineering disciplines and to ever-lower levels in the educational process*. This means mostly the production of textbooks and popularizations of the theory to make it available to people who are not particularly well trained in mathematics. These endeavors turn out to be surprisingly difficult. People who are well trained mathematically find it easier to think in terms of mathematics than to explain to a layman what the mathematics means. It would be an ideal world if everyone were well trained in mathematics and all one had to do was present the definitions, theorems, and proofs, leaving it up to each individual to interpret the theorems and to apply them in everyday practice. Unfortunately, this is impossible; it is necessary for someone to interpret the mathematics and to show how theories can be applied.

Someone recently said, "Systems engineering appears to be an extremely good thing—I only hope it has not arrived too late to save the world." But systems engineering has not arrived until a body of communicable methodology has been established. To this end, a great deal of effort needs to be expended in a relatively short time.

PROBLEMS

9.1. Let $A = \{0, 1, \alpha, a, b\}$, $B = \{0, \beta, a, c\}$, and $C = \{0, 1, b, \alpha, \gamma\}$. Then compute:

(a) $\bigcup \{A, B, C\}$,

(b) $\bigcap \{A, B, C\}$,

(c) $A \cap (B \cup C)$,

(d) $(A \cap B) \cup (A \cap C)$,

(e) $B \sim A$ (where $B \sim A = \{b : b \in B, b \notin A\}$),

(f) $C \sim (A \cup B)$,

(g) $B \sim (A \cap C)$,

(h) $(A \cup B) \sim C$,

(i) $(A \cap C) \sim A$,

(j) $(A \cap B) \sim (A \cap C)$,

(k) $A \times B \times C$ (where $A \times B \times C = \{(x, y, z) : x \in A, y \in B, z \in C\}$),

(l) $\bigtimes \{A, B, C\}$ (where $\bigtimes \{A, B, C\} = \{g : g \in FUNCTIONS(\{A, B, C\},$

$\bigcup \{A, B, C\}$), $g(A) \in A$, $g(B) \in B$, $g(C) \in C\}$, and where if X is a set and Y is a set then $FUNCTIONS(X, Y) = \{f : f \subset X \times Y$, for every $x \in X$ there exists $y \in Y$ such that $(x, y) \in f$; if (x, y), $(x', y') \in f$ and $x = x'$, then $y = y'\}$).

9.2. Prove or find a counterexample:

Let A, B, and C be sets not empty. Then

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

9.3. Prove or find a counterexample:

Let A and B be subsets of a set X. If $X \sim A \subset X \sim B$, then $A \subset B$.

9.4. Prove or find a counterexample:

Let A and B be sets and let f be a function defined on A with values in B. Let B_1 and B_2 be subsets of B. Then $(setinverse(f))(B_1 \cup B_2) = ((setinverse(f))(B_1)) \cap ((setinverse(f))(B_2))$ (where, for every $C \subset B$, $(setinverse(f))(C) = \{a : a \in A, f(a) \in C\}$).

9.5. Prove or find a counterexample:

Let P be a set not empty, and let $f, g \in FUNCTIONS(REALS, P)$; let $r, s \in REALS$; and let $h = step(f, r, g)$. Then $h = translation(segmentation(translation(f, r), translation(g, r)), -r)$ and $translation(h, s) = step(translation(f, s), r - s, translation(g, s))$ (where, if $n \in POSITIVEINTEGERS$, $\{e_0, \ldots, e_n\} \subset FUNCTIONS(REALS, P)$, and $\{t_1, \ldots, t_n\} \subset REALS$, such that $t_1 < \cdots < t_n$, then $step(e_0, t_1, e_1, \ldots, e_{n-1}, t_n, e_n) \in FUNCTIONS$-$(REALS, P)$ and is defined as follows for every $t \in REALS$:

$$(step(e_0, t_1, e_1, \ldots, e_{n-1}, t_n, e_n))(t) = \begin{cases} e_0(t) & \text{if } t < t_1, \text{ or} \\ e_1(t) & \text{if } t_1 \le t < t_2, \text{ or} \\ \cdots & \cdots, \text{ or} \\ e_i(t) & \text{if } t_i \le t < t_{i+1} \\ & \text{for } i \in INTEGERS[1, n-1], \text{ or} \\ \cdots & \cdots, \text{ or} \\ e_n(t) & \text{if } t_n \le t). \end{cases}$$

9.6. Prove or find a counterexample:

Let P_1 and P_2 be sets not empty. Let $\phi \in FUNCTIONS(P_1, P_2)$ and let F_1 be an admissible set of input functions with values in P_1. Let μ be a function defined on F_1 with values in $FUNCTIONS(REALS, P_2)$ defined as follows for every $f \in F$: $\mu(f) \in FUNCTIONS(REALS, P_2)$, so that for every $t \in REALS$

$$(\mu(f))(t) = \phi(f(t)).$$

Let $F_2 = RANGE(\mu)$. Then F_2 is an admissible set of input functions with values in P_2.

9.7. Prove or find a counterexample:

If F, $G \in ADMISSIBLES(P)$, where P is a set not empty, then $F \cap G \in ADMISSIBLES(P)$.

9.8. Prove or find a counterexample:

Let P be a set not empty and let F, $G \in ADMISSIBLES(P)$. Then $F \cup G \in ADMISSIBLES(P)$.

9.9. Prove or find a counterexample:

Let $a \in REALS$ and $F = \{f: f \in FUNCTIONS(REALS, REALS), \int_0^t (exponent(a \times (t - \tau)) \times f(\tau) \ d\tau \in REALS$ for every $t \in REALS\}$. Then $F \in ADMISSIBLES(REALS)$.

9.10. Let $Z = (S, P, F, M, T, \sigma)$ be an assemblage where

$S = \{a, b\}$;
$P = \{0, 1\}$;
$F = ADMISSIBLESET(\{constant(REALS, 0), constant(REALS, 1)\})$,
$M = RANGE(\sigma)$;
$T = NONNEGATIVEINTEGERS$; and

σ is defined as follows for every $f \in F$, $t \in T$, and $x \in S$:

$\sigma(f, t)(x)$
$= x$ if $t = 0$,
$= \sigma(constant(REALS, f(t - 1)), 1)\sigma(f, t - 1)(x)$ if $t \neq 0$, where
$\sigma(constant(REALS, 0), 1) = \{(a, a), (b, a)\}$,
$\sigma(constant(REALS, 1), 1) = \{(a, b), (b, a)\}$.

Enumerate M.

9.11. Let Z be a discrete system, $Z = (S, P, F, M, T, \sigma)$, where

$S = \{a, b, c\}$;
$P = \{0, 1, 2\}$;
$\sigma(constant(REALS, 0), 1) = \{(a, b), (b, c), (c, a)\}$;
$\sigma(constant(REALS, 1), 1) = \{(a, b), (b, b), (c, a)\}$;
$\sigma(constant(REALS, 2), 1) = \{(a, c), (b, c), (c, b)\}$.

Compute the state of the system Z at time 8, given the input function $step(constant(REALS, 0), 2, constant(REALS, 2), 3, constant(REALS, 0), 4, constant(REALS, 1), 6, constant(REALS, 2))$, and the initial state b.

9.12. Let Z be the system $Z = (S, P, F, M, T, \sigma)$, where

$S = REALS$;
$P = REALS$;
$F = \{f: f \in FUNCTIONS(REALS, REALS), \quad \int_0^t f(\tau) \ d\tau \in REALS$ for every $t \in REALS\}$;
$M = RANGE(\sigma)$;
$T = NONNEGATIVEREALS$;
and, for every $f \in F$, $t \in T$, and $x \in S$,
$\sigma(f, t)(x) = x + \int_0^t f(\tau) \ d\tau$.

Compute the state of the system Z at time 8.67, with the input function $\{(t, 3t^2): t \in REALS\}$, and the initial state 4.5.

9.13. Prove or find a counterexample:

Let $Z = (S, P, F, M, T, \sigma)$, where
$S = P \times P$;
P is an arbitrary set not empty;
$F = FUNCTIONS(REALS, P)$;
$M = RANGE(\sigma)$;
$T = NONNEGATIVEINTEGERS$;
and, for every $f \in F$, $t \in T$, and $(x, x') \in S$,

$$\sigma(f, t)(x, x') = \begin{cases} (x, x') & \text{if } t = 0, \text{ or} \\ (f(0), x) & \text{if } t = 1, \text{ or} \\ (f(t-1), f(t-2)) & \text{if } t \geq 2; \end{cases}$$

then Z is a system. Furthermore, let $\zeta \in FUNCTIONS(S, P)$ be defined as follows for every $(x, x') \in S$: $\zeta(x, x') = x'$. Then ζ is an output function for Z with values in P, and, for every $f \in F$, $t \in T$, and $(x, x')tS$,

$$\zeta(\sigma(f, t)(x, x')) = f(t - 2) \quad \text{if } t \geq 2$$

(i.e., Z is a two-unit delay).

9.14. Prove or find a counterexample:

Let P be a set not empty, and let $n \in POSITIVEINTEGERS$. Then there exist a discrete system $Z = (S, P, F, M, T, \sigma)$ and an output function ζ for Z such that
$F = FUNCTIONS(REALS, P)$,
and, for every $f \in F$, $t \in T$, and $x \in S$,

$$\zeta(\sigma(f, t)(x)) = f(t - n) \quad \text{if } t \geq n.$$

9.15. Prove or find a counterexample:

Let P be a set not empty, and let $d \in POSITIVEREALS$. Then there exist a system $Z = (S, P, F, M, T, \sigma)$ and an output function ζ for Z such that
$F = FUNCTIONS(REALS, P)$,
$T = NONNEGATIVEREALS$,
and, for every $f \in F$, $t \in T$, and $x \in S$,

$$\zeta(\sigma(f, t)(x)) = f(t - d) \quad \text{if } t \geq d.$$

9.16. Let $Z = (S, P, F, M, T, \sigma)$ be a differentiable system, where
$S = REALS$;
$P = REALS$;
$F \supset \{constant(REALS, p): p \in P\}$;
$T = NONNEGATIVEREALS$;

and, for every $p \in P$, $x \in S$,

$$\frac{d(\sigma(constant(REALS, p), \tau)(x))}{d\tau}(0) = p \times exponent(-x).$$

Let $\zeta \in FUNCTIONS(S, REALS)$ be the output function for Z, defined as follows for every $x \in S$: $\zeta(x) = exponent(x)$.

Compute the output of Z with respect to ζ at time 2, given the input function, $\{(t, exponent(t)): t \in REALS\}$, and the initial state, $\sqrt{2}$.

9.17. Let $C = (\{Z_1, Z_2, Z_3\}, \{((Z_i, Z_j), \emptyset): i, j \in \{1, 2, 3\}\}, \{((Z_i, Z_j), \emptyset): i, j \in \{1, 2, 3\}\})$, where, for each $i \in \{1, 2, 3\}$,

Z_i is a discrete system;
$Z_i = (S_i, P_i, F_i, M_i, T_i, \sigma_i)$, where
$S_i = \{a_i, b_i, c_i\}$,
$P_i = \{0, 1\}$,
$F_i = FUNCTIONS(REALS, P_i)$,
$\sigma_i(constant(REALS, 0), 1) = \{(a_i, a_i), (b_i, a_i), (c_i, b_i)\}$,
and
$\sigma_i(constant(REALS, 1), 1) = \{(a_i, b_i), (b_i, c_i), (c_i, a_i)\}$.
Let $Z^* = RESULTANT(C)$.

Compute the state of the system Z^* at time 3, given the input function, $step(constant(REALS, (0, 1, 0)), 1, constant(REALS, (1, 0, 0)), 2, constant(REALS, (1, 1, 0)), 3, constant(REALS, (1, 1, 1))))$, and the initial state (a_1, b_2, b_3).

9.18. Let $C = (\{Z_1, Z_2\}, \{((Z_1, Z_1), \emptyset), ((Z_1, Z_2), \emptyset), ((Z_2, Z_1), \{P_2\}), ((Z_2, Z_2), \emptyset)\}, \{((Z_1, Z_1), \emptyset), ((Z_1, Z_2), \{(a, I), (b, II)\}), ((Z_2, Z_1), \emptyset), ((Z_2, Z_2), \emptyset)\})$, where

$Z_1 = (S_1, P_1, F_1, M_1, T_1, \sigma_1)$ is a discrete system such that
$S_1 = \{a, b\}$,
$P_1 = \{1, 2\}$,
$F_1 = FUNCTIONS(REALS, P_1)$,
$\sigma_1(constant(REALS, 1), 1) = \{(a, a), (b, a)\}$,
$\sigma_1(constant(REALS, 2), 1) = \{(a, b), (b, b)\}$;
$S_2 = (S_2, P_2, F_2, M_2, T_2, \sigma_2)$ is a discrete system such that
$S_2 = \{A, B\}$,
$P_2 = \{I, II\}$,
$F_2 = FUNCTIONS(REALS, P_2)$,
$\sigma_2(constant(REALS, I), 1) = \{(A, B), (B, A)\}$,
$\sigma_2(constant(REALS, II), 1) = \{(A, A), (B, B)\}$.
Let $Z^* = RESULTANT(C)$.
Compute $\sigma^*(constant(REALS, 1), 1)$ and $\sigma^*(constant(REALS, 2), 1)$.

9.19. Let $C = (\{Z\}, \{(Z, Z), \{\{I, II\}\})\}, \{((Z, Z), \{(a, I), (b, I), (c, II)\})\})$, where $Z = (S, P, F, M, T, \sigma)$ is a discrete system such that
$S = \{a, b, c\}$,
$P = \{1, 2\} \times \{I, II\}$,
$F = FUNCTIONS(REALS, P)$,
$\sigma(constant(REALS, (1, I)), 1) = \{(a, a), (b, a), (c, a)\}$,
$\sigma(constant(REALS, (1, II)), 1) = \{(a, a), (b, b), (c, b)\}$,
$\sigma(constant(REALS, (2, I)), 1) = \{(a, b), (b, a), (c, c)\}$,
$\sigma(constant(REALS, (2, II)), 1) = \{(a, c), (b, a), (c, c)\}$.
Let $Z^* = RESULTANT(C)$.
Compute $\sigma^*(constant(REALS, 1), 1)$ and $\sigma^*(constant(REALS, 2), 1)$.

GLOSSARY OF SYMBOLS (IN ORDER OF APPEARANCE)

$FUNCTIONS(A, B)$ is the set of all functions defined on the set A with values in the set B.

$REALS$ is the set of real numbers.

$NEGATIVEREALS = \{t: t \in REALS, t < 0\}$.

$NONNEGATIVEREALS = \{t: t \in REALS, t \geq 0\}$.

$identity(S)$ is the identity function defined on the set S, that is, $identity(S) = \{(x, x): x \in S\}$.

$\#(I)$ is the cardinal number of the elements in set I.

$T[r, t) = \{s: s \in T, r \leq s < t\}$, where T is given as a subset of $REALS$ and r, $t \in REALS$. $T[r, t]$, $T(r, t]$, and $T(r, t)$ are defined similarly.

$restriction(f, B) = \{(b, f(b)): b \in B\}$, where, for some A and C,

$$f \in FUNCTIONS(A, C) \quad \text{and} \quad B \subset A.$$

$projection(\mathscr{L}) \in FUNCTIONS(\bigtimes \mathscr{A}, \bigtimes \mathscr{L})$, where $\mathscr{A}$ is a set of sets, $\mathscr{L}$ is a set of sets, $\mathscr{L} \subset \mathscr{A}$, and for every $x \in \bigtimes \mathscr{A}$, $A \in \mathscr{L}$, $((projection(\mathscr{L}))(x))(A) = x(A)$.

$INTEGERS = \{t: t \in REALS, t \text{ is an integer}\}$.

$NONNEGATIVEINTEGERS = INTEGERS \cap NONNEGATIVEREALS$.

$POSITIVEREALS = \{t: t \in REALS, t > 0\}$.

REFERENCES

1. Bellman, R., *Adaptive Control Processes: A Guided Tour*. Princeton, N.J.: Princeton University Press, 1961.

2. Bertalanffy, L. von, "General System Theory." *General Systems Yearbook*, Vol. 1, pp. 1–10, 1956.

3. Birkhoff, G. D., *Dynamical Systems*. American Mathematical Society Colloquium Publication, Vol. 9, 1927.

4. Booth, T. L., *Sequential Machines and Automata Theory*. New York: Wiley, 1967.

5. Cross, B. K., *A Systems Analysis of the Open Pit Mine Design Problem*. Master's Thesis. University of Arizona, Tucson, 1965.

6. Doob, D. L., *Stochastic Processes*. New York: Wiley, 1953.

7. Emshoff, J. R., and R. L. Sisson, *Design and Use of Computer Simulation Models*. New York: Macmillan, 1970.

8. Gupta, S. C., *Transform and State Variable Methods in Linear Systems*. New York: Wiley, 1966.

9. Hall, A. D., *A Methodology for Systems Engineering*. Princeton, N.J.: Van Nostrand, 1962.

10. Hartmanis, J., and R. E. Stearns, *Algebraic Structure Theory of Sequential Machines*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

11. Hille, E., and R. S. Phillips, *Functional Analysis and Semigroups*. American Mathematical Society, Providence, R.I., 1957.

12. Hunt, B. R., "Analysis and Synthesis of Networks of Finite State Machines." *Journal of Systems Engineering*, Vol. 1, No. 1, pp. 9–45, January 1961.

13. Kestenbaum, R., "Systems Engineering." *Electronics*, Vol. 42, No. 9, pp. 68–75. April 1969.

14. Klir, G. J., *An Approach to General Systems Theory*. New York: Van Nostrand Reinhold, 1969.

15. Krohn, K., and J. Rhodes, "Algebraic Theory of Machines. I. Prime Decomposition Theorem for Finite Semigroups and Machines." *Transactions of the American Mathematical Society*, Vol. 116, No. 9, pp. 450–464, April 1965.

16. Lee, Y. W., *Statistical Theory of Communication*. New York: Wiley, 1960.

17. Loomis, L. H., *An Introduction to Abstract Harmonic Analysis*. Princeton, N.J.: Van Nostrand, 1953.

18. McCulloch, W. S., and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115–133, 1943.

19. Naylor, T. H., J. L. Balintfy, D. S. Burdick, and K. Chu, *Computer Simulation Techniques*. New York: Wiley, 1966.

20. Nemytskii, V. V., and V. V. Stepanov, *Qualitative Theory of Differential Equations*. Princeton, N.J.: Princeton University Press, 1960.

21. Norling, R. A., *A Systems Model for Stream Polution Management*. Master's Thesis. University of Arizona, Tucson, 1965.

22. Oren, T. I., *GEST: General System Theory Implementor, A Combined Digital Simulation Language*. Doctoral Dissertation. University of Arizona, Tucson, 1971.

23. Pollin, J. M., and J. L. Sanders, "Theoretical Foundations for Teleological Systems." *Journal of Systems Engineering*, Vol. 1, No. 1, pp. 57–89, January 1969.

24. Pontryagin, L., *Topological Groups*. Princeton, N.J.: Princeton University Press, 1964.

25. Pontryagin, L. S., et al., *The Mathematical Theory of Optimal Processes*. New York: Wiley, 1962.

26. Shannon, C. E., and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1959.

27. **Talavage, J. R.,** "Design of Adaptive Pattern Recognition Systems on the Basis of General System Theory." *Journal of Systems Engineering*, Vol. 1, No. 3, pp. 397–424, July 1969.

28. **Turing, A. M.,** "On Computable Numbers with an Application to the Entscheidungs Problem," *Proceedings of the London Mathematical Society*, Ser. 2, Vol. 42, pp. 230–265, 1936.

29. **Wadsworth, R. B.,** *A Mathematical Model for Instructional Automata.* Doctoral Dissertation. University of Arizona, Tucson, 1969.

30. **Wagner, H. M.,** *Principles of Operations Research.* Englewood Cliffs, N.J.: Prentice-Hall, 1969.

31. **Wagner, P.,** *Programming Languages, Information Structure and Machine Organization.* New York: McGraw-Hill, 1968.

32. **Wiener, N.,** *Cybernetics.* New York: Wiley, 1961.

33. **Wymore, A. W.,** *A Mathematical Theory of Systems Engineering: The Elements.* New York: Wiley, 1967.

34. **Wymore, A. W.,** "A Model Curriculum in Systems Engineering." *Journal of Systems Engineering*, Vol. 1, No. 2, pp. 291–324, April 1969.

35. **Wymore, A. W.,** "Discrete Systems and Continuous Systems." In *Advances in Mathematical System Theory*, P. C. Hammer (Editor), University Park, Pa.: Pennsylvania State University Press, 1969.

36. **Yakowitz, S. J.,** *Mathematics of Adaptive Control Processes.* New York: American Elsevier, 1969.

37. **Zadeh, L. A.** and **C. A. Desoer,** *Linear System Theory.* New York: McGraw-Hill, 1963.

38. **Goodkin, H.,** *Applications of General System Theory to the Scheduling of Oceanic Transports.* Master's Thesis. University of Arizona, Tucson, 1971.

Part IV

Various Aspects of Formal Systems Theories

10. Topological Concepts in the Mathematical Theory of General Systems

JOSEPH V. CORNACCHIO

School of Advanced Technology, State University of New York at Binghamton

EDITOR'S COMMENTS

This chapter, which provides a link between Chapter 9 and Chapter 12, is most profitably read if the reader has some familiarity with elementary concepts and principles from algebra and topology. Suitable references for this purpose are:

Stoll, R. R., *Set Theory and Logic*. San Francisco: W. H. Freeman, 1961.
Hall, D. W., and G. L. Spencer, *Elementary Topology*. New York: Wiley, 1955.
Mansfield, M., *Introduction to Topology*. Princeton, N.J.: Van Nostrand, 1963.

Readers who like solved problems may prefer to learn topology from the books:

Lipschutz, S., *General Topology*. New York: Schaum, 1965.
Greever, J., *Theory and Examples of Point-Set Topology*. Belmont, Calif.: Brooks/Cole, 1967.
Reference [14] is recommended for a deeper study of topology.

Section 10.2 contains brief outlines of specific examples of systems from a wide range of disciplines, and it is suggested that the reader consult the sources listed in the references to this chapter for more detailed discussion and insight. The notation used in Section 10.2 is consistent with the respective sources. For a current and more complete discussion of the material in Sections 10.4 and 10.5, the reader should consult Chapters 8, 9, and 12 of this volume.

10.1. INTRODUCTION

The role of topological concepts in abstract mathematical models of general systems has received little attention in the literature to date, considering the importance of such concepts in the representation of many basic systems properties. Mathematical theories of general systems proposed so far (see Section 10.4) contain essentially only the set-theoretic structure of the underlying specific systems as part of the general theory. Thus, when such a general system model is restricted to subsume a class of specific systems containing topological structure as a fundamental property, it becomes necessary to introduce *additional* structure into the restricted model to adequately represent the specific class at hand. Such a theory lacks the degree of completeness at the abstract general level required of a *general systems* theory. In particular, the theory omits a construct—namely, topological structure—which represents an "essential trait" [3] of the class of specific systems which are presumably to be subsumed by the general system model. These observations may be compared with the aims of the extensive efforts being devoted to general system research [17]:

1. To investigate the isomorphy of concepts, laws, and models in various fields and to help in useful transfers from one field to another;
2. To encourage the development of adequate theoretical models in the fields which lack them.

To motivate the need for topological structure in formal mathematical models of general systems, we pursue an inductive approach by giving examples of classes of specific systems in which topological concepts play a fundamental role in the system description. The topological structure will be seen to be that of the "classical" topological space. The motivation is then pursued by determining a natural topological structure for the set of input functions for a general system model due to Wymore [12]. This is

achieved by a formal deductive approach and yields the interesting result that the natural structure is not the classical topological space [14] but a generalized closure space [10, 13], which we have denoted the *H*-Appert space.

The role of such generalized closure spaces in mathematical theories of general systems must yet be explored. At this time, their principal applications appear to lie in the finitary systems of mathematics [10]. Applications in system-theoretic models of "real-world" phenomena have to be established.

10.2. TOPOLOGICAL STRUCTURES AND CLASSES OF SPECIFIC SYSTEMS

In this section we provide examples of specific systems models in which the use of topological concepts is basic to the representation of fundamental system properties. By representing a broad class of such specific systems, we attempt to motivate the need for topological structure as part of the abstract framework of any mathematical theory of general systems. Thus the motivation is along inductive lines [3, p. 274]. The viewpoint taken is that a mathematical theory of general systems which purports to subsume the characteristics of *specific* (particular) systems should contain not only the basic structure and mappings characterizing the classes of specific systems, but also the topological properties which are a fundamental "trait" of these systems [3].

We will establish, by giving examples, that there is a broad range of classes of specific systems in which topological concepts play a fundamental role in the system description and that such concepts are basic to the consideration of approximation, continuity, optimization, connectivity, etc., in these systems. The examples are taken from the natural sciences (physics), the social and behavioral sciences (anthropology, economics, social psychology), engineering (dynamical physical systems, control theory, functional systems), and computer science (automata theory). These systems are seen to utilize topological concepts in the "classical" sense [10], that is, topological spaces [14] are the natural structure involved. The spectrum of disciplines is broad, and the level of mathematical development of the topological concepts in the systems cited ranges from "minimal" (application of concepts at verbal or descriptive level) to "formally complete" (topological structure is part of the system model, and theorems describing basic system properties exist).

Our review of the application of topological methods in classes of specific systems begins with those in the field of engineering. Although the literature of applications in this field is very extensive, we will discuss works which either are of a fundamental nature or are representative of efforts in general.

10.2.1. Engineering

Dynamical physical systems and control theory. In the model of dynamical systems based on the concept of a state space [18] and used as the basis for studying stability properties of linear dynamical physical systems [19], a dynamical system is defined [18] as a mathematical structure satisfying the following axioms:

(a) There is given a state space, Σ, and a set of values of time θ at which the behavior of the system is defined; Σ is a *topological space.*

(b) There is given a set, Ω, of functions of time, defined on θ, which are the admissible inputs to the system; Ω is a *topological space.*

(c) Every output of the system is a function (R is the real line)

$$\psi: \theta \times \Sigma \to R.$$

(d) The future states of the system are determined by a state-transition function:

$$\varphi: \Omega \times \theta \times \theta \times \Sigma \to \Sigma.$$

(e) The functions ψ and φ are *continuous* with respect to the *topologies* defined for Σ, θ, and Ω and the induced *product topologies.*

Topological considerations which arise for the subclass of this class of specific systems corresponding to linear real, finite-dimensional, continuous time systems [19, p. 155] include the preservation of *stability properties*, which requires, as a necessary condition, *topological equivalence.* In these cases, the topologies are those induced by the Euclidean norm on the finite-dimensional spaces and the resultant topologies induced on the product spaces. Also of fundamental importance, for the case at hand, is the concept of system identification, which requires the existence of certain continuous matrices as necessary and sufficient conditions for the realization of systems governed by differential equations of the type [19]

$$\frac{dx}{dt} = F(t)x + G(t)u(t),$$

$$y(t) = H(t)x(t),$$

defined for $t \in (-\infty, \infty)$, where x, y, u are vector functions of time t, and F, G, H are *continuous* matrix functions of time. These equations are called the *dynamical equations* of the system, and form the basis for much of the work in the extensive research into control theory.*

* For additional material, including recent references and generalizations, see [11].

For a detailed study of linear time-varying systems, including the role of the state space, and topological considerations involving connectedness and continuity, see [20].

A fundamental concept in dynamical physical systems and one which has also been the object of significant study centers around the notion of *approximation* and its derivatives: *optimality* or *optimization*. From an intuitive viewpoint, these concepts relate to the question of determining the "most desirable" among a set of possible alternatives in the solution of a problem. From a mathematical point of view, we may regard this area as the determination of the minimal elements of a partially ordered set [26]. Inherent in the concept of optimization in the context of dynamical systems is that of *approximation*, which in turn, at the fundamental level, can be rigorously described by the constructs utilized in the theory of topological spaces. The description involves the imposition of additional structure on maps between locally convex real *topological* vector spaces [26, p. 490 ff.].

As a final example in dynamical systems and control theory, we refer to the work of Halkin [21], who has presented a generalization yielding a unified axiomatic and *topological* analysis of control problems for a very large class of systems, including the systems described by differential and difference equations. In this work a dynamical system is a pair

$$((Y, \mathcal{Y}), R),$$

where $(Y, \mathcal{Y})$ is a *topological space*, and (Y, R) is a reflexive forward-ordered set having certain additional properties, defined in terms of the subsets of the topology Y. The set Y, called the *event* space, has elements called *events*. On the basis of these fundamental concepts, Halkin proceeds to develop "propositions" involving the determination of *optimal* solutions and *optimal* trajectories for "dynamical polysystems" [21, p. 5].

It is important to point out that the rigorous and quite general mathematical results described here and elsewhere in this section are possible for this specific class of systems precisely because of its *specific* nature, namely, the systems are well defined mathematically and have assumed structure to begin with, for example, finite-dimensional, real, linear, etc. The magnitude of the work required to bring into a mathematical theory of general systems a topological structure which subsumes the material discussed here and yet represents other "classes of specific systems" in a nontrivial manner begins to assume large proportions.

Functional systems. Turning now to another subclass of engineering applications, we consider the work by Brilliant on *functional systems* [7]. Since we will be discussing this topic in some detail in Section 10.3, our description here will be even less detailed than the format followed so far. The work on

functional systems described in [7] is particularly relevant to the objectives of this study since, in addition to the formal introduction of a topological space, Brilliant has also motivated and explicitly determined the relevant open sets which define the topology [14]. Thus his work serves as an instructive example of the manner in which such structure may be introduced into a given set-theoretic model of a specific class of systems to rigorously represent intuitive notions of *approximation* and *continuity*.

In particular, approximation and continuity for such functional systems are accounted for by introducing a *topology* on the set of input functions to the system, and defining either concept in terms of the *neighborhood* (or base) structure of the topology. Using such a derived structure, Brilliant is able to rigorously state and prove results concerning the *continuity* of *linear time-invariant* systems [7, p. 12] and the conditions sufficient for the *approximation* of *continuous* (time-invariant) systems by a class of *polynomial systems* [7, pp. 19–26]. We will return to this work in greater detail in Section 10.3.

10.2.2. Computer Science

The systems of interest in this field are the discrete or countable systems which serve as models for digital computation.

In the work by Arbib [23, 27], the need for topological structure appears even in this case, where the discrete nature of the systems might suggest that such structure would have ancillary relevance. In particular, with reference to the need for representing continuity, Arbib suggests that the problem may be posed as that of determining a relevant topology for a discrete set which is not the discrete topology [14], that is, a definition of continuity in a form such that it is meaningful for finite automata [23, p. 179].

Following this motivation, Arbib has introduced the concept of a *tolerance*, and, from that, the notion of *continuity* of functions mapping one *tolerance space* into another. The pertinent theoretical structure is as follows. Given a set X, a relation ξ on X is a *tolerance* if ξ is reflexive and symmetric. A *tolerance space* is the pair (X, ξ). As an example, let X be the Euclidean plane and ξ be all pairs of points less than ε apart; or let X be the visual field and let ξ be the visual acuity tolerance, that is, all pairs of points that are indistinguishable.

The relationship of this concept of tolerance to automata theory is made more specific by introducing additional structure as follows. Suppose that $T = \{0, 1, 2, \ldots\}$ and that (X, ξ) is a tolerance space. Then a *motion* in X is a function

$$m\colon T \to X.$$

The motion is ξ-*continuous* if $(m(t), m(t + 1)) \in \xi$ for all $t \in T$, that is, there are no "detectable jumps in the motion." These concepts allow for the introduction into the formal system model of a discrete automaton an "intuitively acceptable idea of continuity" [23, p. 180].

Extensions of the notion of tolerance are then made to include the concept of ξ-*continuous* functions between tolerance spaces by defining a function

$$f: X \to Y,$$

where (X, ξ_x), (Y, ξ_y) are tolerance spaces to be ξ-continuous if

$$(x_1, x_2) \in \xi_x \Rightarrow (f(x_1), f(x_2)) \in \xi_y.$$

With this much structure at hand, it becomes possible to consider the *stability* of such discrete systems defined in terms of states of the automaton. Subclasses of automatons—"1-tolerance" automata—are shown to possess a *stability* property not possessed by a larger but similar class ("n-tolerance automata"), namely, that small differences in initial state cannot give rise to large differences in state at later times [23, p. 181]. Having introduced a topological machinery and the related approximation mechanism, it becomes possible to rigorously consider the introduction of *optimization* into the system and ultimately to state precisely the *optimal control problem* for automata [23, p. 183]. As a final step in the assimilation of topological concepts in this discrete system, the derived topological concepts of closure, interior, and boundary are defined in terms of the tolerance relation ξ. For example, if $S \subset X$ is a subset in which X is a tolerance space, then the ξ-closure of S is the set

$$\bar{S} = \{x \,|\, (x, y) \in \xi, \quad \text{for some } y \in S\},$$

etc. By introducing this structure into the set-theoretic model of an automaton, Arbib is able to state and prove a result which is the analog for automata theory of the Pontryagin maximum principle of optimal control [23, p. 184].

Another example of the role of topological concepts in automata systems where countable sets are involved is the study of the relationships between continuity and realizability of sequence transformations [25]. The fundamental set defining the system of interest is the Cartesian product

$$X_{n+1} = X_n \times X_1, \quad n = 1, 2, \ldots,$$

where

$$X_1 = \{0, 1\}.$$

Elements of X_n, that is, n-tuples of 0's and 1's, are called *blocks of length n*, with the set of *infinite right sequences* of 0's and 1's denoted by X, that is,

$$X \subset X^* = \bigcup_{k=0}^{\infty} X_k$$

is the set of mappings from the positive integers, I^+, into $\{0, 1\}$. Thus $x \in X$ implies that x can be written

$$x = x(1)x(2)x(3)\ldots, \quad x(i) \in I_1, i \in I^+,$$

or in the usual fashion

$$x = x_1 x_2 x_3 \cdots$$

with

$$x_i = x(i), \quad i \in I^+.$$

With this notation, the equality of elements in X, say $x, y \in X$, is defined by

$$x = y \quad \text{if and only if } x(i) = y(i), \quad i \in I^+.$$

By utilizing the set-theoretic structure available to this point, the entrance of topological structure is achieved by introducing a *metric* on X, that is, for $x, y \in X$, the *distance* between x and y is given by

$$d(x, y) = \begin{cases} 0, & \text{if } x = y, \\ 1/k, \text{ where } k = \min\{i \in I^+ \,|\, x(i) \neq y(i)\}, & \text{if } x \neq y. \end{cases}$$

Again the approach here is similar to that of Arbib as described above, whereby the topological structure is introduced via an auxiliary construct, in this case a metric which induces a topology on X. In particular, (X, d) is a topological space [14].

The principal property of interest in this work is *compactness*, which is utilized strongly in determining conditions sufficient to *realize* a specific class of *sequence transformations* in terms of the class of logic nets having no feedback [25, p. 561]. The mathematical results are as follows. Since (X, d) is a *topological space*, one may consider the *continuity* of sequence transformations on X, that is, functions

$$\varphi: X \to X.$$

Hellerman et al. show that, in the topology induced by the metric d, if φ is any sequence transformation which is *continuous* at $x \in X$, then, since X is *compact*, φ is *uniformly continuous* on X. This, in turn, brings in the fundamentally associated concept of *approximation* in terms of the metric d, by assuring that a measure of "closeness," specified by

$$d(\varphi(x), \varphi(x^1)) < \frac{1}{n}, \quad n \in I^+,$$

if some x, $x^1 \in X$ holds for some $m \in I^+$ if

$$d(x, x^1) < \frac{1}{m}.$$

The principal result of interest in the work discussed here, which illustrates how a topological structure may yield (by rigorous methods) systems properties of profound consequences, is that sufficient conditions for the *realization* of a sequence transformation in terms of logic nets (or " modified " finite automata [25, p. 563]) include the property that the sequence transformation be *continuous*. The point being made here, again, is that the topological structure is fundamental to the basic system concepts of *continuity* and *approximation*; in this case, such structure also yields a system *synthesis* criterion—a fundamental consideration in any system theory.

10.2.3. Social and Behavioral Sciences

Anthropology. Kinship studies in anthropological systems have been cast into mathematical models in which the basic structuring is accomplished via topological spaces [28]. For example, the question of whether various *separations* of kinsmen by semantic features corresponds to separations in the structure of the semantic space can be investigated by first defining a *topology* on a set of kin terms [28, p. 71] (using geneological relations), and then comparing the properties of the topology to the properties derived from the semantic analysis [28, pp. 73 ff.]. The specific property of the topology is that of the *connectedness* of the associated set and the resultant system of *components*. Another approach toward the analysis and modeling of kinship systems involves the introduction of a *topology*, T, on the Cartesian product X^2 of the set X of members of a given tribe [28, p. 49], that is, the pair (X^2, T) is considered as a *topological space* representing a known kinship relation. These approaches are of quite recent origin and are the subject of active study by mathematically inclined anthropologists [28].

The increasing use of topological structures in anthropology, and the social sciences in general, is suggested by Leach [29], who points out [29, p. 7] that, since topology is a nonmetrical form of mathematics, it deserves special attention from social scientists. In particular, he asserts that anthropological kinship relationships should be considered as constituting a *neighborhood system*, that is, a topological space [29, p. 7]. Again we see that a fundamental topological concept of interest from the anthropologist's viewpoint is *connectedness*, with the implication that such concepts are relevant to "regularities of pattern among neighboring relationships" [29, p. 7]. Thus the interest in patterns of relationships generates a need for topological structure.

Economics. The role of topological structures here is of course firmly established, especially if one considers the heavy emphasis in contemporary economics on optimization. In such cases, the topological structure is not usually singled out in the system description since the relevant sets are the real line or Cartesian products of subsets of the real line, where the topology is the *usual topology* on the real line or the induced *product topology.* It is useful, however, to point out a fundamental work, that of Debreu, related to the nonmetrical* concept of utility [1]. The problem considered in this work considers a result generated by the difficulties encountered in testing the von Neumann-Morgenstern axioms for the existence of cardinal utility in some specific situations [1, pp. 16-17]. The topological statement of the problem reduces, in the specific case denoted by the term "stochastic objects of choice" [1, p. 17], to the determination of *topological conditions—connectivity, separability,* and *closedness*—for a set S ("commodity bundles") such that a utility function can be defined on S. Specifically, given a set S and a preordering relation $\lesssim$ on $S \times S$, a *utility function* is a real-valued, order preserving function u on $S \times S$ such that

$$u(a, b) = \tfrac{1}{2}[u(a, a) + u(b, b)]$$

for every a and $b \in S$.

The problem is to find conditions on S and $\lesssim$ which will guarantee the existence of a utility function defined as above. Debreu makes the following "assumptions" [1, p. 18]

(a) S is connected and separable.
(b) $\lesssim$ is a complete preordering of $S \times S$ such that $S \times S$ and

$$\{(a, b) \in S \times S \,|\, (a, b) \gtrsim (a', b')\}$$

and

$$\{(a, b) \in S \times S \,|\, (a, b) \lesssim (a', b')\}$$

are *closed* for every $(a', b') \in S \times S$.
(c) $[(a_1, b_2) \lesssim (a_2, b_1)$ and $(a_2, b_3) \lesssim (a_3, b_2)] \Rightarrow [(b_3, a_1) \lesssim (b_1, a_3)]$.

Using these assumptions, Debreu establishes that there is a *continuous* utility function determined up to an increasing linear transformation [1, pp. 18–19].

The need for topological structure to account for optimization, approximation, stability, and optimal control is well documented in the literature of mathematical economics [30]. In the area of optimal resource allocation

---

* We can simply identify the concerns suggested by this adjective to be the problems in which the real line and its usual Euclidean metric are not necessarily the basic structure involved.

models, the extensive use of topological structures, namely linear topological spaces, is exhibited in the article by Hurwicz, entitled "Programming in Linear Spaces" [31, pp. 4–102].

Social psychology. The study of perception and the concept of structure [6] involves the construction of models wherein the topological concepts of connectedness, boundary, and continuity have made an intuitive entry. Some particular applications of these concepts appear in the gestalt theory of perception [6], where the correspondence between stimulus patterns and the associated brain patterns is viewed as being *topological* in the sense that the correspondence preserves relationships of "betweennesses" and "adjacences" in a perceived figure, rather than exact sizes, angles, or shapes [6, p. 134].

A derivative of this gestalt approach, which eliminates physiological considerations and focuses instead on the phenomenological concept, called the field, is the field theory of behavior [6], due to Lewin [32; 33, p. 113], with special emphasis directed toward motivation and goal-directed action. This concept of the field is such that "its medium is conceived not as metrical but as topological in its spatial character" [6, p. 148; 32]. Within this field-theoretic structure, the concept of *life space* [24, p. 151; 32; 37, p. 407] is introduced in terms of a *boundary line* called the "Jordan curve." This life space is additionally characterized by the principle that there exists within it a region or connected regions in which "locomotion" could occur. Thus the topological concepts of interior, boundary, connectedness, and components are already intuitively present in the conceptual structure.

The importance of the topological aspect of the model to the behavioral applications to which it is directed is expressed by the observation [6] that one of Lewin's greatest contributions was introducing, through topology and the Jordan curve, the concept of the behavioral aggregate as something that is highly flexible with respect to its spatial properties, is unified and self-delimited, and yet possesses many possible degrees of freedom. The topological system of Lewin is considered as one of the factors that made field notions popular among psychologists. Through the dynamism of field concepts, the topological system has helped to make perception and cognition a basis for modern systems of social psychology [6].

10.2.4. Natural Sciences

Physics. The use of topological concepts and methods in contemporary theoretical physics is comprehensive and profound, reflecting the position of the field as the model *par excellence* of a mature mathematical science. Thus we choose here to give an outline of some recent work concerning the concept of *elementary length* as applied in theoretical studies of quantum-mechanical

particle scattering [5]. Preliminary efforts have utilized a "natural topology" to account for effects of an elementary length, where such a length is considered as a fundamental number, λ, such that all length measurements are integer multiples of λ. The physical interpretation is that the distance between two particles cannot be measured more accurately than λ and also that the distance will be a multiple of λ. Although the concept has been of continual theoretical interest since the days of Pythagoras [5], recent interest has peaked because of anomalous results obtained in scattering experiments involving electron-positron pairs, that is, results not explainable by the currently accepted theory of scattering as described by quantum electrodynamics [5, 2]. In an attempt to explain this anomalous situation, the concept of a *topological potential* due to the existence of an elementary length has been proposed [4]. The question of which class of topologies, when placed on the particle coordinate system, is consistent with certain natural requirements of an elementary length has been addressed [5]. It has been demonstrated that there is essentially only one such class, thereby also placing on a firm basis the ad hoc topology initially assumed in earlier attempts to explain the scattering anomaly [5, 4].

The introduction of the elementary length topology is based on the definition of a *length*, and of an *elementary length*, on a normed linear space $(X, \| \cdot \|)$, where $\| \cdot \|$ is the norm, as a *function*

$$f_\lambda : X \to \{n\lambda \,|\, n = 0, 1, 2, \ldots\}$$

satisfying certain requirements which are the formalization of the intuitive properties of a length. These requirements lead to the result that a function f on X is a *length* if and only if it has one of two forms [5]:

(1) $f(x) = \lambda \left[\dfrac{\|x\| - a}{\lambda} \right], \quad \lambda > 0, 0 \le a \le \lambda$

(2) $f(x) = \lambda \left\{ \dfrac{\|x\| - a}{\lambda} \right\}, \quad \lambda > 0, 0 < a \le \lambda$

with $[\mu]$ and $\{\mu\}$, where $\mu \in R$, defined as the smallest integer not less than μ and the smallest integer greater than μ, respectively. The concept of an *elementary length* is formally represented as a function on X of the form

$$f(x) = \lambda \left[\frac{\|x\|}{\lambda} \right] \quad \text{for all } x \in X, \lambda > 0.$$

To obtain the physically meaningful subsets of X which will also serve as a base for the elementary length topology, an *open ball* is defined as a subset of X given by

$$B_{\lambda, a}(x, r) = \{y \in X \,|\, f_\lambda(y - x) < r\}.$$

However, an attempt to utilize the open balls as a base for an elementary length topology [14] leads to the result that the collection of open balls corresponding to physically significant and accessible subsets of X yields the discrete topology on X. The discrete topology, however, is unsatisfactory for describing the physical situation. This contradiction between the properties required for the collection of open balls to be a base for the elementary length topology and that of physical significance leads ultimately to the conclusion that the lengths on X do not generate a satisfactory topology in the physical sense [5]. Thus the basic set on which the topological structure is to be placed must be changed to account for the physical reality that the lengths introduced must represent the distance between *two* particles, a single particle being experimentally unobservable.

A Cartesian product $X \times X$, representing the coordinate spaces for two particles, is therefore introduced. The derived coordinate spaces X_c, X_r, representing the center of mass of the two particles and their relative position, respectively, are then determined. At this stage it is possible for a physically meaningful collection of open balls to serve also as a base for an elementary length topology. The results are summarized by the following [5]:

The balls $B(x, n\lambda) = \{y \mid f_\lambda(y - x) < n\lambda\}$ generate a nondiscrete topology on X_r: the elementary length topology on X_r, where X_r is the set defined by the canonical map $T: X \times X \to X_c \times X_r$ with

$$T(x_1, x_2) = \left(\frac{x_1 + x_2}{2}, x_1 - x_2\right), \quad x_1, x_2 \in X.$$

Endowing $X \times X$ with the *weakest topology* which makes T continuous yields the elementary length topology on the two-particle space. A base for this topology has elements of the form

$$\{(x, y) \in X \times X \mid |x + y - z_1| < a, f_\lambda(x - y, z_2) < b, z_1, z_2 \in X, a, b \in R\}.$$

The topology so generated is observed to be that introduced in an "ad hoc" fashion in the original attempt to describe the scattering anomaly in terms of a topological potential [4].

Thus the structure of a topological space is seen, in this instance, to represent a fundamental physical property in this area of contemporary theoretical physics.

10.3. TOPOLOGICAL SPACES IN FUNCTIONAL SYSTEMS

An example of the introduction and use of topological spaces in a class of specific systems is the work [7] on *time-invariant*, *physically realizable*, *functional* systems. Such systems are characterized as follows. The input to the

system is a function $f: R \to R$ such that $f(t)$, $t \in R$, is the value of the input at time t. The output of the system at time t depends on the input for only those values of time up to t, that is, for values of $f(\alpha)$, with $\alpha \leq t$. The time-invariance property of the system is accounted for by assuming that the input history up to time t can be represented by a function u_t:

$$u_t : R^{++} \to R,$$

defined for every $t \in R$ and a given input f by

$$u_t(\tau) = f(t - \tau), \quad \tau \geq 0.$$

The assumption of physical realizability is accounted for by restricting τ to be nonnegative. The output of the system at time t is assumed to be a *functional* [22] of the input represented by u_t (corresponding to a given input function, f) and is expressed by

$$g(t) = h(u_t),$$

where $g(t)$ is the output at time t, $g(t) \in R$. It is assumed, furthermore, that the inputs to the system are *bounded* [7]: an input function is bounded by $A \in R^+$ if, for all $t \in R$, $|f(t)| < A$. Thus it follows that, if f is bounded by A, then, for any t, u_t (as determined by f) is also bounded by A. The formal representation of the class of inputs to the system is given by the set $\mathrm{PBI}(A) \subset R^{R^{++}}$, given by $\mathrm{PBI}(A) = \{u \mid u: R^{++} \to R, |u(\tau)| \leq A, t \in R, u \text{ is Lebesque measurable}\}$.* The Lebesque measurability, of u is assumed to eliminate, as possible inputs, functions which have no physical interpretation in the context of the application of the theory being developed here [7]. The interpretation of the set $\mathrm{PBI}(A)$ is that, for a given input function $f: R \to R$ and $t \in R$, such that the corresponding $u_t \in \mathrm{PBI}(A)$, $h(u_t)$ is the output of the functional system h at time t. We will now restrict our attention to the collection of functions $\mathrm{PBI}(A)$, considered as the set of input functions for the systems of interest.

The topological considerations which arise within this class of systems stem from the desire to address the following question [7]: Given the set of input functions, $\mathrm{PBI}(A)$, what structure defined over this set will account for the intuitive notion of approximation in the following cases: (1) in the sense of generating an approximating *table of values* [7], that is, the approximation of all input-output pairs for the system by a finite collection of such pairs; and (2) in the sense of approximation of an arbitrary member of a subclass of functional systems by a member of a "simpler" subclass of functional systems within the class? In both cases, the definition and representation of

---

* PBI represents "Past of Bounded Input" [7].

a *continuous* system play a fundamental role. We outline here the approach followed in [7] to address these questions. On the set PBI(A) a *neighborhood* $N_{T,\delta}(u)$, is defined for every $u \in$ PBI(A) and $T, \delta > 0$ by

$$N_{T,\delta}(u) = \left\{ v \in \text{PBI}(A) \Big| \ \Big| \int_0^x [u(\tau) - v(\tau)] d\tau \Big| < \delta, \quad \text{for } 0 \le x \le T \right\}.$$

On the set R, a neighborhood $N_\varepsilon(x)$ is defined for every $x \in R$ and $\varepsilon > 0$ by

$$N_\varepsilon(x) = \{ y \in R | \ |y - x| < \varepsilon \},$$

that is, $N_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$.

The motivation for and the interpretation of these neighborhoods are contained in [7]; here we may observe that the pair (T, δ) can be considered a *tolerance* in the sense that, given such a pair, all those $v \in N_{T,\delta}(u)$ for a given u are "equal" to u within the tolerance (T, δ), where "equality within the tolerance (T, δ)" is defined by $v \in N_{T,\delta}(u)$ for any $u, v \in$ PBI(A). We may also interpret $v \in N_{T,\delta}(u)$ as v being an approximation (to within T, δ) to u. Notice that the definition of $N_{T,\delta}(u)$ implies that, if $v \in N_{T,\delta}(u)$, then $u \in N_{T,\delta}(v)$, that is, if v is an approximation to u, then u is an approximation to v. Similar interpretations concerning system output may be made for the neighborhoods $N_\varepsilon(x)$. For the collections of neighborhoods

$$\mathscr{B} = \{ N_{T,\delta}(u) | T, \delta > 0, u \in \text{PBI}(A) \}$$

and

$$\mathscr{N} = \{ N_\varepsilon(x) | \varepsilon > 0, x \in R \},$$

a functional h on PBI(A) is defined* as *continuous* if, for any $\varepsilon > 0$, there exists a pair (T, δ), $T, \delta > 0$ such that

$$h(N_{T,\delta}(u)) \subset N_\varepsilon(h(u))$$

for any $u \in$ PBI(A), where

$$h(N_{T,\delta}(u)) = \{ h(v) | v \in N_{T,\delta}(u) \}.$$

Any time-invariant, physically realizable system represented by the functional h is said to be *continuous* (A) [7]. These considerations are represented in the diagram of Figure 10.1.

The introduction of the collections $\mathscr{B}$ and $\mathscr{N}$ and the definition of a continuous system now allow for the consideration by formal methods of the questions posed earlier. The principal results are as follows [7]:

---

* The definition of continuity here is given in terms of the base of a topology [14].

Input set   Output set
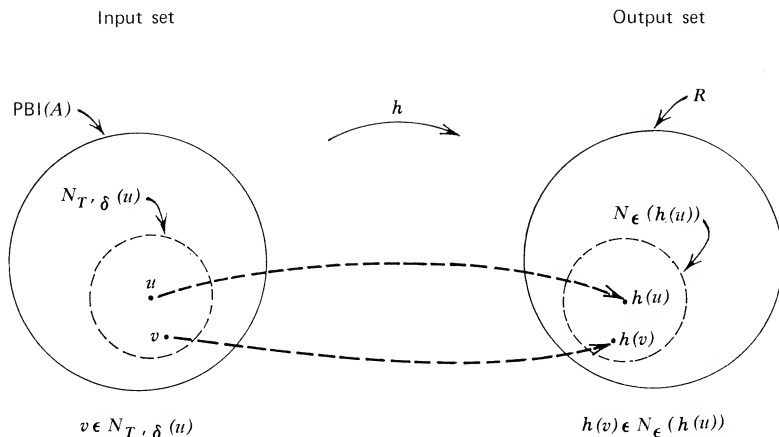
Figure 10.1. Illustration of the neighborhood structure for a functional system h.

1. The collection $\mathscr{B}$ forms a *base* [14] for a topology: the RTI (" Recent Time Integral ") topology.

2. The pair $(\mathrm{PBI}(A), \mathrm{RTI})$ is thus a topological space, which is also compact [4].

3. Since $(\mathrm{PBI}(A), \mathrm{RTI})$ is compact, the input-output behavior of a continuous (A) system h can be represented in the form of a " table of values." This is achieved by first giving a tolerance ε in R^+ and then choosing a finite set of inputs, say $u_i \in \mathrm{PBI}(A)$, $i = 1, \ldots, n$, with a corresponding set of outputs $h(u_i)$ in R. The output, $h(u)$, for any $u \in \mathrm{PBI}(A)$ can be approximated to within ε by the output $h(u_i)$ for some $u_i \in \mathrm{PBI}(A)$ such that $u \in N_{T,\delta}(u_i)$. (That there exists such a finite collection of $N_{T,\delta}(u_i)$, where $h(u) \in h_\varepsilon(h(u_i))$—corresponding to the finite set of u_i—which covers $\mathrm{PBI}(A)$, is assured by the fact that $\mathrm{PBI}(A)$ is compact in the RTI topology [14].)

4. There exists a subclass of continuous functionals defined on $\mathrm{PBI}(R)$ and called *polynomial systems* [7], such that the output of any continuous system can be approximated arbitrarily closely in terms of polynomial systems by using the approximation criteria described above.

Although we have outlined only the principal considerations involved in the work on functional systems, the role of topological spaces in the representation of systems-theoretic concepts of approximation and continuity is evident.

10.4. MATHEMATICAL MODELS OF GENERAL SYSTEMS

In this section, we present a brief description of abstract models of general systems as proposed by Mesarovic [9] and Wymore [12]. The purpose is twofold: to illustrate the principal set-theoretic structure characterizing these two axiomatic approaches to general systems models, and to provide background for Section 10.5.

In the Mesarovic approach [9], a *general system*, S, at the highest level of generality, is defined as a relation on a given collection of sets, $V = \{V_i \,|\, i \in I\}$, called *objects* of the system (I is an indexing set for the collection), that is,

$$S \subset \bigtimes V.$$

Further structure is introduced by assuming that the collection V is partitioned into two subcollections such that

$$S \subset X \times Y,$$

where

$$X = \bigtimes \{V_i \,|\, i \in I_x\},$$
$$Y = \bigtimes \{V_i \,|\, i \in I_y\},$$

and $\{I_x, I_y\}$ is a partition of I. In this case, X and Y are called the *input* and *output* objects, respectively, of the system S. Within this *binary relation representation* of S, the concept of *state object* is developed by introducing a (partial) algebraic structure in S. A *state space* for such a representation is then developed by specifying an equivalence relation on the union of all state objects of the system.

A further specification of the binary relation representation is obtained by considering *complete time systems* in which the input and output objects, X and Y, are collections of functions

$$X = A^T, \qquad Y = B^T,$$

where T is a linearly ordered set (the *time set*), A and B are sets called input and output *alphabets* (or spaces), and S is defined by

$$S \subset A^T \times B^T.$$

The principal considerations in such complete time systems involve the concept of the state, introduced in the general structure above and restricted to this more specific class of system. In particular, such considerations include [9], for example, the existence of *initial state objects*, and *initial state representation* for complete time systems, definitions of *state-transition relations*

and *functions*, determination of the state space for time systems, and conditions for time systems to be *state determined*, *nonanticipatory*, and so on.

The conceptual approach underlying the mechanism for generating specific classes of systems included in the general relation-theoretic structure (i.e., a mechanism for the classification of systems) is that of *constructive specification* of a system. In this approach, the specific properties which define S and provide its description, or structure, are derived from *auxiliary functions* or simpler systems. The specification proceeds by some particular method, for example, by induction in the present work.

The role of topological concepts in this theory has not yet been explored, although reference has been made [9] to the introduction of topological structure in the general system model as a means of further specification of the system.

In the axiomatic approach to a mathematical theory of general systems due to Wymore [12], a general system, Z, is defined as a set

$$Z = \{S, P, F, M, T, \sigma\},$$

where S and P are nonempty sets, called the *state* and *input state* sets, respectively; F is a subset of P^R with certain algebraic closure properties described in the next section* and is called the set of *input functions* of Z; $M \subset S^S$ is the set of *transition functions* of Z and contains the identity function, ω, on S; $T \subset R$ is the *time scale*; and $\sigma: F \times T \to M$ (onto) is the system *state-transition function*. For a given input function $f \in F$, a time $t \in T$, and a state $x \in S$, $\sigma(f, t)(x) \in S$ is interpreted as the state of the system at time t, given the initial state x and input function f. In addition, requirements are imposed on σ which are interpreted to represent the intuitive properties possessed by "real-world" systems, namely,

(i) ("initial state consistency"): For every $f \in F$

$$\sigma(f, 0) = \omega \in M;$$

(ii) ("composition property"): If $t_1, t_2 \in T$ such that $t_1 + t_2 \in T$, then for every input function,† $f \in F$,

$$\sigma(f, t_1 + t_2) = \sigma(f \to t_1, t_2) \circ \sigma(f, t_1);$$

(iii) ("causality"): If $\tau \in T$, $\tau \geq 0$, and $f, g \in F$ such that $f \,|\, [0, \tau) = g \,|\, [0, \tau)$, then

---

* The admissible set as defined in [12] has been generalized in Section 10.5 to allow for the possibility that $F = \phi$; this has the effect of modifying the topology generated by the admissible set operator $\mathscr{G}$ and is also discussed in the following section.

† For $f \in F$, $\tau \in T$, $f \to \tau : R \to P$ is defined by [12, Chapter 2] $(f \to \tau)(t) = f(t + \tau)$. $f \to \tau$, is the translation of f by τ and is discussed in more detail in Section 10.5.

$$\sigma(f, \tau) = \sigma(g, \tau);$$

similarly, for $\tau < 0$ and $f \,|\, [\tau, 0) = g \,|\, [\tau, 0)$,

$$\sigma(f, \tau) = \sigma(g, \tau).$$

The general system model does not require an "output" as part of its specification [12, Chapter 2], although the existence of an output or output functions is accounted for by defining any nonempty set Q as an *output* set and any function

$$\zeta \colon S \to Q$$

as an *output function* for Z with values in Q. Intuitively, the system Z can be represented as in Figure 10.2.

The evolution of the system is formally described by various "trajectory" functions [12, Chapter 2]: the *time trajectory* of Z,

$$e_{f, x} \colon T \to S,$$

defined for each $x \in S$ and $f \in F$ by

$$e_{f, x}(t) = \sigma(f, t)(x), \quad \text{for all } t \in T;$$
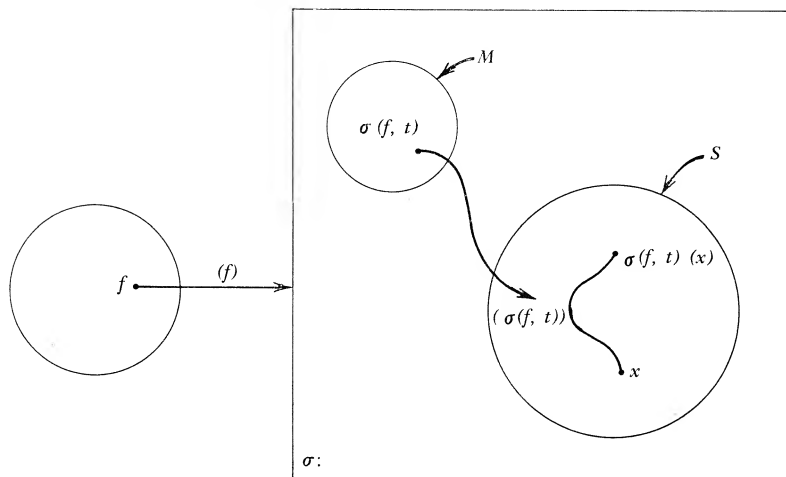
Figure 10.2. Intuitive representation of the general system model due to Wymore [12]: F is the set of input functions, with $f \in F$ a specific input function; S is the set of system states with $x \in S$ the initial state; σ is the system state transition function; $\sigma(f, t)$ is the transition function defined by $(f, t) \in F \times T$ and is shown as a member of M—the set of transition functions of Z; and, finally, $\sigma(f, t)(x) \in S$ is the state of the system at time t, illustrated as the mapping which carries the initial state x into the state at time t, $\sigma(f, t)(x)$.

the *input trajectory* of Z,

$$i_{x,t} : F \to S,$$

defined for each $t \in T$ and $x \in S$ by

$$i_{x,t}(f) = \sigma(f, t)(x) \quad \text{for all } f \in F;$$

and the *output trajectory* of Z, for an output set Q and output function

$$O_{x,f} : T \to Q,$$

defined for each $x \in S, f \in F$, by

$$O_{x,f} = \zeta \circ e_{f,x}.$$

The Wymore model also formally accounts for such concepts as coupling [12, Chapter 5], subsystems and components [12, Chapter 6], and systems classifications considerations [12, Chapter 4], although we will not make use of them here.

In [12] Wymore has also introduced a dichotomy of the class of general systems represented by the model, namely, the classes of discrete and continuous systems. Although we will have more to say in the next section on the topological properties of the Wymore model, it is useful to point out here that the approach we take to this aspect of systems classification is not that of establishing the dichotomy on the nature of the time scale T, as Wymore has done. A particular example which illustrates the distinction is the work by Brilliant discussed in Section 10.3, where consideration of the continuity of "continuous time" systems is established by the nature of the mapping *effected on* the input *functions* by the system, independently of the time scale.

In simple terms, not all continuous time systems are necessarily considered as continuous from this viewpoint.

10.5. CLOSURE SPACES IN A MATHEMATICAL THEORY OF GENERAL SYSTEMS

As stated earlier, our interest in the role of topological concepts in the mathematical theory of general systems has led us to examine the model of a general system due to Wymore [12] in order to determine its intrinsic, or natural, topological structure. This section is concerned with some results of that study, in particular, a relationship between the work of Hammer on extended topology [10, 13] and the general system model of Wymore.

Although a formal systems model can, in an ad hoc manner, incorporate a topological structure by simply postulating that the sets of interest carry a classical topology, that is, that they are to be a priori considered topological

spaces, we have taken the approach that it is of more fundamental value to inquire as to the existence of a *natural topology* determined by the intrinsic algebraic and set-theoretic properties of the structure itself. If such a topology were suggested, subsequent considerations would explore its system-theoretic implications concerning approximation, continuity, and so forth. The pursuit of this approach with respect to the general systems model due to Wymore has yielded some interesting results concerning a topological structure intrinsically determined by the theory [36].

Thus, in this section we wish to establish that:

(i) Within the framework of the elements of a mathematical theory of general systems as developed by Wymore, the method of generating admissible sets of input functions suggests naturally a function, the admissible set operator, which satisfies, in general, *not* the well-known Kuratowski closure axioms, but the closure axioms introduced by Hammer in his work [10, 13] on extended topology.

(ii) As a consequence of (i), the topology suggested by the theory, that is, the topology associated with the admissible set operator, is not, in general, a classical topology, but the extended topology of Hammer.

(iii) Closed subspaces of the Hammer topological space of (ii) represent a topological extension of the set of input functions for a Wymore general system.

10.5.1. Preliminary Definitions and Results

In the following development and throughout the rest of this section* let P be a fixed, but arbitrary, nonempty set unless the contrary is specified; let R be the set of reals; and define P^R by

$$P^R = \{f \,|\, f \colon R \to P\}$$

Definition 10.1 [12]. For $f, g \in P^R$, $r \in R$,

(a) the *translation* of f by r is the function $f \to r \colon R \to P$, defined by

$$f \to r(t) = f(r + t), \quad t \in R;$$

(b) the *segmentation* of f and g is the function $f \,|\, g \colon R \to P$ defined by

$$f \,|\, g(t) = \begin{cases} f(t), & t < 0 \\ g(t), & t \geq 0 \end{cases}$$

for every $t \in R$. The next definition introduces algebraic closure in terms of translation and segmentation.

---

* When not specified otherwise, the notation used in this section conforms with that of [14].

closed subsets of P^R. We begin by immediately defining the admissible set operator.

Definition 10.4. The *admissible set operator* is the function $\mathscr{G}: \mathscr{P}(P^R) \to \mathscr{P}(P^R)$, defined by

$$\mathscr{G}(A) = \cap \{F \mid (F \in \mathscr{A}) \wedge (A \subset F)\}$$

for every $A \in \mathscr{P}(P^R)^*$.

For purposes of brevity, we introduce a neologism by defining a sub-collection of A for each $A \subset P^R$ as follows.

Definition 10.5. For every $A \subset P^R$, the *containing collection of A* is the set $\mathscr{A}_A$, defined by

$$\mathscr{A}_A = \{F \mid F \in \mathscr{A}) \wedge (A \subset F)\}.$$

The preceding definitions establish the following remark.

Corollary 10.1. Let $A \subset P^R$; then $\mathscr{G}(A) = \cap \mathscr{A}_A$. The next result demonstrates that the admissible set operator is a function into A, that is, the image under $\mathscr{G}$ of every $A \subset P^R$ is an admissible set of input functions.

Theorem 10.1. For every $A \in \mathscr{P}(P^R)$, $\mathscr{G}(A) \in \mathscr{A}$.

Proof. Let $A \in \mathscr{P}(P^R)$. Since $P^R \in \mathscr{A}$ (Lemma 10.1), and $A \subset P^R$, we have $P^R \in \mathscr{A}_A$. Therefore $\cap \mathscr{A}_A \subset P^R$ and thus $\mathscr{G}(A) \subset P^R$.

To establish that $\mathscr{G}(A)$ is closed under translation and segmentation, assume that $\mathscr{G}(A) \neq \varnothing$ since, if $\mathscr{G}(A) = \varnothing$, then $\mathscr{G}(A) \in \mathscr{A}$ by Lemma 10.1. Let $f, g \in \mathscr{G}(A) = \cap \mathscr{A}_A$ and $r \in R$. Let $F \in \mathscr{A}_A$; then $f, g \in F$. Since F is closed under translation and segmentation, $f \to r \in F$ and $f \mid g \in F$. Furthermore, since $F \in \mathscr{A}_A$ was arbitrary, we have

$$f \to r, \quad f \mid g \in \cap \mathscr{A}_A = \mathscr{G}(A).$$

Finally, since $f, g \in \mathscr{G}(A)$ and $r \in R$ were arbitrary, $\mathscr{G}(A)$ is closed under translation and segmentation. Thus $\mathscr{G}(A)$ is an admissible set of input functions; therefore $\mathscr{G}(A) \in \mathscr{A}$.

Definition 10.6 [12]. Let $A \subset P^R$; then the admissible set of input functions *generated* by A is the set $\mathscr{G}(A)$.

We now introduce the Kuratowski closure axioms and the definition of a Kuratowski closure operator. We will then show that, for arbitrary non-empty P, the admissible set operator satisfies all the Kuratowski closure

---

* The definition given here is a function-theoretic formalization of a result in [12, Chapter 2], extended to the definition of an admissible set of input functions as used in this work. The name and symbol for the function are due to [12].

axioms except that for distributivity over finite unions—*additivity*—and is therefore, in general, not a Kuratowski closure operator.

Definition 10.7 [5]. Let X be a set and K a function, $K: \mathscr{P}(X) \to \mathscr{P}(X)$; then K is a K-closure operator on X if and only if the following statements, the Kuratowski closure axioms, are true:

  (i) $K(\varnothing) = \varnothing$;
  (ii) $A \subset K(A)$, for every $A \subset X$;
  (iii) $K \circ K(A) = K(A)$, for every $A \subset X$;
  (iv) $K(A \cup B) = K(A) \cup K(B)$, for every $A, B \subset X$.

Theorem 10.2. The admissible set operator, $\mathscr{G}$, satisfies (i), (ii), and (iii) of the Kuratowski closure axioms.

Proof.
  (i) We have that $\mathscr{G}(\varnothing) = \cap \mathscr{A}_\varnothing$. Since $\varnothing \in \mathscr{A}$ (Lemma 10.1), $\varnothing \in \mathscr{A}_\varnothing$, and therefore $\cap \mathscr{A}_\varnothing \subset \varnothing$. Thus*

$$\mathscr{G}(\varnothing) = \varnothing.$$

Let $A \subset P^R$ in (ii) and (iii) below.
  (ii) Let $F \in \mathscr{A}_A$; then $A \subset F$, and therefore

$$A \subset \cap \mathscr{A}_A = \mathscr{G}(A).$$

  (iii) From (ii),

$$\mathscr{G}(A) \subset \mathscr{G}(\mathscr{G}(A)).$$

Furthermore, noting that $\mathscr{G}(A) \in \mathscr{A}$ (Theorem 10.1) and $\mathscr{G}(A) \subset \mathscr{G}(A)$, we have

$$\mathscr{G}(A) \in \mathscr{A}_{\mathscr{G}(A)}.$$

Thus

$$\mathscr{G}(\mathscr{G}(A)) = \cap \mathscr{A}_{\mathscr{G}(A)} \subset \mathscr{G}(A).$$

Therefore

$$\mathscr{G} \circ \mathscr{G}(A) = \mathscr{G}(A).$$

---

* This property of $\mathscr{G}$ follows directly from the fact that $\varnothing \in A$, which in turn results from the assumption that $\varnothing$ is not excluded as an admissible set of input function (cf Lemma 10.1). If nonempty sets were the only valid candidates for admissible sets of input functions, we would have the result that, for a given nonempty P, $\mathscr{G}(\varnothing)$ would not necessarily be empty; furthermore, $\mathscr{G}(\varnothing)$ would be different, in fact, for different P.

Theorem 10.3. Let A, $B \subset P^R$; then

$$\mathcal{G}(A) \cup \mathcal{G}(B) \subset \mathcal{G}(A \cup B).$$

Proof. From Theorem 10.2(ii), $A \cup B \subset \mathcal{G}(A \cup B)$ and thus

$$A \subset \mathcal{G}(A \cup B).$$

Furthermore, since $\mathcal{G}(A \cup B) \in A$ (Theorem 10.1), it follows that

$$\mathcal{G}(A \cup B) \in \mathscr{A}_A.$$

Therefore

$$\mathcal{G}(A) = \cap \, \mathscr{A}_A \subset \mathcal{G}(A \cup B).$$

By interchanging A and B in the preceding argument, we have $\mathcal{G}(B) \subset \mathcal{G}(A \cup B)$. Thus $\mathcal{G}(A) \cup \mathcal{G}(B) \subset \mathcal{G}(A \cup B)$.

Remark 10.1. There exists a nonempty set P, and sets A_1, $A_2 \subset P^R$ such that $\mathcal{G}(A_1 \cup A_2) \not\subset \mathcal{G}(A_1) \cup \mathcal{G}(A_2)$.

To establish the validity of this remark, let $P = \{p_1, p_2\}$ where $p_1 \neq p_2$. Now let $A_1 = \{c_{p_1}\}$, $A_2 = \{c_{p_2}\}$, where, given a nonempty set Q, for each $q \in Q$, $c_q : R \to Q$ is the *constant function* [1], defined by $c_q(t) = q$ for all $t \in R$.

Observe that A_1, $A_2 \in \mathscr{A}$, since A_1, $A_2 \subset P^R$ and both are closed under translation and segmentation ($c_q \to r = c_q$ for all $q \in Q$ and $r \in R$). Furthermore, $A_1 \subset A_1$ and $A_2 \subset A_2$, and therefore $A_1 \in \mathscr{A}_{A_1}$, $A_2 \in \mathscr{A}_{A_2}$. Thus

$$\mathcal{G}(A_1) = \cap \, \mathscr{A}_{A_1} \subset A_1, \qquad \mathcal{G}(A_2) = \cap \, \mathscr{A}_{A_2} \subset A_2.$$

Since we also have $A_1 \subset \mathcal{G}(A_1)$, $A_2 \subset \mathcal{G}(A_2)$ by Theorem 10.2(ii), it follows that

$$\mathcal{G}(A_1) = A_1, \qquad \mathcal{G}(A_2) = A_2.$$

Thus, $\mathcal{G}(A_1) \cup \mathcal{G}(A_2) = A_1 \cup A_2 = \{c_{p_1}, c_{p_2}\}$. Furthermore, since $\mathcal{G}(A_1 \cup A_2)$ must be closed under segmentation (Theorem 10.1), and since c_{p_1}, $c_{p_2} \in A_1 \cup A_2 \subset \mathcal{G}(A_1 \cup A_2)$, it follows that $c_{p_1} | c_{p_2} \in \mathcal{G}(A_1 \cup A_2)$. However, it will now be shown that $c_{p_1} | c_{p_2} \notin \mathcal{G}(A_1) \cup \mathcal{G}(A_2)$. By the definition of segmentation, we have

$$c_{p_1} | c_{p_2}(t) = \begin{cases} p_1, & t < 0 \\ p_2, & t \geq 0 \end{cases}$$

and therefore

$$c_{p_1} | c_{p_2}(t) \neq c_{p_2}(t), \quad \text{for } t < 0$$

$$c_{p_1} | c_{p_2}(t) \neq c_{p_1}(t), \quad \text{for } t \geq 0$$

that is,

$$c_{p_1} | c_{p_2} \neq c_{p_1} \quad \text{and} \quad c_{p_1} | c_{p_2} \neq c_{p_2}.$$

Thus

$$c_{p_1} | c_{p_2} \notin \{c_{p_1}, c_{p_2}\} = \mathscr{G}(A_1) \cup \mathscr{G}(A_2).$$

We have shown, therefore, that for P, A_1, and A_2 as defined above

$$\mathscr{G}(A_1 \cup A_2) \not\subseteq \mathscr{G}(A) \cup \mathscr{G}(A_2).$$

10.5.3. Extended Closure and the Admissible Set Operator

Theorem 10.2 and the preceding remark have established that, for arbitrary P ($P \neq \varnothing$), the admissible set operator $\mathscr{G}$ on P^R satisfies (i), (ii), and (iii) of the Kuratowski closure axioms but does not, in general,* satisfy (iv), although containment is obtained in the sense of Theorem 10.3. Hammer [13, 10] has introduced and studied the properties of an *extended topology*, in which the primitive concept is that of a *closure function* [10, p. 147] that possesses properties (ii) and (iii) of the Kuratowski closure axioms (Definition 10.7), but differs from a *K*-closure operator in two respects which are of interest when considered in the light of the properties characterizing the admissible set operator:

(a) The null set is not necessarily closed.†

(b) The closure function does not necessarily distribute over finite unions, but *does satisfy* containment in precisely the sense of Theorem 10.3.

In addition to Hammer's innovative work, Rio [16] studied and extended fundamental properties of the Hammer topological system, including further considerations of (a) and (b) above. Thus, as will be shown, the admissible set operator $\mathscr{G}$ is a closure function on P^R in the sense of Hammer, with the

---

* Remark 10.1 establishes that, in fact, the only case in which (iv) is satisfied is the one in which P is the unit set $\{p\}$ and thus $P^R = \{c_p\}$. Since such a set of input states is a degenerate case from a systems-theoretic viewpoint and because we do not wish to carry along the euphemism "in general," we will not so qualify remarks to which exception can be taken on this account only.

† Since the precise meaning of "closure" is not yet needed, we have not defined the term as used here; for purposes of the discussion to follow, it will suffice to interpret the use of the noun "closure" as follows: Given a set X and any closure operator, d, on X, then *closure* in X of $B \subset X$ is $d(B)$. This *topological closure* is to be distinguished from *algebraic closure*, discussed in Definition 10.2. Qualifying phrases (e.g., "under segmentation") used with the latter will accomplish this.

additional property that the null set is closed,* that is, $\mathscr{G}(\varnothing) = \varnothing$. It is interesting to note further that in the case in which the null set is closed for both operators they differ only with respect to the *additivity* property, (iv) of Definition 10.7, and it is this relaxation of additivity which is a principal issue in Hammer's development of extended topology [13].

We now present some of the formal structure of Hammer's closure operator and the associated topological space, and explore its relationship to the admissible set operator.

Definition 10.8. Let X be a set and h a function $h: \mathscr{P}(X) \to \mathscr{P}(X)$; then h is an *H-closure operator* on X if and only if the following statements, the *Hammer closure axioms*, are true:

(a) $A \subset h(A)$ for every $A \subset X$;
(b) $A \subset B \to h(A) \subset h(B)$ for every $A, B \subset X$;
(c) $h \circ h(A) = h(A)$ for every $A \subset X$.

Hammer [13, 10] has used the terms *enlarging, isotonic, and idempotent* to describe properties (a), (b), and (c), respectively. Thus his definition of a *closure function* [10, p. 147] as an enlarging, isotonic, idempotent function coincides with the definition given above of an H-closure operator.

Lemma 10.3. Let X be a set and $h: \mathscr{P}(X) \to \mathscr{P}(X)$; then the following statements are equivalent:

(I) h is an H-closure operator on X;
(II) h has the following properties:
   (a) $A \subset h(A)$ for every $A \subset X$,
   (b') $h(A) \cup h(B) \subset h(A \cup B)$ for every $A, B \subset X$,
   (c) $h \circ h(A) = h(A)$ for every $A \subset X$.

Proof. The equivalence between II(a) and II(c) and the enlarging and idempotent properties, respectively, is established immediately from the definitions. Assume now that $A, B \subset X$. If h is isotonic, it follows that $h(A) \subset h(A \cup B)$ and $h(B) \subset h(A \cup B)$, thus establishing II(b'). If, conversely, h satisfies II(b') and $A \subset B$, then $h(A) \cup h(B) \subset h(A \cup B) = h(B)$; thus $h(A) \subset h(B)$ and h is isotonic.

The preceding lemma establishes that statement II in the lemma is an equivalent characterization of an H-closure operator.

---

* This additional property is not an intrinsically fundamental characterization of $\mathscr{G}$, but is rather a consequence of the fact that we did not exclude the null set from the collection of the admissible set of input functions (Definition 10.3) for a general system. This may be compared with the approach of Wymore [2, Definition 2.1], who requires admissible sets of input functions to be nonempty.

In the context of this general definition of a closure operator, we may introduce a generalization of the (classical) topological space, namely, an *H-Appert* space,* as follows.

Definition 10.9. An *H-Appert* space is an ordered pair (X, h), where X is a set and h is an *H*-closure operator.

This definition is analogous to that of a (classical) topological space in term of a set X and a Kuratowski closure operator on X.

Having now at our disposal the *H*-Appert space, (X, h), we consider next the question of the existence of a *relative*, or *induced*, *H*-closure operator on subsets of X. This is an important systems-theoretic consideration since, given a set of *input states*, P, for a Wymore general system, the set of *input functions* for the system is a subset of P^R, say F, and thus we are interested ultimately in closure in $F \subset P^R$, which will generate a *natural* subspace of (X, h) in the sense that the subspace inherits its properties from (X, h). In terms of the *H*-closure operator, h, we may state the question as follows: Given an *H*-Appert space (X, h) and $A \subset X$, is the *induced function*, say h_A, defined by associating the closure in A of $D \subset A$ with the intersection of A and the closure in X of D, an *H*-closure operator on A; that is, is (A, h_A) an *H*-Appert space? This question was answered in the affirmative by Rio [16].

Because of the importance of this result we present a proof (Theorem 10.4) based on Rio's work and using the notation and definitions introduced in Rio's paper.

First we have the following definition.

Definition 10.10. Let (X, h) be an *H*-Appert space and $A \subset X$; then the *relative closure operator* on A is the function

$$h_A : \mathscr{P}(A) \to \mathscr{P}(A)$$

defined by

$$h_A(B) = h(B) \cap A \quad \text{for every } B \subset A.$$

The next important result [16, Theorem 2.7] establishes that the relative closure operator is an *H*-closure operator on the subset A.

Theorem 10.4 [16]. Let (X, h) be an *H*-Appert space, and $A \subset X$; then the *relative closure operator*, h_A, is an *H*-closure operator on A.

Proof. We must establish that $h_A : \mathscr{P}(A) \to \mathscr{P}(A)$ is enlarging, isotonic, and idempotent according to Definition 10.8.

---

* The terminology here is derived from Hammer [10, p. 149], who defines an *Appert space* in terms of an *H*-closure operator and another "dual function" determined by h: the latter is not considered here.

(a) Let $D \subset A$; then

$$h_A(D) = h(D) \cap A.$$

Since h is enlarging, $D \subset h(D)$, and thus

$$D \subset h(D) \cap A = h_A(D).$$

(b) Let $D, E \subset A$ so that $D \subset E$. We have, since h is isotonic,

$$h(D) \subset h(E).$$

Thus

$$h_A(D) = h(D) \cap A \subset h(E) \cap A = h_A(E).$$

(c) Let $D \subset A$ and consider $h_A(h_A(D))$:

$$h_A(h_A(D)) = h_A(h(D) \cap A) = h(h(D) \cap A) \cap A.$$

Since h is enlarging, we have

$$h(h(D) \cap A) \subset h(h(D)), \quad h(A)$$

or

$$h(h(D) \cap A) \subset h(h(D)) \cap h(A).$$

Thus since h is also idempotent, we may write

$$h_A(h_A(D)) \subset h(D) \cap h(A) \cap A.$$

Using the enlarging property of h again, we have

$$h_A(h_A(D)) \subset h(D) \cap A = h_A(D).$$

Thus

$$h_A(h_A(D)) \subset h_A(D).$$

However, since $h_A(D) \subset A$, (a) above implies

$$h_A(D) \subset h_A(h_A(D)),$$

therefore

$$h_A \circ h_A(D) = h_A(h_A(D)) = h_A(D).$$

From Theorem 2.11 follow immediately the important results concerning subspaces of an H-Appert space.

Theorem 10.5. Let (X, h) be an H-Appert space and $A \subset X$; then the ordered pair (A, h_A) is an H-Appert space.

Proof. From Theorem 10.4 h_A is an H-closure operator.

Definition 10.11. A *subspace* of an H-Appert space (X, h) is a pair (A, h_A), where A is a subset of X.

Theorem 10.6. A subspace of an H-Appert space is an H-Appert space.

Proof. Let (X, h) be an H-Appert space; then from Definition 10.11 a subspace of (X, h) is a pair (A, h_A), where $A \subset X$. From Theorem 10.5, the pair (A, h_A) is an H-Appert space. Since the pair (X, h) and $A \subset X$ were arbitrary, the result follows.

Having introduced some of the formal structure of the H-closure operator and the associated H-Appert space, we now consider the relationship between this structure and the admissible set operator.

It is immediately evident from Theorem 10.2 and 10.3 and Lemma 10.3 that the following result holds.

Theorem 10.7. Let $P \neq \emptyset$ be given; then the admissible set operator $\mathscr{G}$ is an H-closure operator on P^R.

Proof. From Theorems 10.2 and 10.3, the admissible set operator satisfies (a), (c), and (b′) of statement II in Lemma 10.3. Thus, from Lemma 10.3, $\mathscr{G}$ is an H-closure operator on P^R.*

The following theorems are an immediate consequence of Theorem 10.7.

Theorem 10.8. Let $P \neq \emptyset$ be given; then the ordered pair $(P^R, \mathscr{G})$ is an H-Appert space.

Proof. From Theorem 10.7, $\mathscr{G}$ is an H-closure operator on P^R, and thus the result follows from Definition 10.9.

Theorem 10.9. Let $P \neq \emptyset$ be given, $A \subset P^R$, and $\mathscr{G}$ be an H-closure operator on P^R; then

   (i) $\mathscr{G}_A$, the relative closure operator on A, is an H-closure on A.
   (ii) $(A, \mathscr{G}_A)$ is a subspace of $(P^R, \mathscr{G})$.
   (iii) $(A, \mathscr{G}_A)$ is an H-Appert space.

Proof. (i) From Theorem 10.8 $(P^R, \mathscr{G})$ is an H-Appert space, and from Theorem 10.4, observing that $A \subset P^R$, it follows that $\mathscr{G}_A$ is an H-closure operator on A.
   (ii) The result follows from Theorem 10.8 and Definition 10.11.
   (iii) The result follows from (ii) and Theorem 10.6.

---

* This result can also be established as a corollary to Lemma 10.1, Theorem 10.1, and a result given by Rio [16, Theorem 5.2].

From a systems-theoretic viewpoint, that is, interpreting P as a given set of input states for a Wymore system, notice that, since $F \in \mathscr{A}$ implies $\mathscr{G}(F) = F$, we have the result that the closed sets in $(P^R, \mathscr{G})$ are the admissible sets of input functions for the system. Thus we have a *natural* topological characterization, in terms of the admissible set operator $\mathscr{G}$, of the admissible set of input functions for a Wymore system having input state set P.

Thus the closed subspaces $(F, \mathscr{G}_F)$ of $(P^R, \mathscr{G})$ are a topological extension, in the sense of Hammer, of the admissible sets of input functions for a Wymore general system.

In summary, we have established in this section that, for the Wymore system model with input state set P:

1. The admissible set operator $\mathscr{G}$ is an H-closure operator on P^R having $\varnothing$ as a closed set.

2. $(P^R, \mathscr{G})$ is an H-Appert space.

3. For any subset A of P^R, $(A, \mathscr{G}_A)$ is a subspace of $(P^R, \mathscr{G})$, where $\mathscr{G}_A$ is the relative closure operator on A.

In addition, it has been demonstrated that the H-Appert space introduced is unique, is independent (in the abstract) of P, and requires the formalization of only *one* additional function-theoretic concept, namely, that of the admissible set operator.

It has also been established that, except for singleton input state sets, the strongest closure structure associated with the admissible set operator is the one which characteries the H-closure operator.

Finally, through the identification of admissible sets of input functions for a Wymore system as the closed subspaces of $(P^R, \mathscr{G})$, the work on generalized closure spaces may be introduced naturally into the axiomatic structure of this general system model. In particular, the use of this generalized topological structure may be explored at the abstract level to account for the topological aspects underlying the systems-theoretic properties of approximation, continuity, optimization, and so forth, as discussed earlier. An intrinsic connection between H-Appert spaces and the Wymore theory having been established, the possible use of such structure in other formal abstract models also suggests itself.

10.6. CLOSURE SPACES IN GENERAL SYSTEMS THEORY: SOME PERSPECTIVES ON FUTURE WORK

We have arrived at the following point: in Sections 10.2 and 10.3 we motivated the position that topological structure should be an intrinsic component of any abstract theory of general systems—the method used was

inductive via a brief description of the role of such structure in a broad class of specific systems; in Section 10.5 an examination of the weakest (most general) closure properties possessed by the set-theoretic structure representing one approach [12] to an abstract model of general systems led to a topological structure of greater generality than the (classical) topological space generated by the usual Kuratowski closure axioms. We now indicate three immediate questions which arise from these efforts.

1. Although we have developed a natural topological extension of the set of input functions for a Wymore general system as an H-Appert space, the use of such structure in representing the systems-theoretic properties of approximation, continuity, optimization, and so on in the Wymore model needs to be explored.

2. From an inductive viewpoint, it is clear from the exhibits presented in Section 10.2 that topological structure is a fundamental trait of an extremely broad class of systems, yet the structure involved in almost all cases is that of the (classical) topological space (the one exception being the work on tolerance automata [16]). Therefore, the question of the need for the weaker (more general) structure characterizing the H-Appert space should be explored. That such general structure is needed for the class of finitary systems in mathematics has been advocated in the work of Hammer [10]; however, for general models representing systems which are not the deductive ones of mathematics, the question is open. Examples and insight are needed here.

3. Hammer [10] has explored for Appert spaces the nature of the topological concepts of connectedness, separation, continuity, convergence, and so on in this generalized closure space. With the representation of admissible sets of input functions for the Wymore system as an H-Appert space generated by the admissible set operator, the interpretation of these extended topological notions from a systems-theoretic viewpoint should be explored. Again, specific system examples would provide insight to be used at the abstract level.

The search for answers to these questions and for understanding—both formal and intuitive—of these areas is of interest as part of the goal of introducing into abstract models of general systems a relevant topological structure.

ACKNOWLEDGMENT

PROBLEMS

10.1. Let d be a function,

$$d: (R \times R) \times (R \times R) \to R,$$

such that

$$d((x_1, y_1), (x_2, y_2)) = (x_1 - x_2)^2 + (y_1 - y_2)^2.$$

Verify that distance d satisfies the following properties:

    (i) $d((x_1, y_1), (x_2, y_2)) \geq 0$;
    (ii) $d((x_1, y_1), (x_2, y_2)) = 0$ if and only if $(x_1, y_1) = (x_2, y_2)$;
    (iii) $d((x_1, y_1), (x_2, y_2)) = d((x_2, y_2), (x_1, y_1))$;
    (iv) $d((x_1, y_1), (x_3, y_3)) \leq d((x_1, y_1), (x_2, y_2)) + d((x_2, y_2), (x_3, y_3))$.

Find another function, $d_1: (R \times R) \times (R \times R) \to R$, which also satisfies these four properties.

Any ordered pair (X, d), where $d: X \times X \to R$ and d satisfies the properties mentioned, is called a *metric space*.

10.2. Consider $d: (R \times R) \times (R \times R) \to R$ of Problem 10.1. This is the usual definition of distance in the Cartesian plane. Now take a circle with center (a_1, a_2) and radius δ. All points of the circle except those on the circumference constitute a set. Let us call such a subset of $R \times R$ an *open ball*—"open" because the points on the circumference are not included. Since a ball can be identified uniquely by its center and radius, we name an open ball with center (a_1, a_2) and radius δ, $B((a_1, a_2), \delta)$. Convince yourself that $B((a_1, a_2)\delta)$ can be written as

$$B((a_1, a_2), \delta) = \{(x, y): (x, y) \in R \times R \quad \text{and} \quad d((a_1, a_2), (x, y) < \delta\}.$$

Now consider a set $\mathcal{T}$ whose elements are all such open balls and subsets generated by arbitrary union of these open balls. Show that

$$((R \times R) \times (R \times R), \mathcal{T})$$

satisfies all the properties necessary to be a topological space.

10.3. Reformulate Problems 10.1 and 10.2 for the Euclidean distance

$$d: R \times R \to R$$

where $d(x, y) = |x - y|$ and verify the assertions in the formulated questions. Remember that the elements of the domain of d are of the form (x, y), not $((x_1, x_2), (y_1, y_2))$. Can we talk in terms of circles any more? What is the interpretation now?

10.4. Consider the function represented by the graph in Figure 10.3. Determine whether the function is continuous at (i) a, (ii) b, and (iii) c. Explain, using the definition of continuity given.
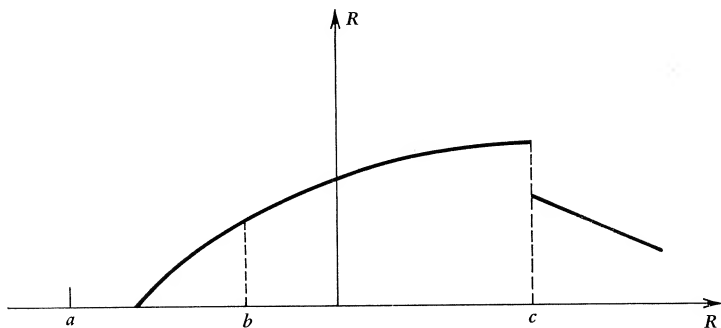
Figure 10.3. Illustration for Problem 10.4.

10.5. Let X be the set $X = \{f \mid f: [0, 1] \to R, f \text{ is continuous}\}$. Let d be the function $d: X \times X \to R$, where

$$d(x, y) = \int_0^1 |x(t) - y(t)| \, dt, \quad x, y \in X.$$

Show that d is a distance function, that is, d satisfies the four conditions listed in Problem 10.1.

10.6. Consider the relation ξ on R defined by

$$(a, b) \in \xi \quad \text{if and only if} \quad |b - a| \leq 1.$$

Show that (R, ξ) is a tolerance space.

10.7. A merchant buys a large variety of commodities from retailers. For each item that he purchases, the merchant knows the average price. He buys an item if the price is within 5% of this average. (He is guided by quality and demand considerations.) Let ξ be a relation on R^{++} such that

$$(a, b) \in \xi, \quad |a - b| < 0.05a.$$

Here a is the average price and b is the current price. Is (R, ξ) a tolerance space?

10.8. Verify that d as defined in Section 10.2,

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1/k, \text{ where } k = \min\{i \in I^+ \mid x(i) \neq y(i)\}, & \text{if } x \neq y, \end{cases}$$

is a distance function on X.

GLOSSARY OF SYMBOLS

R	The real line
PBI(B)	Set of all functionals with past of bounded input, bounded by B
$N_\varepsilon(x)$	Set of all points within distance ε of x
$\underset{}{\times}\,\mathscr{C}$	Cartesian product of collection C
B^A	Set of all functions from A to B
$f \circ g$	Composition of functions f and g
$f \mid A$	Restriction of function f to subset A; segmentation of functions f and A
$\mathscr{P}(A)$	Power set of A
$\mathscr{A}$	Set of admissible functions
$\mathscr{G}$	Admissible set operator
$\mathscr{A}_B$	Containing collection of set B
$\mathscr{G}(A)$	Admissible set of input functions generated by A
h	H-closure operator; system functional
h_A	Relative H-Appert closure operator on set A
(A, h)	H-Appert space with h an H-closure operator on A
$\lVert \cdot \rVert$	Norm of $\cdot$
I^+	Set of positive integers
R^+	Set of positive reals
R^{++}	Set of nonnegative reals
$\lesssim$	Preordering relation

REFERENCES

1. **Debreu, G.,** "Topological Methods in Cardinal Utility Theory." In *Mathematical Methods in the Social Sciences*, K. Arrow, S. Karlin, and P. Suppes (Editors), pp. 16–26. Stanford, Calif.: Stanford University Press, 1960.

2. **Schweber, S.,** *Introduction to Relativistic Quantum Field Theory*, Section 6. New York: Harper & Row, 1961.

3. **Klir, G. J.,** *An Approach to General Systems Theory*. New York: Van Nostrand Reinhold, 1969.

4. **Atkinson, D.,** and **M. Halpern,** "Non-usual Topologies on Space-time and High-energy Scattering." *Journal of Mathematical Physics*, Vol. 8, pp. 373–387, 1967.

5. **Gudder, S. P.,** "Elementary Length Topologies in Physics." *Journal of the Society for Industrial and Applied Mathematics*, Vol. 16, pp. 1011–1019, 1968.

6. **Allport, F. H.,** *Theories of Perception and the Concept of Structure*. New York: Wiley, 1955.

7. **Brilliant, M. B.,** "Theory of the Analysis of Nonlinear Systems." *Technical Report* 345, Massachusetts Institute of Technology, Research Laboratory of Electronics, March 1968.

8. **Wymore, A. W.,** "Discrete Systems and Continuous Systems." In *Advances in Mathematical Systems Theory*, P. C. Hammer (Editor). University Park, Pa.: Pennsylvania State University Press, 1969.

9. **Mesarovic, M. D.,** "Auxiliary Functions and Constructive Specification at General Systems." *Mathematical Systems Theory*, Vol. 2, pp. 203–222, 1968.

10. **Hammer, P. C.,** "Continuity," "Filters in General," "Approximation Spaces," "Charts in Elemental Mathematics." In *Advances in Mathematical Systems Theory*, P. C. Hammer (Editor). University Park, Pa.: Pennsylvania State University Press, 1969.

11. **Kalman, R. E., P. L. Falb,** and **M. A. Arbib,** *Topics in Mathematical Systems Theory*. New York: McGraw-Hill, 1969.

12. **Wymore, A. W.,** *A Mathematical Theory of Systems Engineering—the Elements*. New York: Wiley, 1967.

13. **Hammer, P. C.,** "General Topology, Symmetry and Convexity." *Transactions of the Wisconsin Academy of Sciences, Arts and Letters*, Vol. 43, pp. 221–225, 1955.

14. **Dugundji, J.,** *Topology*. Boston: Allyn and Bacon, 1966.

15. **Kelley, J. L.,** *General Topology*, corr. ed., Princeton, N.J.: Van Nostrand, 1961.

16. **Rio, S. T.,** *On the Hammer Topological System*. Doctoral Dissertation. Ann Arbor, Mich.: University Microfilms, 1959.

17. Society for General Systems Research, *General Systems Bulletin*, March 1971.

18. **Zadeh, L.,** "The Concept of State in System Theory." In *Views on General Systems Theory*, M. D. Mesarovic (Editor). New York: Wiley, 1964.

19. **Kalman, R.,** "Mathematical Description of Linear Dynamical Systems," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 1, pp. 152–192, 1963.

20. **Balakrishnan, A. V.,** "State Space Theory of Linear Time-Varying Systems." In *System Theory*, L. A. Zadeh and E. Polak (Editors). New York: McGraw-Hill, 1969.

21. **Halkin, H.,** "Topological Aspects of Optimal Control of Dynamical Polysystems." *Bell Telephone System Monograph* 4947, 1964.

22. **Volterra, V.,** *Theory of Functionals and of Integral and Integro-Differential Equations* New York: Dover, 1959.

23. **Arbib, M.,** "Automata Theory: The Rapprochment with Control Theory." *In Topics in Mathematical System Theory*, R. E. Kalman, P. L. Falb, and M. A. Arbib (Editors). New York: McGraw-Hill, 1969.

24. **Pontryagin, L. S.,** et al. *The Mathematical Theory of Optimal Processes*. New York: Wiley, 1952.

25. **Hellerman, L., W. L. Duda,** and **S. Winograd,** "Continuity and Realizability of Sequence, Transformations." *IEEE Transactions on Electronic Computers*, Vol. EC–15, No. 4, pp. 560–569, August 1966.

26. **Falb, P. L.,** and **E. Polak,** "Conditions for Optimality." In *System Theory*, L. A. Zadeh and E. Polak (Editors). New York: McGraw-Hill, 1969.

27. **Arbib, M.,** "A Common Framework for Automatic Theory and Control Theory." *Journal of the Society for Industrial and Applied Mathematics*, Vol. 3, Ser. A, pp. 206–222, 1965.

28. **Hoffman, H.,** "Mathematical Anthropology." In *Biennial Review of Anthropology*, B. J. Siegel (Editor). Stanford, Calif.: Stanford University Press, 1970.

29. **Leach, E. R.,** *Rethinking Anthropology*. London School of Economics Monographs on Social Anthropology, No. 22. London: the Althone Press, University of London, 1966.

30. **Lancaster, K.,** *Mathematical Economics*. New York: Macmillan, 1968.

31. **Hurwicz, L.,** "Programming in Linear Spaces." In *Studies in Linear and Non-Linear Programming*, K. J. Arrow, L. Hurwicz, and H. Uzawa (Editors). Stanford, Calif.: Stanford University Press, 1958.

32. **Lewin, K.,** *Principles of Topological Psychology*. New York: McGraw-Hill, 1936.

33. **Ashby, W. R.,** *An Introduction to Cybernetics*. London: Chapman and Hall, 1956.

34. **Rashevsky, N.,** *Mathematical Biology of Social Behavior*, rev. ed. Chicago: University of Chicago Press, 1959.

35. **Hall, D.,** and **G. L. Spencer,** *Elementary Topology*. New York: Wiley, 1955.

36. **Cornacchio, J. V.,** *General Systems Theory: Mathematics, Models, and Methods*. Manuscript: Seminar notes, School of Advanced Technology, State University of New York at Binghamton, 1969–70.

37. **Rashevsky, N.,** *Mathematical Biophysics*, Vol. II, 3rd rev. ed. New York: Dover, 1960.

11.   Relative Explanations of Systems

LARS LÖFGREN

Electrical Engineering Department, University of Lund, Sweden

EDITOR'S COMMENTS

The prerequisites for this chapter are completely included in its two appendices. However, the reader should feel comfortable with mathematics to be able to read these appendices. Generally, the books specified in the Editor's Comments to Chapter 8 will be useful for Chapter 11 as well. The reader should also take advantage of the Glossary of Symbols (p. 405).

11.1. INTRODUCTION

System is frequently used as a synonym for *order*. This may not be considered a very good explanation of "system" but is, nevertheless, a genuine explanation. Indeed, any effective explanation will have to be finite and must leave a final explanans ("order," in the above example) without explanation, provided that the explanation is not circular. In fact, this is the reason for introducing (unexplained) axioms and axiomatic descriptions as explanations. Hence it is natural to look at the problem of *explicability* as a problem of *formalizability*. This approach is taken in the following sections, where various systems, like learning systems, self-reproducing systems, and evolutionary systems, are explained in terms of formalizability (i.e., in a metalanguage for formal theories).

Instead of elaborating upon what a "system" is, let us here in the introduction indicate an evolutionary effect of this concept. We first have to admit that the search for a system or order in nature is really what constitutes basic research in the natural sciences. Among the classical orders found, we have Mendeleev's periodic system and von Linné's *Systema Naturae*.

The findings of such research permit us to order our surroundings even further. For example, we may synthesize technological systems, erect educational systems, and even compete with natural variation and selection in biological evolution.

The larger and more complex these artificial systems or controllable natural systems are, the more difficult it can be to decide what effects they will have. "Complexity," in fact, may be the opposite of "order." Although the artificial systems are systems in the sense of being created by systematic methods, they may not be orderly enough to permit us to understand their

effects. Components may have emergent properties which are not revealed until they are synthesized in a large system, provided, of course, that we do not have a systems theory that is powerful enough to predict these properties. What is desirable is a developing general systems theory.

The evolutionary effects of research can be looked at from many angles; one is as follows. The order which the scientific discoveries reveal will be talked about not only by the scientists themselves, but also by their colleagues in other fields, and eventually will be taught to a new generation of scientists. In this way, a potential is created for asking deeper questions about the systems revealed, and finally new or related orders and systems may be found. The potential will obviously be greater if the scientists of the various disciplines have been taught to express their findings in a " common scientific language " so that they can easily be understood and in turn can readily understand the nature of their colleagues' findings.

What do we mean by a " common scientific language "? Does a general systems theory qualify as one? Let us briefly discuss the first question and then, concerning the second, say simply that our interpretation of a general systems theory is that it is a steadily evolving, common scientific language.

A large catalog of empirical scientific facts is usually not considered a theory unless it is systematized. The more order that can be found, the shorter an effective description of the catalog can be. If the description is sufficiently short, it may qualify as a theory. It then usually happens that the regularities found extend beyond the listings of the catalog. Hence the theory, as a description, can be very short compared with a particularization of all the scientific facts that it contains. Indeed, these facts may be so numerous that it will be impossible to communicate them to another scientist. However, it may well be possible to communicate effectively the description, or theory, so that the receiver at will can particularize any desired fact contained.

The problem here is determining how the regularities of the catalog should be formulated so that the receiver will know how to interpret them. Even if the description of the facts is supplied with a description (i.e., explanation) of the first description, this second description in turn may have to be described. A practical way out of this regression seems to be to improve the educational system so that, for example, a basic course in mathematical logic is taught to all potential scientists. Such a course should contain a sizable part on formal systems and model theory. A reason for such an educational program will be given in Section 11.2, where a hypothesis is formulated to the effect that everything that can be effectively explained can be formalized. There are many ways of formalizing theories, depending on how the logical basis is chosen. However, I believe that agreement on a suitable logical basis constitutes far less of a problem for a group of scientists than to interact without formalizing their ideas.

As will be further argued in Section 11.2, there is a kind of universality about the concept of effective explicability, comparable to effective computability. As we have witnessed, the modern computer has had unmistakably unifying, interdisciplinary effects. It is my guess that a widespread use of effective explicability could raise these effects to the plane of theory development, and in this way serve as a basis for a common scientific language. It is not suggested, of course, that scientists should be trained in formalizing their ideas to such an extent that they start to think only formally. This may be highly effective, but perhaps somewhat mechanical. Indeed, our inner ways of producing new ideas may be incomprehensible to ourselves, and we could not then, of course, hope to explain them effectively to a colleague. Instead, what is suggested is that, as soon as a scientist believes that he has produced a theory for some phenomenon, he should try to formalize the theory so as to make it effectively communicable.

We want, then, to interpret the concept of general systems theory not necessarily as a formal theory, but rather as a common scientific language in the above sense, that is, as a metalanguage in which we can discuss the effects of various logical bases for effective explanations of certain types of systems phenomena. In the following sections we will use some mathematical logic as such a metalanguage and discuss in it the effective explanation of learning systems and various other particular systems.

The mathematical logic to be used is independently developed in Appendices I and II, "Effective Computability" and "Formal Systems," at the end of this chapter. Reference to a definition, theorem, corollary, or example in the appendices is indicated by affixing the letter A to the reference number. Thus, Definition 11.1A refers to the definition numbered 11.1 in the appendices.

11.2. RELATIVELY EFFECTIVE EXPLANATIONS

Asked to explain the fact that the implication relation is reflexive, a teacher may give a proof of the theorem $W \Rightarrow W$ in the formal theory $\mathscr{L}$ of the propositional calculus (cf. Ex. 11.8A, p. 397). Thereby he provides an explanation relative to $\mathscr{L}$ which may satisfy some students but not all. As the questioning continues, he may be asked to also explain $\mathscr{L}$ and, furthermore, to explain why $\mathscr{L}$ can be interpreted as a fragment of the English language where the implication relation occurs. At each such further question, all the teacher can do is give an explanation relative to some sufficiently well-known concepts.

The explanation of $W \Rightarrow W$ relative to $\mathscr{L}$ illustrates a relatively effective explanation. It is effective in relation to the interrogator in the same sense that a program is effective in relation to a computer. It permits the interrogator to

check (like a computer) every step in the chain of arguments and thus to convince himself that the property in question really can be explained in relation to $\mathscr{L}$.

Many explanations are relatively effective even if not given as a sequence of special symbols. The statement "If the sun is shining then either it is raining or the sun is shining," for example, may be effectively explained in similar verbal statements to a person who understands (has learned or is pre-programmed with) the postulates for the sentential connectives. Also, statements which require proper postulates from the natural sciences can often be effectively explained (or falsified) in relation to accepted postulates and rules of inference. However, a complete reference to all postulates and rules involved is seldom made, and so it may sometimes be very difficult for a physicist to understand a biologist and vice versa. For the same reason, it may even be difficult for one physicist to understand another. Such difficulties, however, are avoidable if the explanation is a p-explanation according to the following definition, because a p-explanation, as we shall see in Theorem 11.1, admits effective reconstruction of missing references.

Definition 11.1. A sequence E is a *p-explanation in $\mathscr{S}$* (relative to $\mathscr{S}$) if E is a proof sequence in $\mathscr{S}$ and $\mathscr{S}$ is r-formal (cf. Def. 11.18A, p. 396).

Let us emphasize that many authors simply denote as formal a theory which here is defined as r-formal (recursively formal). However, other authors mean by "formal theory" a more general concept. Since in this chapter there are places where a theory has to be defined as in Definition 11.18A, we feel it necessary to complicate the notation with the term "r-formal."

Theorem 11.1. With every r-formal theory $\mathscr{S}$, there is a Turing machine which decides (i.e., it can be effectively decided) whether or not an alleged explanation is a p-explanation in $\mathscr{S}$. This is true even if the explanation does not explicitly explain which axioms and rules of $\mathscr{S}$ have been used.

Proof. According to the proof of Tm. 11.41A, p. 396, $P_{\mathscr{S}}$, the set of proofs in $\mathscr{S}$ is recursive if $\mathscr{S}$ is r-formal. Hence, by Tm. 11.3A, p. 375, with every r-formal theory $\mathscr{S}$ there is a Turing machine which accepts the proofs in $\mathscr{S}$, that is, the p-explanations in $\mathscr{S}$, and rejects all other alleged explanations.

There is no doubt, according to Theorem 11.1, that a p-explanation in an r-formal theory really can be agreed upon in an effective way, and in this sense is effectively understandable.

Conversely, there are reasons to question whether an explanation cannot be effectively understandable (relatively effective) even if it is an explanation in a theory which does not have all of the constraints imposed on an r-formal theory. However, in what follows, we shall be able to give support for an hypothesis in the opposite direction.

Explanation Hypothesis I. If an explanation E is effectively understandable (relatively effective), then it is understandable in terms of the rules (explanation arguments) and axioms (postulates) which constitute an r-formal theory $\mathscr{S}$, such that E is a p-explanation in $\mathscr{S}$.

Let us first remark that this hypothesis cannot itself be proved because the meaning of "effectively understandable" is only intuitively understood. We can, however, give the following arguments for the hypothesis.

A first consequence of the hypothesis is that an explanation sequence must be finitely long to be effectively understandable. An argument is simply that each explanation step requires a nonzero time to be understood and the whole explanation would never come to an end, that is, never be understood, if it were not finite. However, this does not prevent nonfinite objects from being effectively understood. For example, the set of integers can be effectively understood by giving a p-explanation relative to an r-formal number theory.

Next, let us examine an objection against the hypothesis, namely, that Theorem 11.1 may hold even if the theory $\mathscr{S}$ is not r-formal but, more generally, has a recursively enumerable (instead of recursive) set of axioms. Let us assume that the set of axioms A in the theory $\mathscr{S}$ is recursively enumerable so that $A = \{f(n): n \in N\}$, where the enumerating function f is recursive. Furthermore, let us agree to attach the integer n, which generates an axiom $a(=f(n))$ to the axiom itself as soon as the axiom occurs in a proof sequence. Then we can effectively decide about an alleged proof sequence, whether an $\langle a, n \rangle$-occurrence is true in the sense that $a = f(n)$, because f is recursive. Under this convention of writing a proof, the crucial argument in the proof of Theorem 11.1 stands even though A is recursively enumerable and eventually not recursive. Indeed, we use this convention in Appendix II for the proof of $W \Rightarrow W$ in the propositional logic $\mathscr{L}$ (see Ex. 11.8A, p. 397). Instead of writing this proof as

$$((W \Rightarrow ((W \Rightarrow W) \Rightarrow W)) \Rightarrow ((W \Rightarrow (W \Rightarrow W)) \Rightarrow (W \Rightarrow W)))\,(W \Rightarrow ((W \Rightarrow W) \Rightarrow W))((W \Rightarrow (W \Rightarrow W)) \Rightarrow (W \Rightarrow W))(W \Rightarrow (W \Rightarrow W))(W \Rightarrow W),$$

which can be effectively checked to be a proof of $W \Rightarrow W$ in $\mathscr{L}$, we attach the axiom number to each axiom occurrence in the proof in Appendix II (and even supply further explanations).

Now, the main point of our argument in support of Explanation Hypothesis I is the following. Suppose that the explanation E is effectively understandable because of the convention of writing $\langle a, n \rangle$ instead of (a) for the axioms. Then, although the set of axioms may be recursively enumerable, the set of axiom-number pairs is recursive because the predicate $a = f(n)$ is computable. Hence the new explicit way of writing the explanation sequence E is equivalent to a sequence written in the conventional way, where indeed the set of axioms is recursive, that is, the reference theory $\mathscr{S}$ is r-formal.

Another natural objection to Explanation Hypothesis I is that Theorem 11.1 may hold even if the theory $\mathscr{S}$ is not r-formal, but instead has rules of inference that are semicomputable (Def. 11.11A, p. 380) instead of computable. Let $R(W_{i_1}, W_{i_2}, \ldots, W_{i_n})$ (Def. 11.17A, p. 396) be a semicomputable rule, that is, the set $\{\langle W_{i_1}, W_{i_2}, \ldots, W_{i_n}\rangle : R(W_{i_1}, W_{i_2}, \ldots, W_{i_n})\}$ is recursively enumerable (Tm. 11.17A, p. 385). Let $\psi_z(x)$ (Def. 11.1A, p. 380) be the recursive function that enumerates this set so that $\psi_z(m) = \tau(W_{m_1}, W_{m_2}, \ldots, W_{m_n})$, where τ is the n-tuple function according to Def. 11.12A, p. 382. To have Theorem 11.1 hold in this more general case, we may agree always to write $\langle W_{m_1}, z, m\rangle$ in a proof where we previously wrote W_{m_1} as the consequence of $W_{m_2}, W_{m_3}, \ldots, W_{m_n}$ according to the rule R (computed by z). Indeed this is the type of explanation we supplied to the proof of $W \Rightarrow W$ in Appendix II (see Ex. 11.8A, p. 397), although redundant in this case. We can always check effectively whether a given triple $\langle W_{m_1}, z, m\rangle$ is true in the sense that $\psi_z(m) = \tau(W_{m_1}, W_{m_2}, \ldots, W_{m_n})$ because z can be effectively decoded into a Turing machine which computes $\psi_z(m)$. Hence the new explicit way of writing the explanation sequence is equivalent to writing a sequence according to the conventional way, where indeed the rules of inference are computable. The hypothesis is thus supported also on this point.

A third natural objection to the hypothesis concerns the finiteness of the set of rules of inference in an r-formal theory. Could we not have the proof of Theorem 11.1 stand with a nonfinite, but computable, set of computable rules of inference? Let S be such a computable set of computable rules. Then S is the extension of a computable predicate $P(z)$ saying that z is the code number of one of the computable rules of inference, say

$$R_z(Y, X_1, X_2, \ldots, X_n).$$

However, since $P(z)$ is computable,

$$R^*(Y, z, X_1, X_2, \ldots, X_n) =_{\text{def}} R_z(Y, X_1, X_2, \ldots, X_n)$$

is a single computable rule of inference that works on proofs where, instead of a consequence Y, we write the pair $\langle Y, z\rangle$. Provided that we accept this way of writing proofs, we have an affirmative answer to our previous question. We also have the somewhat surprising result that the whole computable set of rules can be replaced by a single rule of inference, namely,

$$R^*(Y, z, X_1, X_2, \ldots, X_n).$$

But if we do not affix a z to a consequence Y when writing a proof, the answer would be negative. We would then have to decide upon

$$\exists z\, R^*(Y, z, X_1, X_2, \ldots, X_n)$$

in order to know whether an alleged Y really is a consequence according to some rule of inference. And, although R^* is computable, this predicate

$\exists z \, R^*$ may not be computable (Tm. 11.12A, p. 382). As soon as the set of rules is finite, however, the existential quantifier is bounded and the predicate is computable.

Finally, let us consider the case of a recursively enumerable set S of semi-computable rules of inference. Let $f(m)$ be the recursive function that enumerates S; $f(m) = k$ indicates that k is the code number of the recursive function $\psi_k(v)$ that enumerates the extension of the mth rule of inference, R_m. Hence $f(m) = k$ and $\psi_k(v) = \tau(Y_{v_1}, Y_{v_2}, \ldots, Y_{v_n})$ implies that $R_m(Y_{v_1}, Y_{v_2}, \ldots, Y_{v_n})$ holds. If we write our proof with a triple $\langle Y, m, v \rangle$ instead of a single Y as consequence, we can effectively check whether or not

$$\psi_{f(m)}(v) = \tau(Y, X_1, X_2, \ldots, X_n),$$

that is, Theorem 11.1 will still hold. At this point we realize that $\psi_{f(m)}(v) = g(m, v) = h(\tau(m, v))$, where g and h are computable functions of two arguments and one argument, respectively (τ is the pair function of Def. 11.12A, p. 382). Hence it suffices to write $\langle Y, \mu \rangle$ instead of $\langle Y, m, v \rangle$ in the proofs. We can still effectively compute $h(\mu) = \tau(Y, X_1, X_2, \ldots, X_n)$ and thus effectively check Y as a consequence of the rules of inference, even though these rules now are only semicomputable and the set of rules recursively enumerable. As in the previous case it is necessary to affix a μ to each consequence Y. If we did not, we would have to verify Y by deciding on

$$\exists \mu [h(\mu) = \tau(Y, Y_1, \ldots, Y_n)].$$

This predicate, although semicomputable, need not be computable, that is, Theorem 11.1 would fail.

All our arguments thus support Explanation Hypothesis I in the sense that the stipulation of the reference theory $\mathscr{S}$ as r-formal (Def. 11.18A, p. 396) is precisely what makes the proof of Theorem 11.1 stand. It is true that we equivalently could require that $\mathscr{S}$ possess only one rule of inference. However, the greater ease of writing proofs in a theory with a few but finite number of rules seems to be advantageous.

In the following sections Explanation Hypothesis I will be complemented with hypotheses in the reverse direction, namely, that concerning the explanatory and predictive power of formal theories.

11.3. SYNTACTIC INFORMATION OF A THEORY

In [1] Bar-Hillel and Carnap develop a concept of semantic information. Let W_1 and W_2 be two well-formed formulas (wffs) in a formal system such that $W_1 \Rightarrow W_2$ is logically true.* Then, in a semantic sense, W_1 asserts all that

---

* A wff W is logically true, denoted $\vDash W$, if W is tautology (according to the propositional calculus) or if W is true for every interpretation (according to first-order theories). For further clarifications of this semantic concept, see, for example, [16].

is asserted by W_2 and probably more. In other words, the information, Inf(W_1) (conceived as a set), carried by W_1 includes the information, Inf(W_2), carried by W_2 as a (perhaps improper) part. This stipulation, that is,

$$[\text{Inf}(W_2) \subset \text{Inf}(W_1)] \equiv [\vDash(W_1 \Rightarrow W_2)],$$

is basic to the development in [1] of semantic information theory. Among the models suggested for Inf(W) is the set of all wffs W' such that $\vDash(W \Rightarrow W')$ and such that not $\vDash W'$. The implied wffs which are logically true are deleted because they carry no specific semantic information.

In a context of formalizability where the syntactic concepts dominate the semantic, it is natural to conceive of information as a syntactic rather than semantic concept. We could then say that, if $W_1 \Rightarrow W_2$ is a theorem in a formal theory $\mathscr{S}$, then W_1 implies (asserts) everything that is implied (asserted) by W_2 and possibly more. In other words, the syntactic information (conceived as a set) carried by W_1 includes the syntactic information carried by W_2 as a (perhaps improper) part. However, instead of defining syntactic information in terms of the special connective $\Rightarrow$, which may even be missing in some formal systems, it seems more natural to consider the concept of deducibility. A wff W_2 is said to be deducible in $\mathscr{S}$ from the hypothesis W_1, denoted $W_1 \vdash_{\mathscr{S}} W_2$, if W_2 is a theorem in $\mathscr{S}'$, obtained from $\mathscr{S}$ by augmenting the axioms with the wff W_1. Thus, if $W_1 \vdash_{\mathscr{S}} W_2$, the wffs deducible in $\mathscr{S}$ from the hypothesis W_2 are among the wffs deducible in $\mathscr{S}$ from the hypothesis W_1. Hence W_1 is at least as informative as W_2 in relation to $\mathscr{S}$ in the sense that W_1 is a stronger hypothesis than W_2 (or is at least as strong). Every wff that is asserted by W_2 in relation to $\mathscr{S}$ is also asserted by W_1 in relation to $\mathscr{S}$. This motivates the following definitions.

Definition 11.2. $I(W, \mathscr{S})$, a *syntactic information of a wff W in a formal theory* $\mathscr{S}$, is the following set of wffs in $\mathscr{S}$: $I(W, \mathscr{S}) = \{X: W \vdash_{\mathscr{S}} X, X \notin T_{\mathscr{S}}\}$.

We delete the wffs X that belong to $T_{\mathscr{S}}$, the theoremhood of $\mathscr{S}$, because a theorem in $\mathscr{S}$ can be deduced from the axioms alone and hence also from any hypothesis. Therefore a theorem does not contribute to the information carried by a specific wff, W.

Our aim is, however, to define a syntactic information, not of the particular wffs, but of a formal theory as such. Later we want to use this information concept to measure the explicatory power of a formal theory.

Let us consider what usually happens in the process of formalizing experimental findings. We first seek some rules or regularities in the findings. Such rules permit us to describe the findings, not as separate facts, but as a formal theory, say $\mathscr{S}$. Usually the rules of inference in $\mathscr{S}$ are of a logical nature, fixed in advance of the formalization. The same may be true of some axioms, called the logical axioms in $\mathscr{S}$. The rest of the axioms in $\mathscr{S}$, called proper

axioms, properly result from the experimental findings. Thus, in the process of formalizing the findings, we begin with a logical basis $\mathscr{B}$, that is, a theory containing all logical axioms and usually all rules of inference. Next, the experimental findings lead us to formulate and investigate an hypothesis, say a wff h_1. This hypothesis is reasonable if its syntactic information relative to $\mathscr{B}$ agrees with the information which has already been obtained experimentally or which is to be obtained by further experiments (an interpretation step is obviously involved here; because of lack of space, we want to suppress this step as much as possible). If h_1 is reasonable, it will be added to the axioms of $\mathscr{B}$ as a proper axiom, thus forming $\mathscr{S}'$. Next, an hypothesis h_2 is investigated in relation to $\mathscr{S}'$. If the syntactic information of h_2 relative to $\mathscr{S}'$ agrees with the information obtained by experiments, h_2 may be added to the axioms of $\mathscr{S}'$, thus forming $\mathscr{S}''$, and so on. If in this way a theory $\mathscr{S}$ is obtained such that the union of these successive syntactic informations, that is, $T_{\mathscr{S}} - T_{\mathscr{B}}$, includes all the information that has been experimentally obtained, then $\mathscr{S}$ is a theory for these experiments.

It is reasonable, therefore, to consider $T_{\mathscr{S}} - T_{\mathscr{B}}$ as the syntactic information of the whole theory $\mathscr{S}$ relative to its logical basis $\mathscr{B}$. It may happen that the explanation (formalization) of a certain phenomenon requires an hypothesis that will not have the function of an axiom, but of a proper rule of inference. In that case $\mathscr{S}$ should contain both logical and proper axioms and both logical and proper rules, whereas $\mathscr{B}$ contains only the logical axioms and only the logical rules of inference.

Definition 11.3. Let $\mathscr{S}$ be a formal theory with the *logical basis* $\mathscr{B}$, that is, $\mathscr{B}$ is a subtheory of $\mathscr{S}$ containing only the logical axioms of $\mathscr{S}$ and only the logical rules of inference of $\mathscr{S}$. Then $\text{In}(\mathscr{S}, \mathscr{B})$, the *syntactic information of $\mathscr{S}$ relative to $\mathscr{B}$*, can be expressed as $\text{In}(\mathscr{S}, \mathscr{B}) = T_{\mathscr{S}} - T_{\mathscr{B}}$.

It is true that the question of whether an axiom or rule of inference should be regarded as logical or proper is not syntactic but rather semantic. However, $\text{In}(\mathscr{S}, \mathscr{B})$, as a function (not as a function value), does not depend on how the logical basis of a theory is determined. In that sense, that is, as a function, $\text{In}(\mathscr{S}, \mathscr{B})$ is a proper syntactic concept. This motivates the name "syntactic information of $\mathscr{S}$ relative to $\mathscr{B}$" for $\text{In}(\mathscr{S}, \mathscr{B})$.

Let us illustrate by example the semantic question of distinguishing between a logical and a proper axiom. Let $\mathscr{S}$ be a first-order theory of groups, that is, a formal theory which explains the mathematical concept of a group. For one such theory the reader is referred to [16]. There is no doubt about the classification of the axioms into logical and proper (properly mathematical), with exceptions for axioms which concern the equality relation. If these axioms are considered logical, one speaks of the group theory as a first-order theory with equality. If instead these axioms are considered proper (properly

mathematical), one speaks of the group theory as a first-order theory. Let $\mathscr{B}_=$ and $\mathscr{B}$ denote the logical bases, respectively; $\mathscr{B}_=$ contains the axioms for the equality relation, whereas $\mathscr{B}$ does not. Then $\mathrm{In}(\mathscr{S}, \mathscr{B}_=)$ will be properly included in $\mathrm{In}(\mathscr{S}, \mathscr{B})$. There is no doubt that $\mathrm{In}(\mathscr{S}, \mathscr{B}_=)$ properly informs about the group concept, whereas $\mathrm{In}(\mathscr{S}, \mathscr{B})$ also contains information that can be classified as logical as well as mathematical.

In formalizing a physical or biological theory, it may be more natural to separate a logical basis and to classify axioms (and rules) as properly physical or properly biological. There are, however, difficult questions involved here too.

11.4. EXPLICATORY POWER OF A THEORY

According to Explanation Hypothesis I, every effectively understandable explanation is a p-explanation in an r-formal theory. We want now to examine a converse problem: Given an r-formal theory, what understanding does it provide, that is, what is its explicatory power? In this general context, let us examine the following excerpt from Popper [19]:

> . . . a full understanding of a theory would mean understanding of all its logical consequences. But these are infinite in a non-trivial sense: there are infinitely many situations of infinite variety to which a theory might be applicable; that is to say, upon which some of its logical consequences may bear; and many of these situations have never been thought of; their possibility may not even have been discovered. But this means that nobody, neither its creator nor anybody who has tried to grasp it, can have a full understanding of all the possibilities inherent in a theory; which shows again that the theory, in its logical sense, is something objective and objectively existing—an object that we can study, something that we try to grasp. It is no more paradoxical to say that theories or ideas are our products and yet not fully understood by us than to say that our children are our products and yet not fully understood by us, or that honey is the product of the bee, and yet not fully understood by any bee.

The richness of applications of a theory, to which Popper refers, may certainly be considered a measure of its explicatory power. It may be difficult to define, however, and instead we want to associate the explicatory power of a formal theory $\mathscr{S}$ with its syntactic information, $\mathrm{In}(\mathscr{S}, \mathscr{B})$. Here the logical basis $\mathscr{B}$ reflects the preassigned logic which helps explain the proper theorems. Such an association will permit a discussion, on a syntactic level, of what power might be needed to explain certain phenomena (e.g., evolution). In that direction, let us formulate the following hypothesis.

Explanation Hypothesis II. With every r-formal theory $\mathscr{S}$ and with every contained logical basis $\mathscr{B}$, the nonrecursive properties of the syntactic

information $\text{In}(\mathscr{S}, \mathscr{B})$ reflect the explicatory power of $\mathscr{S}$ with respect to $\mathscr{B}$ as follows. If $\text{In}(\mathscr{S}', \mathscr{B}') \ll \text{In}(\mathscr{S}'', \mathscr{B}'')$ (cf. Def. 11.16A, p. 393), so that the reducibility function is 1-1, the explicatory power of $\mathscr{S}''$ with respect to $\mathscr{B}''$ is greater than or equal to the explicatory power of $\mathscr{S}'$ with respect to $\mathscr{B}'$. The explicatory powers are equal if also $\text{In}(\mathscr{S}'', \mathscr{B}'') \ll \text{In}(\mathscr{S}', \mathscr{B}')$. The explicatory power of $\mathscr{S}$ is the explicatory power of $\mathscr{S}$ with respect to $\varnothing$, the empty set.

A theory $\mathscr{S}$ with claims to explain a natural phenomenon usually has a logical basis $\mathscr{B}$ such that $\text{In}(\mathscr{S}, \mathscr{B})$ is an infinite set. Since every one of the proper theorems contained carries with it a p-explanation, that is, a proof (which may or may not be effectively obtainable from the theorem), we can choose to say with Popper that the theory may not be fully understood even by its creator. He simply cannot examine each of an infinitude of proofs. However, instead of having card $\text{In}(\mathscr{S}, \mathscr{B})$ measure the explicatory power of $\mathscr{S}$ with respect of $\mathscr{B}$, we want to have this power reflected by the nonrecursive properties of $\text{In}(\mathscr{S}, \mathscr{B})$.

Suppose that a certain phenomenon P has been formalized as a (e.g., biological) theory, $\mathscr{S}'$, and that we ask whether $\mathscr{S}'$ can be translated into a (e.g., physical) theory, $\mathscr{S}''$. This is possible if $\text{In}(\mathscr{S}', \mathscr{B}')$ is strongly recursively reducible to $\text{In}(\mathscr{S}'', \mathscr{B}'')$ so that the reducibility function is 1-1 (cf. Tm. 11.43A, p. 399). Then the explicatory power of $\mathscr{S}''$, as a physical theory, is at least as great as the explicatory power of $\mathscr{S}'$, as a biological theory.

In Section 11.10 we will see how the property of creativity (Def. 11.14A, p. 387) suffices to explain certain forms of the phenomenon of evolution.

11.5. PREDICTIVE POWER OF A THEORY

Let $\{h_1, h_2, \ldots, h_n\}$ be a set of experimentally obtained facts. This set can be considered the theoremhood of a formal theory $\mathscr{S}_0$ with $\{h_1, h_2, \ldots, h_n\}$ as the set of axioms, with no rules of inference, and thus with a logical basis $\mathscr{B}_0$ which is empty. Hence $\text{In}(\mathscr{S}_0, \mathscr{B}_0) = \{h_1, h_2, \ldots, h_n\}$. Such a mere listing of the experimentally obtained facts has no predictive power whatsoever. Not until we make an inference—which we cannot do without rules of inference—can the theoremhood contain a proper statement which predicts the outcome of an experiment.

If the catalog $\{h_1, h_2, \ldots, h_n\}$ is synthesized with the aid of some logical basis, $\mathscr{B}$, the result will be a theory $\mathscr{S}$ such that $\{h_1, h_2, \ldots, h_n\} \subset \text{In}(\mathscr{S}, \mathscr{B})$. Suppose that $\mathscr{S}', \mathscr{B}$ is another such systematization of $\{h_1, h_2, \ldots, h_n\}$. Suppose further that h_{n+1} and h'_{n+1} are two conflicting proper statements such that $h_{n+1} \in \text{In}(\mathscr{S}, \mathscr{B})$ and $h'_{n+1} \in \text{In}(\mathscr{S}', \mathscr{B})$, and such that neither h_{n+1}

nor h'_{n+1} belongs to $\{h_1, h_2, \ldots, h_n\}$. Then, which of the two forecasts, h_{n+1} or h'_{n+1} is the more likely?

The following reasoning will help in answering the question. A specification of the proper axioms and proper rules of inference of a theory may be regarded as a description of its proper theoremhood relative to a common logical basis, $\mathcal{B}$. The more rules concerning the generating part of the theoremhood, $\{h_1, h_2, \ldots, h_n\}$, that are discovered, the shorter can its description be, that is, the shorter will be the string of symbols spelling out the proper axioms and proper rules of inference. If, in the above case, the length of the proper axioms and proper rules of $\mathcal{S}$ is shorter than that of the proper axioms and proper rules of $\mathcal{S}'$, which has the same logical basis as $\mathcal{S}$, then h_{n+1} is more likely than h'_{n+1}, because $\mathcal{S}$ reflects more knowledge about the interrelationships among the h_i's than does $\mathcal{S}'$.

The argument supports the following hypothesis.

Explanation Hypothesis III. Let $\mathcal{S}'$ and $\mathcal{S}''$ be two formal theories with the same logical basis, both of which explain one set of experimental facts with comparatively short and few proper axioms and proper rules of inference. Then, if both $\mathcal{S}'$ and $\mathcal{S}''$ predict further experimental results (as being proper theorems or not), the theory with the simplest proper axioms and proper rules has the greater predictive power, in the sense that its predictions are more likely to agree with further experiments. The simplicity of the proper axioms and proper rules can be measured by the total length of all corresponding wffs (with a convention permitting a rule of inference to be written as a wff).

11.6. ORDER AND RANDOMNESS

Basic to the problem of detecting order and randomness is that of finding a short description of a string of symbols. The more orderly a long string is, the shorter its description can be if it utilizes the descriptive shortcuts permitted by references to the rules. The more randomized a long string is, the longer its description must be because fewer references to rules can be made; at the end the long string has to be described by itself.

The importance of the shortest description problem in connection with our general theme, that is, general systems theory, may be argued as follows. The more complex the man-made systems become, the more difficult it will be to talk of them, their behaviors, and their consequences. However, a man-made system is planned in advance of its building. If its behavior and consequences cannot be sufficiently well described, it will never be built. If the system is intended to be very complex, to describe it effectively but at great length may not suffice. It may be necessary to develop a sufficiently short effective description of the system. This can be done only if the system is orderly enough.

Let us assume a widespread knowledge of logical bases in terms of which it is possible to formalize wide classes of system behaviors. One problem is then to study how short descriptions can be generated with respect to particular bases. In what follows we will approach this problem by studying short descriptions that are effective in relation to a universal Turing machine (Def. 11.15A, p. 389).

Definition 11.4. Let x and z be words on the alphabet of a universal Turing machine, U, with code number u (cf. Def. 11.15A, p. 389). Then x is said to be a *description of z with respect to u* if U computes z from the standard initial tape expression x, that is, $\psi_u(x) = z$. Moreover, x is said to be the *shortest description of z with respect to u* if $\psi_u(x) = z$ and if, for every other y such that $\psi_u(y) = z$, we have that the length of x is shorter than the length of y or, if the lengths are equal, x comes before y in the lexicographical order defined by the alphabet. The *shortest form function $s(z, u)$* has as its value the shortest description of z with respect to u.

The problem of finding the shortest description of a long string z of symbols with respect to a universal Turing machine U, is just the problem of computing the shortest form function, $s(z, u)$. Hence, if $s(z, u)$ were recursive, we would have another Turing machine that solved our problem. In the light of the above comments about the wide applicability of the problem, it will perhaps not come as a surprise to learn that $s(z, u)$ is not recursive.

Theorem 11.2. For no universal Turing machine U is $s(z, u)$ a recursive function of z.

Proof. Assume that there is a universal Turing machine U such that $s(z, u)$ is recursive in z. Consider the function $g(x) = \mu z[s(z, u) > x]$ (cf. Def. 11.4A, p. 372). Because of our assumption, $g(x)$ is partial recursive (Def. 11.5A, p. 373). Furthermore, $g(x)$ is total, because with every x there must be a string z whose shortest description is larger than x. This follows from the obvious fact that the number of distinct shortest descriptions is as large as the number of strings z computable on U, that is, infinite. Hence $g(x)$ is recursive (Def. 11.5A, p. 373). This means (Tm. 11.33A, p. 392) that there exists an argument τ in the domain of ψ_u such that $\psi_u(\tau) = g(\tau) = \mu z[s(z, u) > \tau]$. Thus U, starting from τ, computes a z, whose shortest description with respect to the machine U is larger than τ. This contradiction shows that the above assumption is false, that is, there is no universal Turing machine U such that $s(z, u)$ is recursive in z.

Theorem 11.3. For no universal Turing machine U is there an algorithm for the determination of a z such that $s(z, u)$ is larger than an arbitrarily given x.

Proof. Assume that there is a universal Turing machine U for which there is an algorithm for the computation of $f(x)$ such that $s(f(x), u) > x$. Thus, since $f(x)$ is recursive, there is (Tm. 11.33A, p. 392) an argument τ in the domain of ψ_u such that $\psi_u(\tau) = f(\tau)$. This means that U, when started from τ, produces $f(\tau)$, whose shortest description with respect to the machine U is larger than τ. This contradiction completes the proof of Theorem 11.3.

In [12] and [14] there are more complete results concerning the nonrecursive properties of $s(z, u)$. The exposition given here will suffice, however, for the following discussion of order and randomness.

Order-randomness hypothesis. With every explicable conception of order and randomness, there is a universal Turing machine U such that of two equally long numbers (strings of symbols), z_1 and z_2, z_1 is more ordered (lawful) than z_2 when $s(z_1, u) < s(z_2, u)$, and z_1 is more randomized (lawless) than z_2 when $s(z_1, u) > s(z_2, u)$.

Let $z(v)$ stand for a string of v symbols from an alphabet of, say, n symbols. With a simple dichotomy we could say that half of these n strings are ordered and the other half randomized. According to the hypothesis we should then have $s(z_1, u) < s(z_2, u)$ as soon as z_1 belongs to the ordered $z(v)$-strings and z_2 to the randomized. A definition which more accurately reflects the order-randomness hypothesis is the following.

Definition 11.5. A string z of length v, written on an alphabet of n symbols, is *randomized to a degree r in relation to u* if $s(z, u) \geq s(w, u)$ holds for $r \cdot n^v$ of the n^v strings w of length v, where r is a rational number such that $0 < r \leq 1$.

Hence z is randomized to degree 1 if z has a shortest description that is longer than or equal to the shortest description of every one of the n^v strings of length v. On the other hand, if r is chosen close to zero, z is among the most ordered sequences.

Theorem 11.4. For no universal Turing machine and for no degree of randomization $r > 0$ is there an algorithm for the generation of a string $z(v)$, where v is an arbitrarily given length of the string, such that $z(v)$ is randomized to degree r relative to u.

Proof. Let U be a universal Turing machine with an alphabet of n symbols. Let x be an arbitrarily given number. Since r is nonzero, we can select a v such that $r \cdot n^v > n^x$. This means that, if $z(v)$ is randomized to degree r, then $s(z(v), u)$ is larger than the shortest descriptions of n^x other distinct strings. Hence $s(z(v), u)$ itself must be larger than x, and it follows from Theorem 11.3 that there is no algorithm for the computation of $z(v)$.

Another way of defining randomized sequences is based on von Mises' principle of the excluded gambling system. The reader is referred to Popper [18] for a lucid discussion of this principle. Here a sequence is accepted as randomized if it does not permit a gambler to predict the sequence so as to win. A more accurate definition of randomness is obtained if the gambler is rationalized and replaced by some universal Turing machine.

It is interesting to notice that the two concepts of random sequences, although derived from distinct basic ideas, both imply that randomized sequences cannot be generated effectively. With our definition, the reason is the nonrecursive properties of $s(z, u)$, which were first obtained in [12] and then further studied in [14].

Instead of discussing the principle of the excluded gambling system, let us here just give an example of another kind of game situation, and see how it is connected to the notion of randomization. In the game called "matching pennies," two players each put down a penny and bet as to whether the coins will show the same or different sides. If they show the same side, whether heads or tails, A wins; if they show different sides, B wins. The winning player collects the opponent's coin. Obviously, if A knows B's strategy, A can always win. Any pure strategy employed by either player must then fail (such as always playing heads, or playing heads and tails alternately). Therefore the game is "indeterministic," and "mixed" strategies come into play. Von Neumann [22] showed that A's best hope lay in changing his strategies at random, preserving, however, the overall proportion of $\frac{1}{2}$ to $\frac{1}{2}$ for heads and tails. In this way, his losses would ultimately equal his gains. For the other player the same obviously holds true. Now, because of Theorem 11.4, even if both players were as powerful as Turing machines, neither of them would be able to reach the von Neumann solution. This does not mean, however, that they could not play a balanced game. Indeed they would, if both had the same predictive power (when looked upon as rule finders). And the higher this power, the more randomlike would the two head-tail sequences be. Nevertheless, if the game were played for a sufficiently long time, the sequences would be found to be randomized not even to the smallest degree, r (i.e., to any nonzero degree, however small).

11.7. LEARNING SYSTEMS

In biology we recognize two distinct ways of obtaining a property: it is either acquired or inherited. An algorithm of Woodger [24] is intended to explain the two concepts to such an extent that one can be separated from the other.

In this section and the following ones we will study extensions of the two

ways of producing properties in a context not solely restricted to biology. Let us call these extensions of acquired and inherited characters "learned" and "programmed" characters, respectively. Also, in biology, it is customary to use the word "learning" for acquiring a property (as opposed to inheriting one). Occasionally geneticists speak of genetic programs, and therefore it seems appropriate to use the term "programming" for a generalized type of inheritance.

It should be observed that Woodger's algorithm for the distinction between an acquired and an inherited character cannot be readily applied to distinguish between learning and programming simply because it is based on the specifically biological concept of a genetical system. The following hypothesis will be helpful for a distinction in the general case.

Learning-programming hypothesis. An object A can learn from a surrounding S if A can extract order (regularities) from S. The more order that is extracted (the shorter the description of S produced by A), the more genuine is the quality of learning. The amount of work done by the learning mechanism (the order-extraction mechanism) in A represents the amount of learning done by A and is hence a subjective measure. If A obtains the properties of S without any proper order-extraction work, A is said not to have learned S but to have been programmed by S.

Let us indicate by a few examples how this hypothesis, abstract as it may appear, can explain the most common conceptions of learning as they occur in the everyday language. We often meet "learning" in sentences like "to learn what to do." Usually it is then not "what to do" which is learned. It is primarily a regularity which is extracted, and extracted to such an extent that its internal description will be short enough to yield a predictive power (cf. Explanation Hypothesis III, p. 352). An inference of what to do in order to have certain desires fulfilled can hence be made. Basically, however, it is not what is done which is learned but the underlying regularity, even though it, as well as the subsequent inference, may be wholly unconscious.

Also, learning through teaching can be fruitfully regarded as a regularity extraction. A teacher usually rewards the type of behaviors he wants to have taught. If his rewards are sufficiently attractive, his regularities are likely to be observed by the students and a quick learning may result. Again, the teacher may appeal to the primary learning drift (taste for discoveries), and reveal a difficult relation far enough to whet the appetite of the student. Because of the teacher, the amount of learning done by the student in the latter case may be far less than that done by the researcher who learned the regularity directly with nature as the teacher.

It would seem plausible that almost anything could be learned effectively with a sufficiently good teacher and a sufficiently elaborate learning mechanism. As we shall see, however, this is not the case. There are problems which cannot be learned without so much help from a teacher that the latter will function instead like a programmer.

Consider an effective system A (a machine of Turing type), intended to be universally learning in the sense that it should be able to extract every regularity from a surrounding B as soon as there is an effective explanation of B. According to the learning-programming hypothesis, the learning mechanism of A would have to compute a shortest form function (Def. 11.4, p. 353). Since, according to Tm. 11.2, p. 353, this is not recursive, it follows that A cannot be an effective system. However, as we shall see in a later section, with every B there is associated a self-describing B^*. Obviously there are systems A, and indeed very simple systems, which can "learn" from B^* simply by copying the self-describing part of B^* (to obtain predictive power). But, according to the learning-programming hypothesis, this type of behavior is programming, not learning. Moreover, concerning programming, we know from Tm. 11.27A, p. 390, that there are universal Turing machines which can be programmed to behave like any other Turing machine.

There are several alternatives to the question of how to measure the amount of work done by a learning mechanism. One way is to measure the work in terms of the complexity of an irredundant computer for doing the work. If the computer is sequential, obviously the computation time will also be a factor. Another way would be to measure the work by the predictive power produced by the learning mechanism (Explanation Hypothesis III, p. 352) or by the length of a diagnosing program for the identification of the surrounding as a machine.

11.8. PRODUCTIVE AND REPRODUCTIVE SYSTEMS

In this section and the two that follow we will examine certain phenomena in the general area of productive, reproductive, and evolutionary systems. Concepts like these have their strongholds in biology, where they rarely need to be defined but can instead be directly demonstrated by reference to our most common surroundings. Now, when we intend to embed the concepts in an interdisciplinary systems theory, it becomes necessary to supply definitions. Let us first see, however, how these concepts are used in a biological context. In [7] Dobzhansky, writes as follows about reproduction, self-reproduction, and evolution.

Every species of organism reproduces itself (p. 23).

Heredity is a conservative force. If children and parents were completely identical, evolution could not occur. Heredity, however, is opposed by a process of change, variability. Self-reproduction occasionally results in an imperfect copy of the parental living unit, and the altered copy, called a mutant, then reproduces the altered structure until new mutations intervene (p. 23).

The evolutionary development of the living world has, on the whole, led from simple to more complex forms of life. It is reasonable to suppose that this progression from the simple to the complex was accompanied by an increase in the number of genes which a species carries. Duplication and polyploidy are the only known methods whereby such increase could occur, since the appearance of self-reproducing genes from non-self-reproducing cell structures seems improbable (p. 66).

It would seem that Dobzhansky uses "self-reproduction" and "reproduction" in essentially the same sense, namely, as expressions for the property of producing an eventually imperfect copy of a parental unit.

Let us first see how the underlying concepts, production and productive, can be examined with reference to a logical basis. When can we say that an object (automaton, organism) is productive?

Any property of an automaton A will have to be defined in relation to the properties of its surrounding S. If S had no properties whatsoever, A could be said to have properties relative to A, which, if explicable, would be explained only to A itself and hence would be of no use to us unless we were included in A. We do not want to make such a strong assumption, namely, to consider ourselves as automata. Rather, we want in Section 11.11 to discuss the so-called reduction problem without having made any prejudicial assumptions.

Definition 11.6. An object A *is productive in a surrounding* S if A causes S to produce another entity B, symbolized by $A \to d \underset{S}{\rightarrow} B$ and read as follows: The configuration (output state) d of A forces S to produce B. Here d can be considered a description of B relative to S.

With a liberal interpretation of this definition, we could say that A is productive in S as soon as A has some influence on its surrounding, S. The reason for the actual formulation of Definition 11.6 may be clarified by the following discussion.

Suppose that A is an automaton which is productive in its surrounding S, of which we are part, so that A causes S to produce an entity B which is a number, say two. Then A is capable of taking on an output state d, a materialistic configuration like, for example, the sign 2 written on paper. This d forces the surrounding (us) to conceive of it as the number two (2 forces us to produce the number two in our minds).

In general, let d be a description which has a unique effect on a surrounding

S, such that, for example, d forces S to produce B. Then it is very natural to say, as we have suggested in Definition 11.6, that A, which generates d as an output state, is the cause of B in S, that is, that A produces B in S, or $A \to d \underset{S}{\to} B$.

If A is productive in S, one of two things will happen: either the produced B is again productive in S, or it is not. In the latter case B has no further effects on S.

For example, if B is a number, and this number is neutral with respect to the properties of S, then B is nonreproductive in S. On the other hand, if B is productive in S, we will have the same two alternatives concerning the object C which is produced by B in S, and so on. It is this property of producing offspring that are again productive, and so on, that we want to refer to as being reproductive.

Definition 11.7. An object A *is reproductive in a surrounding* S if there are objects A_i with descriptions d_i such that

$$A \to d_1 \underset{S}{\to} A_1 \quad \text{and} \quad A_i \to d_{i+1} \underset{S}{\to} A_{i+1} \quad \text{for } i = 1, 2, 3, \dots .$$

(An object A *is n-productive in a surrounding* S if the i-index in the above stipulation is delimited to the set $\{1, 2, \dots, n-1\}$.)

Eventually, this definition of reproduction may be considered an abstract form of the following biological definition, found in Henderson [8]: " Reproduction: Continuation of species or race, sexually or through cell rupture, cell-division, budding, spore-formation, conjugation, or parthenogenesis."

In Definitions 11.6 and 11.7 we have the notion of a description d of B in a surrounding S. As explained, this means that d forces S to produce B. Again, this is possible only if S has some properties.

By way of example, let us consider a productive system which produces Turing machines. Previously we have considered certain code numbers, z, for Turing machines (Def. 11.9A, p. 378). Such a code number z is indeed a description of a Turing machine Z in relation to us, who can effectively decode z into a list of quadruples and, furthermore, effectively build a machine Z which has all the characteristics implied by z. (The problem with the potentially infinite tape may be overcome by using a tape-adding mechanism which is influenced by the computation of Z.) The point is that all the decoding and building steps are effective and could be done as well by a machine, M, which operates in a surrounding, S, which is far more simple than the one which contains ourselves (as decoders and builders). In addition to M, S may just contain the types of building blocks which suffice for constructing any specific Turing machine. The properties of these building blocks constitute the property of S which permits a productive system A to produce any Turing machine, Z_i, that is, $A \to d_i \underset{S}{\to} Z_i$. Notice that here the descriptions d_i,

which are effective in relation to the surrounding S, no longer need be the code numbers z_i, which were effective in relation to ourselves. The d_i-descriptions are (sequences of) output states from A which directly force S to produce the Turing machines Z_i. Such a surrounding S will be said to be complete with respect to the class of Turing machines.

Definition 11.8. A surrounding S is said to be *complete with respect to a class C of constructs* if, with each $e \in C$, there is a description d which forces S to produce e.

Definition 11.9. S_{TM} is a surrounding which is complete with respect to the class of Turing machines, so that the description $c(z)$ forces S_{TM} to produce Turing machine Z. Here, z is the code number of Z, and the code function c is recursive.

The reader is referred to von Neumann [23] for discussions of various S_{TM}-surroundings.

11.9. SELF-PRODUCTIVE SYSTEMS

Let us examine the word "self-reproductive" (cf. the excerpt from Dobzhansky in Section 11.8) from a logical point of view. There are two distinct but natural interpretations. One is to say that A is self-reproductive if, first of all, A is reproductive and, second, A reproduces a copy of itself. The other interpretation is that A is self-reproductive if A is reproductive by itself, that is, A is reproductive in a surrounding S with no properties. In accordance with our previous comments, we will not be concerned with such surroundings here. Instead the interested reader is referred to [13] for a study of this more abstract form of self-reproduction, there called complete self-reproduction.

In this chapter we will be concerned only with the first interpretation, namely, that A is self-reproductive in S if A is reproductive in S and produces a copy of itself in S. Notice that we can equivalently say that A is self-reproductive in S if A produces a copy of itself in S, for if A produces A in a surrounding S, we have that there exist objects A_i, namely, $A_i = A$, such that

$$A \to d_1 \xrightarrow{S} A_1 \quad \text{and} \quad A_i \to d_{i+1} \xrightarrow{S} A_{i+1} \quad \text{for} \quad i = 1, 2, \ldots,.$$

Thus A is reproductive according to Definition 11.7. Hence a logically irredundant notation for this phenomenon is that A is self-productive (instead of self-reproductive).

Definition 11.10. An object A *is self-productive* (*self-reproductive*) *in a surrounding S* if A produces a copy of itself in S.

Corollary 11.1. If A is self-productive in S, then A is reproductive in S.

In biology, various forms of self-production are known. Let us here demonstrate that these phenomena can also be effectively explained so as to permit us to build self-producing, although nonliving, automata.

Theorem 11.5. There exists a Turing machine which is self-productive in an S_{TM}-surrounding (cf. Def. 11.9, p. 360).

Proof. According to Definition 11.9, it suffices to prove that there exists a Turing machine E, with a code number e, such that E computes its own description, $c(e)$, relative to S_{TM}. The existence of such an E follows easily by applying the Kleene recursion theorem (Tm. 11.30A, p. 391) to the recursive function $g(y, x) = c(U_1{}^2(y, x)) (= c(y)$; cf. Def. 11.5A, p. 372). We then obtain an e such that $\psi_e(x) = g(e, x) = c(e)$. This e is the code number of a Turing machine E which computes description $c(e)$ such that $c(e)$ forces S_{TM} to produce another Turing machine E.

In [23] von Neumann gives a detailed study of automata which are self-productive in a so-called tessellation surrounding. He discusses various connected problems of mathematical logic (but not explicitly the recursion theorem). In [11] Lee presents a proof of the existence of a self-describing Turing machine. Lee's proof could have been simplified by a direct use of the recursion theorem. In fact, the power of this theorem is so great that the short proof it yields for Theorem 11.5 may give the impression that the existence of a self-producing Turing machine is not very noteworthy. However, if one tries to modify the quadruples of a Turing machine to have it compute its own code number, it is a surprise to learn that such a machine really exists. Furthermore, the proof given for the recursion theorem is constructive in the sense that it indicates how such a self-producing machine can be constructed, given the properties of an S_{TM}-surrounding.

Theorem 11.6. There exists a pair of *symbiotically self-producing* distinct Turing machines, M and N, such that M produces N and N produces M in an S_{TM}-surrounding.

Proof. According to Def. 11.9, p. 360, it suffices to prove that there exists a pair of distinct Turing machines, M and N, with code numbers m and n, respectively, such that $\psi_m(x) = c(n)$ and $\psi_n(x) = c(m)$, where c is the recursive S_{TM}-code (cf. Definition 11.9). By the Kleene projection theorem (Tm. 11.15A, p. 384), there is a recursive function $g(y)$ such that $\psi_{g(y)}(x) = c(y)$ and $c(g(y)) > c(y)$. [With the notations of Theorem 11.15A we would have

$\psi_{z_0}(y, x) = c(y) = \psi_{\varphi(z_0, y)}(x)$, that is, $g(y) = \varphi(z_0, y)$; furthermore, we have chosen $c_1(x) = c_2(x) = c(x)$.] Hence we have both $\psi_{g(y)}(x) = c(y)$ and $\psi_{g(g(y))}(x) = c(g(y))$. The question is now whether we can find a particular y such that, not necessarily $g(g(y)) = y$, but $\psi_{g(g(y))}(x) = \psi_y(x)$. We can, because Tm. 11.31A, p. 392 (the second form of the recursion theorem) ensures a number n such that $\psi_{g(g(n))}(x) = \psi_n(x)$ [the function $g(g(y))$ is recursive in y according to Tm. 11.1A, p. 374]. With m for $g(n)$ we thus have a pair of numbers, m and n, such that $\psi_m(x) = c(n)$ and $\psi_n(x) = c(m)$. Furthermore, since $c(g(y)) > c(y)$, we have that $c(m) > c(n)$. Consequently the machines N and M, which compute $c(m)$ and $c(n)$, respectively, must be distinct. This completes the proof of Theorem 11.6.

In the case of a pair of symbiotically self-producing machines, M and N, we have a self-production cycle of length two; M produces N, which produces another M. The proof of Theorem 11.6 can obviously be modified so as to give the existence of a self-production cycle of any finite length.

Besides being self-productive, a machine M can also perform some other function. It may, for example, compute a function $f(x)$. Then M computes both $f(x)$ and its description $c(m)$ relative to S_{TM}. If the value $f(x)$ is not supposed to be interpreted by S_{TM} as a description of a machine, $f(x)$ may have to be kept apart from the produced description, $c(m)$. Let $\tau(x, y)$ be a corresponding recursive pair function which is effective in relation to S_{TM} so that $\tau(x, y)$ is interpreted by S_{TM} as a neutral value x and a description y of a machine.

Theorem 11.7. With every partial recursive function $f(x)$ there is a self-producing Turing machine which computes $f(x)$.

Proof. By the above comments it suffices to prove that there is a Turing machine M, with description $c(m)$, such that $\psi_m(x) = \tau(f(x), c(m))$. For any partial recursive $f(x)$, $\tau(f(x), c(y))$ is partial recursive in x and y. Hence, by the recursion theorem (Tm. 11.30A, p. 391), there is a number m such that $\psi_m(x) = \tau(f(x), c(m))$. This completes the proof of Theorem 11.7.

11.10. EVOLUTIONARY SYSTEMS

Mayr [15] and other eminent biologists have called the theory of evolution the greatest unifying theory in biology. With the unifying property of general systems theory in mind, we ask whether, in turn, there are systems theories with an explanatory power great enough to permit effective explanations of some theories of evolution as they occur in philosophy, biology, and other sciences.

When talking about evolution in general, it may be appropriate to distinguish between two types: evolution by reproduction and evolution by production. Biological evolution (cf. the excerpt from Dobzhansky in Section 11.8) exemplifies evolution by reproduction. Evolution is here carried on by biological individuals with a life span which is short compared to the duration of the evolution. Some individuals are reproductive, however, and this is how evolution carries on: by successively reproducing new generations. If there are no truly reproducing individuals, an n-productive (Def. 11.7, p. 359) individual and its successive generations may account for an evolution with a duration of n generations.

The evolution of a theory may be taken as an example of evolution by production. Usually the main body of the theory is never completely destroyed, but functions as a productive center around which new hypotheses are generated, examined, and eventually accepted as new axioms. Then it is the whole production phase that carries the evolution.

We will not go into the question of whether there is an effective way of distinguishing between evolution by reproduction and evolution by production. It may well happen that evolution by reproduction can be considered a special case of evolution by production. Let it suffice here to say that biological evolution by reproduction is probably the best documented of all evolutionary phenomena. Hence we could not claim to have a formal theory explain a typical phenomenon of evolution if it cannot explain an evolution by reproduction. This is the only type which will be considered here.

So far we have been able to demonstrate the existence of a very special kind of reproduction, namely, self-production (self-reproduction). If the successive generations of an individual all look the same, they certainly cannot be said to be the carriers of an evolution. A kind of change is invariably associated with evolution. Evolution by reproduction requires a nonperiodic change in the series of new generations. Let us ask whether it is possible to explain effectively an individual with the property not only to be productive, but also to have its offspring productive so that all succeeding generations are also productive and, furthermore, distinct, each from the other. At first glance, the difficulties involved in constructing such an individual seem insurmountable. However, the recursion theorem will help us out here, too.

Let us consider Turing machines that are productive in an S_{TM}-surrounding. Then, what we are looking for is an infinite series of code numbers, $z_1, z_2,$ $z_3, \ldots,$ such that $\psi_{z_1}(x) = c(z_2), \psi_{z_2}(x) = c(z_3), \psi_{z_3}(x) = c(z_4), \ldots,$ and such that the machines Z_i are all distinct. If this series $Z_1, Z_2, Z_3, \ldots$ is going to be explicable, there must be a recursive function $g(i)$, such that $g(i) = z_i,$ that is, $\psi_{g(y)}(x) = c(g(S(y)))$ for all x and $y(S(y)$ is the successor function of Def. 11.5A, p. 373). In order to prove that a solution exists for this partial evolution equation, we need the following theorem.

Theorem 11.8. With every recursive function $c(x)$ with a nonfinite range, there is a recursive function $f(z, y)$ such that $\psi_{f(z,y)}(x) = c(\psi_z(S(y)))$ and $c(f(z, y)) > y$.

Proof. By Tm. 11.27A, p. 390, $h(x, y, z) = {}_{\text{def}} c(\psi_z(S(y)))$ is partial recursive in x, y, z. Hence there is a code number e such that $\psi_e(x, y, z) = c(\psi_z(S(y)))$. Choosing the functions c_1 and c_2 of Tm. 11.15A, p. 348 as c and the identity function, respectively, we obtain from Theorem 11.15A a recursive function $\varphi(e, y, z)$ such that $\psi_{\varphi(e, y, z)}(x) = c(\psi_z(S(y)))$ and such that $c(\varphi(e, y, z)) > y$. Hence $f(z, y) = {}_{\text{def}} \varphi(e, y, z)$ is the desired recursive function.

We can now prove the following theorem.

Theorem 11.9. There is a recursive function $g(y)$ which satisfies the partial evolution equation, $\psi_{g(y)}(x) = c(g(S(y)))$, so that no two of the machines produced, $G_0, G_1, G_2, \ldots$, are alike [machine G_{i+1} has code number $g(i + 1)$; the description $c(g(i + 1))$, which forces S_{TM} to produce G_{i+1}, is produced by machine G_i].

Proof. Applying the recursion theorem (Tm. 11.30A, p. 391) to the function $f(z, y)$ of Theorem 11.8, we get a number m such that $f(m, y) = \psi_m(y)$. Hence according to Theorem 11.8, $\psi_{\psi_m(y)}(x) = \psi_{f(m, y)}(x) = c(\psi_m(S(y)))$. Thus there is a $g(y) = {}_{\text{def}} \psi_m(y)$ such that $\psi_{g(y)}(x) = c(g(S(y)))$ and $c(g(y)) = c(f(m, y)) > y$. If two machines, M_i and M_j $(i \neq j)$, in the produced machine sequence $G_0, G_1, G_2, \ldots$ were alike, they would produce similar offspring and the machine sequence would be periodic. Hence there could be only a limited number of distinct offspring. This contradicts $c(g(y)) > y$, which implies that an unbounded number of distinct descriptions $c(g(i))$ is produced by G_j-machines. Hence there must be also an unbounded number of distinct G_j-machines, that is, no two machines in the sequence can be alike.

Theorem 11.9 with its proof can be interpreted as describing an effective construction of a Turing machine G_0 which in S_{TM} will be the ancestor of a forthcoming steadily changing sequence of reproducing offspring. Whether or not this sequence should be termed evolutionary depends on what further functions the machines G_i may have. None has been specified so far. However, we could say that every machine that would give rise to an effectively explicable evolutionary series of new generations in S_{TM} would have to correspond to a recursive g-solution to the partial evolution equation of Theorem 11.9.

Let us now modify the evolution equation so that it will account for not only a changing reproductive structure, but also a normal behavior for each individual of the evolving sequence. With reference to an S_{TM}-surrounding, this is done with the aid of a pair function, $\tau(y, x)$, whose value is interpreted by S_{TM} partly as a neutral object x and partly as a description y. Thus, if

there is a function g such that $\psi_{g(y)}(x) = \tau(c(g(S(y))), f(y, x))$, we may interpret it as follows: $g(y)$ is the code number of a Turing machine, G_y, which partly produces another Turing machine, G_{y+1}, with code number $g(y + 1)$ and partly computes the function $f(y, x)$. Again, G_{y+1} is reproductive and computes the function $f(y + 1, x)$, and so forth.

Theorem 11.10. With every partial recursive function (normal behavior) $f(y, x)$ there exists a recursive function $g(y)$ which solves the partial evolution equation $\psi_{g(y)}(x) = \tau(c(g(S(y))), f(y, x)), c(g(y)) > y$.

Proof. Upon observing that $\tau(c(\psi_z(S(y))), f(y, x))$ is partial recursive in x, y, z, we can continue just as with the proof of Theorem 11.9. There exists an e such that $\psi_e(x, y, z) = \tau(c(\psi_z(S(y))), f(y, x)) = \psi_{\varphi(e, y, z)}(x)$, where, according to Tm. 11.15A, p. 384, $h(y, z) = {}_{\text{def}} \varphi(e, y, z)$ is recursive and $c(h(y, z)) > y$. Applying Kleene's recursion theorem to $h(y, z)$, we conclude that there is an m such that $h(y, m) = \psi_m(y)$. With $g(y)$ for the recursive function $h(y, m)$ we thus have

$$\psi_{g(y)}(x) = \tau(c(\psi_m(S(y))), f(y, x)) = \tau(c(g(S(y))), f(y, x))$$

and $c(g(y)) > y$, which completes the proof of Theorem 11.10.

Comparing Theorem 11.10 with Tm. 11.7, p. 362, we see that, in regard to normal behaviors, a self-producing machine can do just as much as an evolving reproductive series of machines. It is true that the normal function of Theorem 11.7 is specified with only one variable, whereas in Theorem 11.10 the first variable of the normal function is connected with the generation number so as to account for a change in the normal behavior as well. Such a change can also be accomplished with a pure self-producing structure by having the first machine, M, compute the normal function $f(a)$. Furthermore, the second generation M, which also computes the function f, should now compute on the result $f(a)$ of the previous generation. Thus $f(f(a))$ will be computed, and, at the nth generation, $f^n(a)$.

In [17] Myhill develops a theory for this kind of evolution by self-production. Concerning the choice of the f-function, Myhill chooses $f^n(a)$ to be the description of a nonproductive automaton which enumerates the theorems of a theory T_n such that, for each n, T_{n+1} is richer in content than T_n. In this way he meets with a criterion, namely, the ability to solve a new problem, which is often stipulated for an evolutionary process.

The difference between evolution by self-production (cf. Myhill [17]) and evolution according to the partial evolution equation of Theorem 11.10 may be exemplified as follows. In an evolution by self-production it is not the self-producing structure that evolves, but instead the normal behavior of this structure. Although this new generation computes the same function

as a previous one, it computes on a new value, namely, that produced by the previous generation. Hence this type of evolution can be called *evolution of behavior*, whereas evolution according to Theorem 11.10 is *evolution of structure* (and at the same time evolution of behavior). In biology evolution of behavior may be quite rapid, as when one generation of *Homo sapiens* works on the results of a previous generation. The structural evolution of the species *Homo sapiens*, however, is comparatively very slow.

Finally, let us see how the normal function $f(y, x)$ of Theorem 11.10 can be determined so as to exhibit a specific evolution character. As we know (Tm. 11.41A, p. 396), the theoremhood of an r-formal theory is always recursively enumerable. Hence, with every such theory $\mathscr{S}$ there is a code number z such that $S_{[z]}$ (cf. Def. 11.14A, p. 387) is the theoremhood, say T_z, of the theory $\mathscr{S}$. Again, $S_{[z]}$, that is, T_z, is enumerated by a recursive function $\psi_{\varphi(z)}(x)$ such that $\varphi(z)$ is recursive (Tm. 11.18A, p. 386). What will happen if $f(y, x)$ is chosen as the description of a Turing machine which enumerates the theorems of T_y, that is, $f(y, x) = c(\varphi(y))$? According to Theorem 11.10, the machine G_y will produce both a reproductive machine, G_{y+1}, and the eventually nonreproductive machine, Z_y, which enumerates the theorems of T_y. In turn, G_{y+1} produces G_{y+2} and Z_{y+1}, which enumerates T_{y+1}, and so on.

Next, let us see whether it is possible to specify a recursive function $f(y, x)$ such that a specific sequence of enumerating machines will be produced. If the enumerating machine Z generates only "true" statements, the Z-machine of the next generation enumerates further new "true" statements, and so on, then the sequence of reproducing G-machines could be said to be evolutionary.

Let Q be a set of all true statements of a certain form. Such sets are usually productive (cf. Ex. 11.10A, p. 401). For example, let the predicate $P(x)$ be defined as follows: "When started on the argument x, Turing machine X, with code number x, never halts." Then the set of all true statements of the form $P(x)$ is productive. This follows from the fact that the set

$$K = \{x : \exists y T(x, x, y)\}$$

is creative (cf. Tm. 11.24A, p. 388), that is, $Q = \bar{K} = \{x : \neg \exists y T(x, x, y)\}$ is productive (cf. Def. 11.14A, p. 387). From the productiveness of the set of truths, Q, it follows that there is a partial recursive function $h(x)$ such that

$$[T_z \subset Q] \Rightarrow [h(z) \in Q] \ \& \ [h(z) \notin T_z].$$

By Tm. 11.16A, p. 385, there is a recursive function $r(z)$ such that

$$\exists y [T(z, x, y) \lor x = h(z)] \equiv \exists y T(r(z), x, y).$$

Hence $T_{r(z)} = T_z \cup \{h(z)\}$ is a recursively enumerable set (a theoremhood of an r-formal theory) which contains only true statements (Q-elements) and

one new true statement relative to T_z. Since Q is productive, we can repeat this process indefinitely without exhausting all true statements of Q. In particular, $T_{r(r(z))}$ contains only true statements and one which is new relative to $T_{r(z)}$.

Hence let us define $f(y, x) =_{\text{def}} c(\varphi(r^y(z)))$. Here $r^y(z)$ is the yth composition of r, obtained by primitive recursion (Def. 11.3A, p. 372) from r and hence recursive (Def. 11.5A, p. 373). Then the following evolution characteristic will result. The behavior of the machine G_y, produced at the yth generation, will be partly to produce a reproductive machine, G_{y+1}, and partly to produce the eventually nonreproductive machine, Z_y, with description $c(\varphi(r^y(z)))$. Here Z_y enumerates all truths which have been enumerated by previous generations and, in addition, a truth which has never been revealed before.

Thus we have been able to give an effective description (rather, an outline which, however, can be completed to the level of a blueprint) of a machine G_0, such that G_0 in an S_{TM}-surrounding will be the ancestor of an evolving sequence of offspring. For the description, we have used a property of a creative set, K. Indeed, there are r-formal theories with high explicatory power, so-called creative theories, which are semicomplete with respect to the creative sets (cf. Def. 11.25 A, p. 400). However, in predicting the behavior of this effectively described machine G_0, we have gone still further with the set of truths, $Q(=\bar{K})$. This set, being productive (i.e., not recursively enumerable), cannot be defined as the theoremhood of any r-formal theory. Indeed, the set of all truths ever enumerated by the offspring of G_0, namely,

$$T = \{w \colon \exists y \, \exists v \, T(r^y(z), w, v)\},$$

is recursively enumerable and cannot exhaust all truths of Q.

In light of this general systems discussion (i.e., metadiscussion of, e.g., the set Q), we may well look upon G_0 as a machine which is programmed for evolution (cf. the discussion of Myhill [17]).

11.11. COMMENTS CONCERNING GENERAL SYSTEMS THEORY AND THE REDUCTION PROBLEM

The problem of whether one science, for example, biology, can be reduced to another, such as physics, is known as the reduction problem. The meaning of "reduction" will obviously have to be informal as long as the sciences are informal. On the other hand, if the sciences are formal theories, we suggest that the reduction problem can be formalized as the question of whether one of the theories can be translated into the other in the technical meaning of Def. 11.24, p. 399.

The effects of a general systems theory on the reduction problem may be of various kinds. First, a general systems theory, if sufficiently developed with mathematical logic, may contribute to a widespread formalization of scientific disciplines. In that sense, it will make the reduction problem, at least for sufficiently small disciplines, meaningful.

Second, the interdisciplinary character of a general systems theory is likely to reduce the borders between the sciences. As a result of the systems theory, a conflict between some axioms of two sciences may be revealed. If both are natural sciences, the effect of such a discovery is likely to be a change in the theories such that they may be united by a systems theory.

The more sciences a general systems theory unites, that is, reduces within its explicatory power, the likelier it is that its predictive power will be small unless it develops into a special general systems theory. In this context, the remark by Klir [10] that "no permanent scientific explanation can be made only on the basis of empirical data" is relevant. Indeed, it reflects the importance of using a carefully selected logical basis to obtain both explicatory power and predictive power from a set of experimental facts (cf. Explanation Hypotheses II, p. 350, and III, p. 352).

One area where the reduction concept appears particularly interesting is on that border between automata theory and biology where evolutionary phenomena are discussed. As we have seen in Section 11.10 automata aspects suggest that evolutionlike phenomena can be programmed in the code which transmits structural information from one generation of automata to the next. Are there biological theories of evolution with explicatory powers such that phenomena of this type can be translated into them?

It would seem that recent biological findings concerning repair enzymes in procaryotic cells may account for one such translation. Accordingly, DNA-strands can be encoded to direct the production of an enzyme which to a certain extent repairs DNA-errors. Such a strand can be said to contain a partial program of evolution in the following sense. The repair enzyme cannot possibly repair all errors, but will take care of some; hence they may be classified as "undesirable" errors. Other errors will pass and among them will be the "desirable" mutations. Therefore, due to the DNA-program, the probability of certain changes (mutations) will be higher than the probability of others, and we can speak of a partially programmed evolution.

ACKNOWLEDGMENT

APPENDIX I. EFFECTIVE COMPUTABILITY

Among the prerequisites from mathematical logic which are developed here, formal systems play a central role. Since formalizability is highly dependent on the idea of computability, and because we also need many particular computability results, it seems appropriate to present separate appendices—the first one on effective computability and the second on formal systems.

In regard to effective computability, let us emphasize that a description (instruction) is usually considered effective if it can be executed by a machine. Man himself is then not to be considered, a priori, as a machine. A machine is a perfect deterministic, materialistic device. By hypothesis, a description is effective in an absolute sense, if computable by a so-called Turing machine.

I.1. TURING MACHINES

Abstracting from all powers of man except his ability to compute, Turing [21] has argued that any computation done by a human being could as well be performed on a Turing machine. This remarkably simple device is supplied with a tape, potentially infinite in both directions and divided throughout into squares. Each square may be blank or have printed on it a symbol out of a finite alphabet $\{s_1, s_2, \ldots, s_m\}$. The tape passes through the machine in such a way that in a given situation the machine scans just one square. The machine is capable of being in any one state of a finite set of states $\{q_0, q_1, \ldots, q_n\}$.

Scanning a square, the machine determines its next move from its present state, q_i, and the scanned symbol, s_j. Depending on $\langle q_i, s_j \rangle$, there are four possibilities: a halt in operation, a change of the symbol being scanned, or a change (movement) of the scanned square to the right or to the left. With the exception of the halt, of course, the moves are followed by a specified change of the machine state.

The behavior of a Turing machine can therefore be completely determined by a finite list of quadruples of the form $\langle q_i, s_j, s_k, q_l \rangle$ or $\langle q_i, s_j, R, q_l \rangle$ or $\langle q_i, s_j, L, q_l \rangle$, such that no two quadruples in the list begin with the same pair $\langle q_i, s_j \rangle$. If the machine is in state q_i, and scans the symbol s_j, it prints the symbol s_k and changes its state to q_l if the list contains the quadruple $\langle q_i, s_j, s_k, q_l \rangle$. If instead there is a quadruple $\langle q_i, s_j, R, q_l \rangle$ ($\langle q_i, s_j, L, q_l \rangle$), the machine moves one square to the right (left) and changes its state to q_l. If there is no quadruple beginning with $\langle q_i, s_j \rangle$, the machine stops in state q_i, scanning the symbol s_j.

I.2. COMPUTABLE FUNCTIONS

In regard to computability, there is no loss of generality to assume that the symbol alphabet is binary $\{B, 1\}$, where B stands for a blank square.

Definition 11.1A. The *function* $\psi_z(x_1, x_2, \ldots, x_n)$ *computed by a Turing machine* Z is defined as follows: $\psi_z(m_1, m_2, \ldots, m_n) = m$ if and only if Z, being in state q_0 and scanning the leftmost 1 of the tape expression,*

$$1^{m_1+1}B1^{m_2+1}B\ldots B1^{m_n+1},$$

proceeds according to its quadruples and finally halts with a tape containing m 1's. If instead Z never halts within a finite number of operations when beginning its computation from the argument $\langle m_1, m_2, \ldots, m_n \rangle$, this argument is said not to belong to the domain of ψ_z. A function $f(x_1, x_2, \ldots, x_n)$ is *partially computable* if there exists a Turing machine Z such that

$$f(x_1, x_2, \ldots, x_n) = \psi_z(x_1, x_2, \ldots, x_n).$$

If, in addition, $f(x_1, x_2, \ldots, x_n)$ is a *total* function, that is, is defined for all n-tuples $\langle x_1, x_2, \ldots, x_n \rangle$, it is called *computable*.

It follows that in the process of computing a function for a well-defined argument in its domain, the number of 1's printed on the tape will always be finite.

It should be pointed out that Turing [21] himself did not define computability precisely as in Definition 11.1A. He was interested primarily in computing decimal point expansions of numbers like the transcendentals π and e, which he proved computable. He therefore talked about a machine as computing a sequence if it did not stop.

The exposition chosen here, that of Davis [5], is more natural than Turing's when dealing with the computability of functions with both arguments and values defined in the set of integers, $N = \{0, 1, 2, 3, \ldots\}$. All functions dealt with here are of this type.

Example 11.1A. The Turing machine which is defined by the following list of quadruples:

$$\langle q_0, 1, B, q_0 \rangle, \langle q_0, B, R, q_1 \rangle, \langle q_1, 1, R, q_1 \rangle, \langle q_1, B, R, q_2 \rangle, \langle q_2, 1, B, q_2 \rangle$$

* 1^m symbolizes a sequence of m consecutive 1's, that is, the whole tape expression is

A reason for adding an extra 1 to each block of m 1's becomes apparent upon considering tape expressions for the number 0 and the pair $\langle 0, 0 \rangle$.

computes the function $f(x, y) = x + y$. If the machine starts from the argument $\langle 1, 2 \rangle$, for example, it will perform as follows:

q_0

| | 1 | 1 | | 1 | 1 | 1 | | |

q_1

| | | 1 | | 1 | 1 | 1 | |

q_0

| | | 1 | | 1 | 1 | 1 | | |

q_2

| | | 1 | | 1 | 1 | 1 | |

q_1

| | | 1 | | 1 | 1 | 1 | | |

q_2

| | | 1 | | | 1 | 1 | |
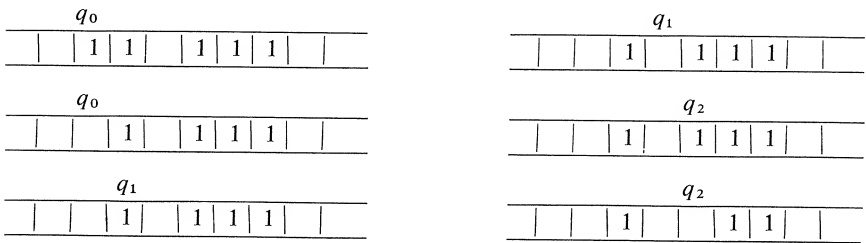
Since there is no quadruple in the list, beginning with the pair $\langle q_2, B \rangle$, the machine halts after having erased two 1's, thus producing $(1 + 2 =)$ three 1's on the tape.

I.3. TURING'S AND CHURCH'S HYPOTHESES

In 1936 Turing [21] and Church [2] advanced hypotheses concerning the nature of effective computability. In the following we have slightly reformulated Turing's hypothesis which was originally stated in terms of computable numbers.

Turing's hypothesis. Every function which can naturally be regarded as computable can be computed by a Turing machine.

Church's hypothesis. Every effectively calculable function is recursive.

The recursive functions will be defined in the next section. As proved by Turing [21], these two hypotheses are equivalent in the sense that the total computable functions are equivalent to the recursive functions (compare also Tm 11.7A, p. 378).

The hypotheses themselves cannot be proved true because of the lack of precise meaning of "naturally be regarded as computable" or "effectively calculable." They may be motivated, however, and Turing [21] advances the following arguments. The behavior of a human computer at any moment is determined by the symbols he is observing and by his "state of mind" at the moment. The number of symbols which he can recognize is finite, for if we were to allow an infinity of symbols there would be symbols differing to an arbitrarily small extent. Again, if we admitted an infinity of states of mind, some of them would be arbitrarily close and would be confused. As we are considering computation by a preassigned method, the possible states

of mind are fixed in advance of naming the particular arguments and we do not allow mathematical invention in the midst of the computer's performance.

These limitations on the behavior of the human computer in the act of computing the value of a number-theoretic function of given arguments are of the same kind as enter in a Turing machine. The machine state corresponds to the computer's state of mind, and the tape to the symbol space (sheets of paper) of the computer.

The human computer may appear less restrictive in behavior than the machine in the following senses. First, he can observe more than one symbol occurrence at a time. Next, he can perform more complicated atomic acts than the machine, and his symbol space need not be a one-dimensional tape. Furthermore, he can choose some other symbolic representation of the arguments and function values than that used in the above definition of a Turing machine. Turing convincingly argues, however, that each of these different modes of operation can be effectively reduced to Turing machine operations. On that point the reader is referred to Turing [21] or to Kleene [9].

I.4. RECURSIVE FUNCTIONS

There is an entirely different way of introducing the computable functions by means of a regular definition of a set of initial functions which, together with a couple of operations, generate all computable functions. When introduced in this way, the functions are called recursive instead of computable.

Although the two operations, composition and minimalization, are sufficient to generate the recursive functions, we will also indicate another operation, primitive recursion, which preserves recursiveness.

Definition 11.2A. The operation of *composition* associates with the functions $f(y_1, y_2, \ldots, y_m)$, $g_1(x)$, $g_2(x)$, $\ldots$, $g_m(x)$, the function

$$h(x) = f(g_1(x), g_2(x), \ldots, g_m(x)).$$

This function $h(x)$ is defined for precisely those x in the domain of each g-function for which $\langle g_1(x), g_2(x), \ldots, g_m(x) \rangle$ is in the domain of f. For the variable x may be substituted an n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$.

Definition 11.3A. The operation of *primitive recursion* associates with the total functions $f(x)$, $g(z, y, x)$ the function $h(z, x)$, such that $h(0, x) = f(x)$, $h(z + 1, x) = g(z, h(z, x), x)$. For the variable x may be substituted an n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$.

Definition 11.4A. The operation of *minimalization* associates with each total function $f(y, x)$ the function $h(x)$, whose value for a given x is the least value

of y, if one such exists, for which $f(y, x) = 0$, and which is undefined if no such y exists. The minimalization operation is written $h(x) = \mu y[f(y, x) = 0]$. The total function $f(y, x)$ is called *regular* if $\mu y[f(y, x) = 0]$ is total. For the variable x may be substituted an n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$. [As a predicate, $f(y, x) = 0$ may be denoted as $P(x, y)$. With $P(x, y)$ for an arbitrary predicate $\mu y\, P(x, y)$ is the function of x, whose value is the least value of y, if one such exists, such that $P(x, y)$.]

The regular definition of the recursive functions is as follows.

Definition 11.5A. A function is *partial recursive* if it can be obtained by a finite number of applications of composition and minimalization beginning with functions of the list:

1. $S(x) =_{\text{def}} x + 1$ the successor function
2. $U_i^n(x_1, x_2, \ldots, x_n) =_{\text{def}} x_i$ the projection functions ($1 \leq i \leq n$)
3. $x + y$ addition
4. $x \dot{-} y =_{\text{def}} \begin{cases} x - y & \text{if } x \geq y \\ 0 & \text{if } x < y \end{cases}$ proper subtraction
5. xy multiplication

A function is *recursive** if it is partial recursive and total. (A function is *primitive recursive†* if it can be obtained by a finite number of applications of composition and primitive recursion, beginning with the functions of the above list.)

Example 11.2A. The function $\alpha(x) = 1 \dot{-} x$, that is, $\alpha(0) = 1$, $\alpha(x) = 0$ for $x > 0$, is (primitive) recursive. This is easily seen by the decomposition:

$$\alpha(x) = S(U_1^1(x) \dot{-} U_1^1(x)) \dot{-} U_1^1(x)$$

which shows that $\alpha(x)$ satisfies Definition 11.5A.

Example 11.3A. The function $[\sqrt{x}]$, defined as the largest integer $\leq \sqrt{x}$, is recursive. First let us observe that $y = [\sqrt{x}]$ is the smallest integer such that $y > \sqrt{x} - 1$, that is, such that $(y + 1)^2 \dot{-} x \neq 0$. Hence, we can decompose $[\sqrt{x}]$ as follows:

$$[\sqrt{x}] = \mu y[(y + 1)^2 \dot{-} x \neq 0]$$
$$= \mu y[\alpha[S(U_2^2(x, y)) \cdot S(U_2^2(x, y)) \dot{-} U_1^2(x, y)] = 0],$$

that is, $[\sqrt{x}]$ satisfies Definition 11.5A.

---

* Davis [5] defines the recursive functions in a slightly different but equivalent way.
† The primitive recursive functions constitute a proper subclass of the recursive functions, and the operation of primitive recursion preserves the property of being recursive; the reader is directed to Davis [5] for proof.

In order to prove that the partial recursive functions are partially computable, it suffices to prove that the initial functions are computable and that the operations of composition and minimalization preserve partial computability. (In the same way, a proof that the operation of primitive recursion preserves computability yields a proof that the primitive recursive functions are computable.)

We have already seen (Ex. 11.1A, p. 370) that initial function number 3 is computable. The computability of the other initial functions can also be established by exhibiting lists of quadruples for Turing machines which compute them.

Example 11.4A. The list of quadruples for a Turing machine which computes initial function number 4, $x \dot- y$, is given below, together with a programming explanation:

$\langle q_0, 1, B, q_0\rangle$	erase 1 on the left
$\langle q_0, B, R, q_1\rangle$	
$\langle q_1, 1, R, q_1\rangle$	go to the right of separating blank
$\langle q_1, B, R, q_2\rangle$	
$\langle q_2, 1, R, q_2\rangle$	go to the right-most 1
$\langle q_2, B, L, q_3\rangle$	
$\langle q_3, 1, B, q_3\rangle$	erase 1 on the right
$\langle q_3, B, L, q_4\rangle$	
$\langle q_4, 1, L, q_5\rangle$	if the y-representing 1's are exhausted, stop; otherwise continue
$\langle q_5, 1, L, q_5\rangle$	go to the left of separating blank
$\langle q_5, B, L, q_6\rangle$	
$\langle q_6, B, R, q_8\rangle$	if the x-representing 1's are exhausted, go to q_8; otherwise
$\langle q_6, 1, L, q_7\rangle$	continue
$\langle q_7, 1, L, q_7\rangle$	go to the left-most 1 and return to q_1
$\langle q_7, B, R, q_1\rangle$	
$\langle q_8, B, R, q_9\rangle$	erase all 1's on the tape if $x < y$ and stop
$\langle q_9, 1, B, q_8\rangle$	

Demonstration of the computability of the remaining initial functions of Definition 11.5A is left as an exercise.

The (partial) computability of composition and minimalization of computable functions can be demonstrated by exhibiting Turing machines which compute them (see Davis [5]). Here we will instead argue their computability by referring to Turing's thesis.

Theorem 11.1A. The operation of composition preserves both the property of being computable and the property of being partially computable.

Argument. Suppose that the functions $g_1(x), g_2(x), \ldots, g_n(x)$ of Def. 11.2A, p. 372, are computable. This means that there are Turing machines $Z_1, Z_2, \ldots, Z_n$, respectively, which compute these functions. We also assume that $f(y_1, y_2, \ldots, y_n)$ is computable by Turing machine Z_0. Then there is obviously an effective way of letting $Z_1, Z_2, \ldots, Z_n$ compute their functions and of transcribing the results in order with separating blanks onto an otherwise blank tape which is fed to Z_0; Z_0 will compute the composition function $h(x)$ of Definition 11.2A. Since the whole procedure is effective, there is, according to Turing's hypothesis, a Turing machine Z which computes this operation of composition. Hence the operation of composition preserves the property of being computable. Suppose next that f and the g_i-functions are partially computable. Then the above argument applies for all x in the domain of $h(x)$. However, if $\langle g_1(x), g_2(x), \ldots, g_n(x) \rangle$ does not belong in the domain of f, Z_0 will never halt. Again, if z does not belong to the domain of some $g_i(x)$, Z_i will never halt. In any one of these cases the machine Z will compute forever if it is instructed to wait for the completion of the $g_i(x)$-computations as well as the $f(g_1(x), g_2(x), \ldots, g_n(x))$-computations before it halts. Hence Z will perform a nonhalting computation from the argument x if and only if x does not belong to the domain of the composition function, which completes the argument of Theorem 11.1A.

Theorem 11.2A. The operation of minimalization preserves the property of being partially computable.

Argument. Suppose that the total function $f(y, x)$ in Def. 11.4A, p. 372, is computable by a Turing machine Z. Thus there is an effective way to compute in succession $f(0, x), f(1, x), f(2, x), \ldots$, and at each step to compare the result with 0. The first occurrence of a 0 can thus be effectively registered, provided that there is such an occurrence, that is, that x belongs to the domain of $\mu y[f(y, x) = 0]$. According to Turing's thesis there is a Turing machine which computes $\mu y[f(y, x) = 0]$ for all x in the domain of this function and which never stops for x not in the domain. The operation of minimalization* thus preserves the property of being partially computable.

Theorem 11.3A. If a function is (partial) recursive, then it is (partially) computable.†

---

* Notice that minimalization, according to Definition 11.4A, must be applied only to total functions. Some authors define minimalization without this restriction, and then minimalization does not preserve the property of being partially computable. There are, for example, partially computable functions $\psi_u(x)$ such that $\mu y(\psi_u(y) = x)$ is not computable although total. Hence such an extended minimalization over a partially computable function may define a nonpartially computable function.

† This notation is a short form for "If a function is recursive, then it is computable; and if a function is partial recursive, then it is partially computable." This form carries throughout the chapter.

Proof. The initial functions in Def. 11.5A, p. 373, are all recursive and hence, also partial recursive. Since the operations of composition and minimalization preserve the property of being partial recursive, it follows from Definition 11.5A that every partial recursive function is partially computable. Finally a function is recursive (computable) if and only if it is partial recursive (partially computable) and total.

Later we shall see that the converse of Theorem 11.3A is also true (see Tm. 11.7A, p. 379).

I.5. RECURSIVE SETS AND COMPUTABLE SETS

So far we have been dealing with computability and recursivity of functions; using the characteristic function of a set, defined below, we can extend these properties to sets.

Definition 11.6A. Let S be a set of n-tuples $\langle x_1, x_2, \ldots, x_n \rangle$. The *characteristic function*, $C_S(x_1, x_2, \ldots, x_n)$, of S is then defined by

$$C_S(x_1, x_2, \ldots, x_n) = 1 \quad \text{if } \langle x_1, x_2, \ldots, x_n \rangle \in S,$$

$$C_S(x_1, x_2, \ldots, x_n) = 0 \quad \text{if } \langle x_1, x_2, \ldots, x_n \rangle \notin S.$$

If the characteristic function of a set is recursive, it is also computable (Theorem 11.3A). There is a Turing machine, then, which determines whether or not any given element (n-tuple) belongs to the set; and, as membership is really that which defines a set, it is reasonable to say that the set in question is computable. Hence the following definitions.

Definition 11.7A. Let S be a set of n-tuples. Then S is *recursive (computable)* if its characteristic function is recursive (computable).

Theorem 11.4A. Let R and S be recursive (computable) sets. Then the sets* $R \cup S$, $R \cap S$, and $\bar{R}$ are also recursive (computable).

Proof. The characteristic functions of the sets $R \cup S$, $R \cap S$, and $\bar{R}$ can be decomposed as follows:

$$C_{R \cup S} = C_R + C_S \doteq (C_R \cdot C_S), \quad C_{R \cap S} = C_R \cdot C_S, \quad C_{\bar{R}} = 1 \doteq C_R.$$

Hence, if R and S are recursive, so are $R \cup S$, $R \cap S$, and $\bar{R}$. Again, if R and S are computable, C_R and C_S are computable. As seen from the above decompositions, then, $C_{R \cup S}$, $C_{R \cap S}$, and $C_{\bar{R}}$ can be obtained by the operation

---

* $\bar{R}$ denotes the complement of R with respect to the set of all n-tuples, N^n.

of composition from computable functions. From Tm. 11.1A, p. 374, we know that composition preserves the property of being computable; hence $R \cup S$, $R \cap S$, and $\bar{R}$ are computable.

I.6. RECURSIVE AND COMPUTABLE PREDICATES

Having defined recursive (computable) sets, we are able to define recursive (computable) predicates via the extension set of a predicate.

Definition 11.8A. The *predicate* $P(x_1, x_2, \ldots, x_n)$ *is recursive* (*computable*) if its extension $\{\langle x_1, x_2, \ldots, x_n \rangle : P(x_1, x_2, \ldots, x_n)\}$ is recursive (computable).

Definition 11.8A implies, for example, that if a predicate is computable there is a Turing machine which can compute whether the predicate is true or false for any given n-tuple.

Theorem 11.5A. Let P and Q be recursive (computable) predicates. Then the predicates $P \vee Q$, $P \ \& \ Q$, and $\neg Q$ are also recursive (computable).

Proof. Let C_P denote the characteristic function of the extension set to a predicate P. Thus

$$C_{P \vee Q} = C_P + C_Q \dot{-} (C_P \cdot C_Q), \quad C_{P \& Q} = C_P \cdot C_Q, \quad C_{\neg P} = 1 \dot{-} C_P.$$

Hence, if P and Q are recursive (computable), $P \vee Q$, $P \ \& \ Q$, and $\neg P$ are also recursive (computable).

Example 11.5A. The predicates $P(x, y)$ and $Q(x, y)$, defined as $P(x, y) \equiv (x \leq y)$ and $Q(x, y) \equiv (x = y)$, are recursive and hence also computable. Let us as a proof first consider $P(x, y)$. The characteristic function of its extension is $C_P = 1 \dot{-} (x \dot{-} y)$. Hence C_P is recursive and by Tm. 11.3A, p. 375, computable, that is, $P(x, y)$ is recursive and computable. On the other hand, $Q(x, y)$ can be written as

$$Q(x, y) \equiv (x = y) \equiv (x \leq y) \ \& \ (y \leq x) \equiv P(x, y) \ \& \ P(y, x).$$

By Theorem 11.5A, $Q(x, y)$ is recursive and hence computable.

I.7. THE TURING MACHINE PREDICATE, $T(z, x, y)$

It is worthwhile to apply the concepts of computability to predicates which themselves make statements about Turing machines and their computability. Since the computable and recursive predicates have been defined for variables

that range over N, the set of integers, such applications require that we arithmetize the theory of Turing machines. Consider, for example, the following predicate: $T_0(Z, x, Y) \equiv$ "When started from a tape printed with the number (n-tuple) x, Turing machine Z performs the computation sequence Y." We cannot really apply Definition 11.8A because two of the predicate variables are defined, not on N, but instead on the set of Turing machines and the set of sequences of tape expressions. First we will have to encode the set of Turing machines onto a set of integers such that, without any ambiguity, we can speak of a Turing machine as though it were an integer. The same applies to the set of sequences of tape expressions as well.

A way of encoding the Turing machines would be to encode their lists of quadruples. One such way is provided by the Gödel enumeration technique; when this is used, it is customary to speak of the *Gödel number* of a Turing machine (see Davis [5]). Given such a number, z, one can effectively compute the list of quadruples that define the corresponding machine Z and vice versa. Without going into this or any other enumeration technique, we assume that there are such effective codes. We will use the term *code number* for an integer, z, that effectively describes a Turing machine, Z.

Likewise we can talk about the code number for the computation Y of a Turing machine, again taking it for granted that there is such an effective code. With these preliminary points now settled, we can define the Turing machine predicate.

Definition 11.9A. The *Turing machine predicate*, $T(z, x, y)$, is true if and only if z and y, respectively, are the code numbers of a Turing machine, say Z, and a computation sequence, say Y, such that Z, when started on x, performs Y. For the variable x may be substituted an n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$.

Thus $T(z, x, y)$ is a predicate about which we may ask whether or not it is recursive.

Theorem 11.6A. The Turing machine predicate, $T(z, x, y)$, is recursive and hence computable.

Argument. A proof of Theorem 11.6A is given in Davis [5] on the basis of a Gödel encoding. Here we will only present an argument utilizing Church's hypothesis. What we want to prove is that, given an arbitrary triple $\langle z, x, y \rangle$, we can effectively decide whether $T(z, x, y)$ is true or false. It then follows from Church's thesis, Definition 11.8A, and Theorem 11.3A that the predicate is recursive and computable. Given a triple $\langle z, x, y \rangle$ we can, according to our assumptions on the encoding, effectively reconstruct a list of quadruples of a Turing machine Z, or decide that z is not a code number of a Turing machine. In the latter case $T(z, x, y)$ is false. Furthermore, we can effectively reconstruct a sequence of tape expressions Y, or decide that y is not a code

number of a sequence of tape expressions. In the latter case $T(z, x, y)$ is false. If the predicate has not yet been proved false, we proceed as follows. We effectively apply the quadruple to the initial tape expression x (cf. Ex. 11.1A, p. 370) and compare the result with Y. If the two do not agree, the predicate is false; otherwise, the comparison is continued to the end of Y. Since Y is defined, the total number of comparisons will be finite and can be effectively executed. [Compare Example 11.5A, showing that $x = y$ is recursive, and Theorem 11.5A, showing that $x \neq y$ is recursive; a comparison between two integers is thus a recursive procedure.] The whole procedure is thus effective, and the theorem follows.

For a full use of the Turing machine predicate we will also need the following definition.

Definition 11.10A. Let y be the code number of a computation sequence Y. Let n be the number of 1's in the final element of Y, that is, n is the number computed by the machine which produces the sequence Y. As a function of y, n will be denoted $U(y)$, that is, $n = U(y)$.

Obviously, $U(y)$ is an effectively computable function of y.

I.8. KLEENE'S NORMAL FORM OF A PARTIALLY COMPUTABLE OR PARTIAL RECURSIVE FUNCTION

By Theorem 11.6A, $T(z, x, y)$ is a recursive predicate. Therefore, by Def. 11.5A, p. 373, $\mu y\, T(z, x, y)$ is a partial recursive function of z and x. Finally, by Definition 11.10A, $U(\mu y\, T(z, x, y))$ is partial recursive.

We notice that, for a given code number z, x is in the domain of $\mu y\, T(z, x, y)$ if and only if x is in the domain of $\psi_z(x)$, the function computed by Turing machine Z. Furthermore, if x belongs to this domain, $\mu y\, T(z, x, y)$ will be the code number of the sequence of expressions computed by Z because the Turing machine is deterministic, that is, it can have at most one computation sequence for each x of its domain. This means that $U(\mu y\, T(z, x, y)) = \psi_z(x)$, a decomposition showing that every (partially) computable function $\psi_z(x)$ is (partial) recursive. It is thus possible to sharpen Tm. 11.3A, p. 375, into the following normal form theorem.

Theorem 11.7A. The function $f(x_1, x_2, \ldots, x_n)$ is partially computable if and only if there is a number z_0 such that

$$f(x_1, x_2, \ldots, x_n) = U(\mu y\, T(z_0, x_1, x_2, \ldots, x_n, y)),$$

that is, if and only if $f(x_1, x_2, \ldots, x_n)$ is partial recursive. This form is called the *Kleene normal form* of a partially computable (i.e., partial recursive) function.

Although for simplicity we have occasionally used Turing's or Church's thesis in the arguments for Theorem 11.7A, this is not necessary. A proof of this important theorem, without using these hypotheses, may be found in Davis [5].

I.9. SEMICOMPUTABLE PREDICATES

To extend our understanding of the absolutely computable predicates we want to demonstrate the existence of predicates that are not computable. Hence we turn to the semicomputable predicates.

Consider a predicate $P(x_1, x_2, \ldots, x_n)$, saying that a certain Turing machine, Z_0, will perform a computation, that is, will halt within a finite time when supplied with a tape printed with $\langle x_1, x_2, \ldots, x_n \rangle$. Suppose that we want to compute $P(x_1, x_2, \ldots, x_n)$ and that our only aid is the machine Z_0, which is presented to us as a black box. We can compute $P(x_1, x_2, \ldots, x_n)$ by printing an n-tuple on the tape of Z_0 and letting Z_0 start its computation. If Z_0 halts, we know that $P(x_1, x_2, \ldots, x_n)$ is true and computable for that particular n-tuple. But what conclusion can be drawn if Z_0 does not halt within a reasonable time? Evidently none. We could have Z_0 compute for some further time, hoping it would eventually halt; but obviously, if Z_0 never halts for a given n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$, we can never find out whether $\langle x_1, x_2, \ldots, x_n \rangle$ belongs to the domain of the partially computable function $\psi_{z_0}(x_1, x_2, \ldots, x_n)$. Thus $P(x_1, x_2, \ldots, x_n)$ is not computable for this n-tuple on the basis of the knowledge about $P(x_1, x_2, \ldots, x_n)$ that we possess.

We are thus led to the following definition of a semicomputable predicate.

Definition 11.11A. A predicate $P(x_1, x_2, \ldots, x_n)$ is *semicomputable* if there exists a partially computable function whose domain is the extension of P, $\{\langle x_1, x_2, \ldots, x_n \rangle : P(x_1, x_2, \ldots, x_n)\}$.

The concept of semicomputability is weaker than the concept of computability in the following sense. Every computable predicate is semicomputable (cf. Theorem 11.8A below), but not all semicomputable predicates are computable (cf. Theorem 11.11A below).

Theorem 11.8A. Every computable predicate is semicomputable.

Proof. Let $P(x_1, x_2, \ldots, x_n)$ be computable. Then $\{\langle x_1, x_2, \ldots, x_n \rangle : P(x_1, x_2, \ldots, x_n)\}$ is the domain of the partially computable function $\mu y [(1 \dot{-} C_P(x_1, x_2, \ldots, x_n)) + y = 0]$.

Every semicomputable predicate can be written in a normal form as follows.

Theorem 11.9A. (Kleene's enumeration theorem). Let $P(x_1, x_2, \ldots, x_n)$ be a semicomputable predicate. Then there is a number z such that

$$P(x_1, x_2, \ldots, x_n) \equiv \exists y \ T(z, x_1, x_2, \ldots, x_n, y).$$

Proof. Let $P(x_1, x_2, \ldots, x_n)$ be a semicomputable predicate. Then, according to Definition 11.11A, there is a Turing machine Z computing a function $\psi_z(x_1, x_2, \ldots, x_n)$ whose domain is

$$\{\langle x_1, x_2, \ldots, x_n \rangle : P(x_1, x_2, \ldots, x_n)\}.$$

On the other hand, the domain of $\psi_z(x_1, x_2, \ldots, x_n)$ is, according to Theorem 11.7A, $\{\langle x_1, x_2, \ldots, x_n \rangle : \exists y \ T(z, x_1, x_2, \ldots, x_n, y)\}$.

A form of a converse of Theorem 11.9A is the following.

Theorem 11.10A. Let $R(y, x_1, x_2, \ldots, x_n)$ be a computable predicate. Then $\exists y \ R(y, x_1, x_2, \ldots, x_n)$ is semicomputable.

Proof. Since $R(y, x_1, x_2, \ldots, x_n)$ is computable,

$$\mu y \ R(y, x_1, x_2, \ldots, x_n)$$

is partially computable. The domain of this partially computable function is the extension of the predicate $\exists y \ R(y, x_1, x_2, \ldots, x_n)$. Hence, by Definition 11.11A, this predicate is semicomputable.

Theorem 11.11A. A predicate $P(x_1, x_2, \ldots, x_n)$ is computable if and only if both $P(x_1, x_2, \ldots, x_n)$ and $\neg P(x_1, x_2, \ldots, x_n)$ are semicomputable.

Proof. Let us assume that $P(x_1, x_2, \ldots, x_n)$ is computable. It follows from Theorem 11.5A that $\neg P(x_1, x_2, \ldots, x_n)$ also is computable. Hence by Theorem 11.8A both $P(x_1, x_2, \ldots, x_n)$ and $\neg P(x_1, x_2, \ldots, x_n)$ are semicomputable. Conversely, let us assume that $P(x_1, x_2, \ldots, x_n)$ and $\neg P(x_1, x_2, \ldots, x_n)$ are semicomputable. Then, by Theorem 11.9A, there are numbers z_1 and z_2 such that

$$P(x_1, x_2, \ldots, x_n) \equiv \exists y_1 \ T(z_1, x_1, x_2, \ldots, x_n, y_1),$$
$$\neg P(x_1, x_2, \ldots, x_n) \equiv \exists y_2 \ T(z_2, x_1, x_2, \ldots, x_n, y_2).$$

For each n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$, either

$$P(x_1, x_2, \ldots, x_n) \quad \text{or} \quad \neg P(x_1, x_2, \ldots, x_n)$$

is true. Hence, for each n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$, it is true that

$$\exists y_1 \ T(z_1, x_1, x_2, \ldots, x_n, y_1) \vee \exists y_2 \ T(z_2, x_1, x_2, \ldots, x_n, y_2),$$

that is,

$$\exists y[T(z_1, x_1, x_2, \ldots, x_n, y) \vee T(z_2, x_1, x_2, \ldots, x_n, y)].$$

There exists, therefore, a smallest such y; that is, the function

$$h(x_1, x_2, \ldots, x_n) = \mu y[T(z_1, x_1, x_2, \ldots, x_n, y) \lor T(z_2, x_1, x_2, \ldots, x_n, y)]$$

is total. Let us recall that the T-predicates are recursive (Theorem 11.6A). The form of h thus reveals that h is partial recursive and total, that is, recursive and computable. Finally,

$$P(x_1, x_2, \ldots, x_n) \equiv T(z_1, x_1, x_2, \ldots, x_n, h(x_1, x_2, \ldots, x_n)),$$

that is $P(x_1, x_2, \ldots, x_n)$ is computable.

It follows from Theorem 11.11A that an eventual semicomputable predicate which is not computable must have a negation which is not semicomputable. There are indeed such predicates.

Theorem 11.12A. The predicate $\exists y\, T(x, x, y)$ is semicomputable but not computable.

Proof. Suppose that $\neg \exists y\, T(x, x, y)$ were semicomputable. Then, by Theorem 11.9A, there would be a number z_0 such that $\neg \exists y\, T(x, x, y) \equiv \exists y\, T(z_0, x, y)$. We need only choose the number z_0 for x to find that a contradiction has resulted.

Another consequence of the noncomputability of the predicate $\exists y\, T(x, x, y)$ is the following.

Theorem 11.13A. Let $f_c(x)$ be the completion of a function $f(x)$, that is, $f_c(x) = f(x)$ for all x in the domain of $f(x)$, and $f_c(x) = 0$ for all other x. Then there exists a partially computable function $f(x)$ with a noncomputable completion $f_c(x)$.

Proof. Let $\varphi(x) = 1 + (x \div x)$, that is, $\varphi(x)$ is the computable function whose value is 1 for all values of x. Let $f(x) = \varphi(\mu y\, T(x, x, y))$, that is, $f(x)$ is partially computable. Obviously $f_c(x)$ is the characteristic function of the predicate $\exists y\, T(x, x, y)$. By Theorem 11.12A, $f_c(x)$ is not computable although total.

Theorem 11.13A may suggest the following comment: since $f(x)$ is partially computable, that is, effectively computable for every x in its domain, there should be an effective way of computing $f_c(x)$, that is, of computing $f(x)$ or of assigning zero as the function value. This, however, requires a decision as to whether or not x belongs to the domain of f, a decision which is not effectively implemented.

Although the negation of a semicomputable predicate is not always semicomputable, the property of being semicomputable is preserved under the &- and $\lor$-operations (Theorem 11.14A below). For the proof of Theorem 11.14A, we need the following definition.

Definition 11.12A. $\tau(x_1, x_2)$ is a recursive 1-1 *pair-encoding function* that encodes the ordered integer pairs $\langle x_1, x_2 \rangle$ onto the set of integers.* $\lambda(y)$ and $\rho(y)$ are recursive *pair-decoding functions* such that $\lambda(\tau(x_1, x_2)) = x_1$ and $\rho(\tau(x_1, x_2)) = x_2$.

Theorem 11.14A. If $P(x_1, x_2, \ldots, x_n)$ and $Q(x_1, x_2, \ldots, x_n)$ are semicomputable, then so are $P(x_1, x_2, \ldots, x_n) \vee Q(x_1, x_2, \ldots, x_n)$ and $P(x_1, x_2, \ldots, x_n)$ & $Q(x_1, x_2, \ldots, x_n)$.

Proof. Let

$$P(x_1, x_2, \ldots, x_n) \equiv \exists y_1\, T(z_1, x_1, x_2, \ldots, x_n, y_1)$$

and

$$Q(x_1, x_2, \ldots, x_n) \equiv \exists y_2\, T(z_2, x_1, x_2, \ldots, x_n, y_2).$$

Then

$$P(x_1, x_2, \ldots, x_n) \vee Q(x_1, x_2, \ldots, x_n)$$
$$\equiv \exists y(T(z_1, x_1, x_2, \ldots, x_n, y) \vee T(z_2, x_1, x_2, \ldots, x_n, y)).$$

Here $T(z_1, x_1, x_2, \ldots, x_n, y) \vee T(z_2, x_1, x_2, \ldots, x_n, y)$ is recursive. Hence, by Theorem 11.10A, $P(x_1, x_2, \ldots, x_n) \vee Q(x_1, x_2, \ldots, x_n)$ is semicomputable. We next turn to

$$P(x_1, x_2, \ldots, x_n) \,\&\, Q(x_1, x_2, \ldots, x_n)$$
$$\equiv \exists y_1 T(z_1, x_1, x_2, \ldots, x_n, y_1) \,\&\, \exists y_2 T(z_2, x_1, x_2, \ldots, x_n, y_2)$$
$$\equiv \exists y_1 \exists y_2 (T(z_1, x_1, x_2, \ldots, x_n, y_1) \,\&\, T(z_2, x_1, x_2, \ldots, x_n, y_2)).$$

With the notations of Definition 11.12A we have

$$[P(x_1, x_2, \ldots, x_n) \,\&\, Q(x_1, x_2, \ldots, x_n)]$$
$$\equiv \exists \tau[T(z_1, x_1, x_2, \ldots, x_n, \lambda(\tau)) \,\&\, T(z_2, x_1, x_2, \ldots, x_n, \rho(\tau))],$$

that is, $P \,\&\, Q$ is obtained by existential quantification over a computable predicate. By Theorem 11.10A, $P \,\&\, Q$ is semicomputable.

I.10. THE KLEENE PROJECTION THEOREM

Two forms of the Kleene projection theorem will be given, namely, Theorems 11.15A and 11.16A.

---

* One such pair-encoding function is $\tau(x_1, x_2) = \frac{1}{2}[(x_1 + x_2)^2 + 3x_1 + x_2]$. The reader is advised to furnish the function values $\tau(x_1, x_2)$ on a Cartesian representation of the domain and to convince himself that each function value $\tau(x_1, x_2)$ effectively determines a unique $\lambda(\tau) = x_1$ and $\rho(\tau) = x_2$.

Theorem 11.15A. There is a recursive function $f(z, y)$ such that

$$\psi_z(y, x_1, \ldots, x_n) = \psi_{f(z, y)}(x_1, x_2, \ldots, x_n).$$

Furthermore, with any two recursive functions $c_1(x)$ and $c_2(x)$ such that the range of $c_1(x)$ is nonfinite, there is a recursive function $\varphi(z, y)$ such that $\psi_z(y, x_1, x_2, \ldots, x_n) = \psi_{\varphi(z, y)}(x_1, x_2, \ldots, x_n)$ and such that $c_1(\varphi(z, y)) > c_2(y)$.

Proof. For each fixed value $\langle z_0, y_0 \rangle$ of $\langle z, y \rangle$, $\psi_{z_0}(y_0, x_1, \ldots, x_n)$ is a partial recursive function of $x_1, x_2, \ldots, x_n$. Hence there is a Turing machine W_0 with code number w_0 which computes it. The important content of the theorem is that $w_0 = f(z_0, y_0)$, where f is recursive. To convince ourselves of this property of f, let us give an effective description of the Turing machine W_0. Beginning with the tape expression $1^{x_1 + 1} B 1^{x_2 + 1} B \ldots B 1^{x_n + 1}$, W_0 should first print $1^{y_0 + 1}$ to the left, thus producing the expression

$$1^{y_0 + 1} B 1^{x_1 + 1} B 1^{x_2 + 1} B \ldots B 1^{x_n + 1},$$

and thereafter scan the left-most 1. From there on W_0 should continue just like the known machine Z_0 with the code number z_0. Hence W_0 is defined by a list of quadruples which begins as follows:

$$\langle q_0, 1, L, q_0 \rangle, \langle q_0, B, L, q_1 \rangle, \langle q_i, B, 1, q_i \rangle, \langle q_i, 1, L, q_{i+1} \rangle,$$
$$\langle q_{y_0+1}, B, 1, q_{y_0+2} \rangle; \quad 1 \le i \le y_0.$$

From there on the list should continue with the quadruple list of Z_0, with, however, the states renumbered so that q_j of Z_0 now is called $q_{y_0 + 2 + j}$. Thus, given $\langle z_0, y_0 \rangle$, there is an effective way of producing $w_0 = f(z_0, y_0)$, that is, $f(z, y)$ is recursive by Church's thesis. A proof without explicit use of this hypothesis may be found in Davis [5]. This completes the proof of the first part of Theorem 11.15A.

For the proof of the second part, let us observe that to each code number $f(z, y)$ there corresponds a unique set of quadruples which defines a computation. Precisely the same computation is in this way associated also with an unlimited number of distinct quadruple lists. If the $f(z, y)$-list is written on $Q = \{q_0, q_1, \ldots, q_n\}$. $S = \{s_1, s_2, \ldots, s_m\}$, it can, for example, be continued with an unlimited amount of quadruples of the type $\langle q_{n+i}, s_{m+j}, s_{m+k}, q_{n+l} \rangle$, $i, j, k, l > 0$, without having the computation affected. Hence, there must be an unlimited number of $\varphi(z, y)$-values such that $\psi_{f(z, y)}(x_1, x_2, \ldots, x_n) = \psi_{\varphi(z, y)}(x_1, x_2, \ldots, x_n)$, and in particular one $\varphi(z, y)$-value such that $c_1(\varphi(z, y))$ is larger than $c_2(y)$, provided that the range of $c_1(x)$ is nonfinite, as is assumed. Furthermore, both $c_1(x)$ and $c_2(x)$ are assumed recursive. Therefore, there is an effective way of obtaining such a $\varphi(z, y)$-value from z, y, $f(z, y)$, $c_1(x)$, $c_2(x)$, where also $f(z, y)$ is recursive. Hence, by Church's thesis, $\varphi(z, y)$ is recursive. This completes the proof of Theorem 11.15A.

Theorem 11.16A. Let $R(w, x_1, x_2, \ldots, x_n)$ be a semicomputable predicate. Then there is a recursive function $f(w)$ such that

$$R(w, x_1, x_2, \ldots, x_n) \equiv \exists y\, T(f(w), x_1, x_2, \ldots, x_n, y).$$

Proof. Theorem 11.15A implies that

$$U(\mu y\, T(z, w, x_1, x_2, \ldots, x_n, y)) = U(\mu y\, T(\varphi(z, w), x_1, x_2, \ldots, x_n, y)),$$

and the corresponding domains are equal:

$$\exists y\, T(z, w, x_1, x_2, \ldots, x_n, y) \equiv \exists y\, T(\varphi(z, w), x_1, x_2, \ldots, x_n, y).$$

Since $R(w, x_1, x_2, \ldots, x_n)$ is a semicomputable predicate, there is a code number, r, such that

$$R(w, x_1, x_2, \ldots, x_n) \equiv \exists y\, T(r, w, x_1, x_2, \ldots, x_n, y).$$

Hence

$$R(w, x_1, x_2, \ldots, x_n) \equiv \exists y\, T(\varphi(r, w), x_1, x_2, \ldots, x_n, y).$$

Here $\varphi(r, w) = f(w)$ is a recursive function of w (cf. Theorem 11.15A), which proves the theorem.

I.11. RECURSIVELY ENUMERABLE SETS

The recursively enumerable sets are closely related to the semicomputable predicates.

Definition 11.13A. A set is *recursively enumerable* if it is the range of a partial recursive function.

Theorem 11.17A. A set S is recursively enumerable if and only if $S = \{x: P(x)\}$, where $P(x)$ is semicomputable.

Proof. Let S be a recursively enumerable set, that is, there exists a partial recursive function $f(x) = U(\mu y\, T(z_0, x, y))$, whose range is S. Hence

$$w \in S \equiv \exists x\, \exists y[T(z_0, x, y)\ \&\ w = U(y)],$$

that is,

$$S = \{w: \exists \tau[T(z_0, \lambda(\tau), \rho(\tau))\ \&\ w = U(\rho(\tau))]\},$$

where τ, λ, ρ are the pair-encoding and pair-decoding functions of Def. 11.12A, p. 383. Thus S is the extension of a semicomputable predicate. Conversely, let us assume that $P(x)$ is a semicomputable predicate, that is, there is a

number z_1 such that $P(x) = \exists y \, T(z_1, x, y)$. Then the extension $S = \{x : P(x)\}$ of $P(x)$ is enumerated by the partial recursive function

$$g(n) = \lambda(\mu\tau[T(z_1, \lambda(\tau), \rho(\tau)) \, \& \, (\tau \geq n)]),$$

which completes the proof.

Theorem 11.18A. A set S is recursively enumerable if and only if $S = \varnothing$ or S is the range of a recursive function.

Proof. Consider first $S = \varnothing$, which is the range of the partial recursive function $\mu y(1 + y + x = 0)$. By Definition 11.13A, S is recursively enumerable. Next let S be the range of a recursive function. Since every recursive function is also partial recursive, it follows from Definition 11.13A that S is recursively enumerable. Conversely, assume that S is a nonempty, recursively enumerable set enumerated by the partial recursive function $g(x)$. Then there is the following effective way of enumerating all the elements of S. Let G be the Turing machine which computes g. First let G make two computations on the numbers 0 and 1, respectively, such that each computation is cut off after one step. If any number is produced, that is, if any computation halts before being cut off, the result is added to a list L. Next repeat the procedure with $n = 2$, that is, G computes two steps on each of the arguments 0, 1, 2. If any new numbers are produced, they are also added to the list L. At stage n, G computes n steps on each of the arguments 0, 1, 2, ..., $n - 1$, n. By this effective enumeration every element of S and no other elements will be added to the list L. By Church's hypothesis there is a recursive function which implements this enumeration, that is, S is the range of a recursive function.

It should be observed that the enumeration technique used in this proof does not contradict the previous result that there are partial recursive functions with a nonrecursive completion (Tm. 11.13A, p. 382; compare also Ex. 11.6A, p. 392). The technique shows only that the enumeration done by a partial recursive function can be performed by a recursive function as well.

Theorem 11.19A. A set is recursive if and only if both S and $\bar{S}$ are recursively enumerable.

Proof. Assume that S is recursive, that is, $S = \{x : P(x)\}$ where $P(x)$ is computable. Then $P(x)$ is also semicomputable, that is, S is recursively enumerable. Since S is recursive, $\bar{S}$ is also recursive and so also recursively enumerable. Conversely, assume that both S and $\bar{S}$ are recursively enumerable, that is, $S = \{x : P(x)\}$ and $\bar{S} = \{x : \neg P(x)\}$, where both $P(x)$ and $\neg P(x)$ are semicomputable. By Tm. 11.11A, p. 381, $P(x)$ is computable (i.e., recursive). Hence S is recursive.

An intuitive argument for Theorem 11.19A is as follows. If both S and $\bar{S}$ are enumerated by recursive functions, that is, by Turing machines, say Z_1 and Z_2, we can effectively decide for each x whether or not $x \in S$, that is, S is recursive by Church's thesis. We only have to let Z_1 and Z_2 work on the arguments 0, 1, 2, ... and see which of them first produces a given number x. If it is Z_1, then $x \in S$. If it is Z_2, then $x \notin S$. The important point is that any given x sooner or later must be generated by Z_1 or Z_2. Conversely, if S is recursive, its characteristic function $C_S(x)$ is recursive and hence also $f(x) = x \cdot C_S(x) + x_0 \cdot (1 \dot{-} C_S(x))$, which enumerates S if x_0 is a particular element of S, for example, the smallest. In the same way $\bar{S}$ is then recursive, that is, recursively enumerable.

Theorem 11.20A. The set $K = \{x\colon \exists y\, T(x, x, y)\}$ is recursively enumerable but not recursive.

Proof. By Theorem 11.12A we know that the predicate $\exists y\, T(x, x, y)$ is semicomputable but not computable. Theorem 11.20A hence follows from Theorem 11.17A.

Although the complement of a recursively enumerable set need not be recursively enumerable, the operations of union and intersection preserve the property of being recursively enumerable.

Theorem 11.21A. If R and S are recursively enumerable sets, so are $R \cup S$ and $R \cap S$.

Proof. Tm. 11.14A, p. 383, and Tm. 11.17A, p. 385, imply the statement.

I.12. PRODUCTIVE AND CREATIVE SETS

If R is a recursively enumerable set that is not recursive, then, as we know, $\bar{R}$ is not recursively enumerable. Hence, with every recursively enumerable set P such that $P \subset \bar{R}$, there is an m such that $m \notin P$ and $m \in \bar{R}$. We inquire whether there is an effective procedure by which, given P, we can obtain such an m and thus extend P to a larger recursively enumerable subset of $\bar{R}$. A recursively enumerable set will be considered given, if a code number n is presented such that $P = \{x\colon \exists y\, T(n, x, y)\}$. This motivates the following definitions.

Definition 11.14A. Let $S_{[n]}$ be the nth recursively enumerable set of integers, that is, $S_{[n]} = \{x\colon \exists y\, T(n, x, y)\}$. A set S of integers is *productive* if there exists a partially recursive *production function* $f(x)$ such that for each n

$$[S_{[n]} \subset S] \Rightarrow [f(n) \in S] \,\&\, [f(n) \notin S_{[n]}].$$

A set R of integers is *creative* if R is recursively enumerable and $\bar{R}$ is productive.

Theorem 11.22A. No recursively enumerable set can be productive.

Proof. Let S be a recursively enumerable set. Then there exists a number m such that $S = S_{[m]}$. By choosing the value m for the variable n in Definition 11.14A we get the contradiction

$$[S_{[m]} \subset S_{[m]}] \Rightarrow [f(m) \in S_{[m]}] \& [f(m) \notin S_{[m]}]$$

to the assumption that S is productive.

Theorem 11.23A. No recursive set can be creative.

Proof. Let S be a recursive set. Then $\bar{S}$ is recursively enumerable, that is, $\bar{S}$ cannot be productive (cf. Theorem 11.22A). Hence S is not creative.

Theorem 11.24A. The set $K = \{x: \exists y\, T(x, x, y)\}$ is creative.

Proof. The form of K reveals that K is recursively enumerable. Furthermore, $\bar{K}$ is productive with the recursive production function $f(x) = x$ for all x, that is, $(S_{[n]} \subset \bar{K}) \Rightarrow (n \in \bar{K}) \& (n \notin S_{[n]})$. This is demonstrated as follows. First $n \in \bar{K} \equiv n \notin S_{[n]}$ because

$$(n \in \bar{K}) \equiv \neg \exists y T(n, n, y) \equiv n \notin S_{[n]}.$$

Next

$$\begin{aligned}
[S_{[n]} \subset \bar{K}] &\equiv [\forall x: (x \in S_{[n]}) \Rightarrow (x \in \bar{K})] \\
&\equiv [\forall x: \exists y T(n, x, y) \Rightarrow \neg \exists y T(x, x, y)] \\
&\Rightarrow [\exists y T(n, n, y) \Rightarrow \neg \exists y T(n, n, y)] \\
&\equiv [\neg \exists y T(n, n, y)] \\
&\equiv [n \in \bar{K}].
\end{aligned}$$

Theorem 11.25A. Let Q be a productive set. Then Q contains arbitrarily large recursively enumerable subsets.

Argument. Since Q is productive, there exists a recursive production function $f(x)$ such that $(S_{[x]} \subset Q) \Rightarrow (f(x) \in Q) \& (f(x) \notin S_{[x]})$. Choose a recursively enumerable subset $S_{[n]}$ of Q. Next construct the set $S_{[n]} \cup \{f(n)\}$ and determine the code number, n_1, of the Turing machine which enumerates it. This determination is done by a recursive function, $\varphi(n) = n_1$, obtained from Tm. 11.16A, p. 385:

$$\exists y[T(n, x, y) \vee (x = f(n))] \equiv \exists y\, T(\varphi(n), x, y).$$

Repeating this process, we get an unlimited sequence of larger and larger recursively enumerable subsets of Q such that $S_{[n]} \subset S_{[\varphi(n)]} \subset S_{[\varphi(\varphi(n))]} \subset \cdots$.

Theorem 11.26A. The set $S = \{z: \forall x\, \exists y\, T(z, x, y)\}$ is productive.

Proof. Let $\psi_{f(z)}$ be a recursive function which enumerates the nonempty, recursively enumerable set $S_{[z]}$, that is, $S_{[z]} = \{\psi_{f(z)}(x): x \in N\}$. Notice that there is an effective method of constructing a Turing machine which computes the recursive enumerating function $\psi_{f(z)}(x)$ for any given z (compare the proof of Tm. 11.18A, p. 386). Next define a partial recursive function $\psi_{g(z)}(x)$ as follows: $\psi_{g(z)}(x) = \psi_{\psi_{f(z)}(x)}(x) + 1$. Let us argue that the function $g(z)$ thus defined is a partial recursive production function (cf. Definition 11.14A) such that $[S_{[z]} \subset S] \Rightarrow [g(z) \in S] \ \& \ [g(z) \notin S_{[z]}]$, which indicates that S is productive. Hence let us assume that $S_{[z]}$ is a nonempty recursively enumerable set such that $S_{[z]} \subset S$. Then $\psi_{f(z)}(x) \in S$ for every x, that is, $\psi_{f(z)}(x)$ is the code number of a recursive function. Hence $\psi_{\psi_{f(z)}(x)}(x) + 1$ is effectively computable for every x, that is, $g(z)$ is the code number of a recursive function, that is, $g(z) \in S$. Furthermore, $g(z)$ itself is partial recursive by Church's thesis. Finally, we must have $g(z) \notin S_{[z]}$, because otherwise there is an n such that $g(z) = \psi_{f(z)}(n)$, that is, $\psi_{\psi_{f(z)}(n)}(x) = \psi_{\psi_{f(z)}(x)}(x) + 1$, which yields a contradiction for $x = n$.

I.13. UNIVERSAL TURING MACHINES

As we know, $g(z, x) = U(\mu y\, T(z, x, y))$ is a partial recursive function, that is, there is a Turing machine, say U_0, which computes $g(z, x)$, that is, $\psi_{u_0}(z, x) = g(z, x)$. This machine is "universal" in the sense that it can compute any partial recursive function $f(x)$. This is immediately clear, for there is a code number z to each partial recursive function $f(x)$, that is, $f(x) = \psi_z(x)$, and obviously $f(x) = \psi_z(x) = \psi_{u_0}(z, x)$. We may regard $\langle z, x \rangle$ as a program which makes U_0 compute $\psi_z(x)$.

Let U_1 be another "universal" Turing machine such that $\psi_{u_1}(\varphi(z, x)) = \psi_z(x)$, that is, U_1 computes ψ_z from the program $\varphi(z, x)$. Now, how complex may the programming function $\varphi(z, x)$ be? If we allow it to be as complex as the function which is to be computed, we could have $\psi_{u_1}(\varphi(z, x)) = \varphi(z, x)$, that is, ψ_{u_1} would be the identity function. But there would be little sense in calling such a machine U_1 (a pair of wires, e.g.) universal, if the work of programming U_1 would require another universal and more complex Turing machine.

There are several reasonable definitions of universal Turing machines which meet with intuitive requirements concerning the relative complexities of the computation and the programming (cf. Davis [3, 4] and Rogers [20]). Rogers' definition may be stated as follows.

Definition 11.15A. A Turing machine U is *universal* if there exists a recursive function φ such that, for all z, $\psi_z(x) = \psi_u(\varphi(z, x))$. [For the variable x may be substituted an n-tuple $\langle x_1, x_2, \ldots, x_n \rangle$.]

In order to argue the plausibility of this definition we will prove the following theorems.

Theorem 11.27A. The Turing machine U_0 which computes $f(z, x) = U(\mu y T(z, x, y))$ is universal.

Proof. Let φ be the recursive pair identity function $\varphi(z, x) = \langle z, x \rangle$ [computed by a machine with, e.g., the single quadruple $\langle q_0, B, B, q_1 \rangle$; this machine leaves the tape expression $1^{z+1}B1^{x+1}$ as it is]. Obviously

$$\psi_{u_0}(\varphi(z, x)) = f(z, x) = \psi_z(x),$$

that is, U_0 is universal.

Theorem 11.28A. The function computed by a universal Turing machine is partial recursive and not recursive.

Proof. Assume that there exists a universal Turing machine which computes a recursive function ψ_u. Then, by Definition 11.15A, $\psi_z(x) = \psi_u(\varphi(z, x))$ would be recursive for each z. But this is false. For example, the function $\psi_{z_1}(x) = \mu y(x + y = 0)$, which is computable by a Turing machine Z_1, is partial recursive but not recursive.

Theorem 11.28A indicates a characteristic difference in the computation function ψ of a universal Turing machine and its programming function φ: ψ_u is not recursive whereas φ is. A difference on a deeper level is revealed by the following theorem.

Theorem 11.29A. Let U be a universal Turing machine. Then its domain $D_{[u]}$ is creative.

Proof. According to Definition 11.15A, we have

$$D_u = \{\varphi(z, x) : \exists y\, T(z, x, y)\}.$$

Since $D_{[z]} = \{x : \exists y\, T(z, x, y)\}$, we have for all x and z,

$$x \in D_{[z]} \equiv \varphi(z, x) \in D_{[u]},$$

where φ, according to Definition 11.15A, is recursive. Suppose that $D_{[n]} \subset \bar{D}_{[u]}$ (compare Def. 11.14A, p. 387), that is, that $\exists y\, T(n, x, y) \Rightarrow x \notin D_u$. Then

$$\exists y\, T(n, \varphi(z, x), y) \Rightarrow \varphi(z, x) \notin D_{[u]}$$
$$\Rightarrow x \notin D_{[z]} \Rightarrow \neg \exists y\, T(z, x, y).$$

This is true for all z and particularly for $z = x$, that is, $\exists y\, T(n, \varphi(x, x), y) \Rightarrow \neg \exists y\, T(x, x, y)$. Because $\varphi(x, x)$ is a recursive function of x, $T(n, \varphi(x, x), y)$ is

recursive in n, x, and y. By Tm. 11.10A, p. 381, $\exists y\, T(n, \varphi(x, x), y)$ is semi-computable, and hence by Tm. 11.16A, p. 385, there exists a recursive function $f(n)$ such that $\exists y\, T(n, \varphi(x, x), y) \equiv \exists y T(f(n), x, y)$. Hence, for each x, $\exists y\, T(f(n), x, y) \Rightarrow \neg \exists y\, T(x, x, y)$. In particular,

$$\exists y\, T(f(n), f(n), y) \Rightarrow \neg \exists y\, T(f(n), f(n), y).$$

Hence $\neg \exists y\, T(f(n), f(n), y)$, that is, $\neg \exists y\, T(n, \varphi(f(n), f(n)), y)$ that is, $\varphi(f(n), f(n)) \notin D_{[n]}$. Furthermore, $\neg \exists y\, T(f(n), f(n), y)$ implies that $f(n) \notin D_{[f(n)]}$. Recalling the identity $x \in D_{[z]} \equiv \varphi(z, x) \in D_{[u]}$ from the beginning of the proof, we conclude that $\varphi(f(n), f(n)) \notin D_{[u]}$. We have thus seen that there exists a recursive function $g(n) = \varphi(f(n), f(n))$ such that

$$(D_{[n]} \subset \bar{D}_{[u]}) \Rightarrow (g(n) \in \bar{D}_{[u]}) \;\&\; (g(n) \notin D_{[n]}),$$

that is, $\bar{D}_{[u]}$ is productive by Definition 11.14A. Furthermore, $D_{[u]}$, being the domain of a partially recursive function, is recursively enumerable. Hence, by Def. 11.14A, p. 387, $D_{[u]}$ is creative.

I.14. RECURSION THEOREMS

The Kleene recursion theorems (Theorems 11.30A and 11.31A) are central in recursive function theory. In particular, they provide methods for handling with elegance many self-reference problems that otherwise would require extensive, complex treatment.

Theorem 11.30A (Kleene's recursion theorem). If $g(y, x_1, x_2, \ldots, x_n)$ is a partial recursive function, then there is a number e such that

$$\psi_e(x_1, x_2, \ldots, x_n) = g(e, x_1, x_2, \ldots, x_n).$$

Proof. Let $g(y, x_1, x_2, \ldots, x_n)$ be partial recursive. Then

$$g(f(y, y), x_1, x_2, \ldots, x_n),$$

where f is the recursive function of Tm. 11.15A, p. 384, is partial recursive and hence computable by a Turing machine Z_0. Let z_0 be the code number of Z_0. Then, by Theorem 11.15A,

$$\begin{aligned}
g(f(y, y), x_1, x_2, \ldots, x_n) &= \psi_{z_0}(y, x_1, x_2, \ldots, x_n) \\
&= \psi_{f(z_0, y)}(x_1, x_2, \ldots, x_n).
\end{aligned}$$

This is true for all y and particularly for $y = z_0$, that is,

$$g(f(z_0, z_0), x_1, x_2, \ldots, x_n) = \psi_{f(z_0, z_0)}(x_1, x_2, \ldots, x_n).$$

Hence $f(z_0, z_0)$ is an e-number, which proves the theorem.

Theorem 11.31A (Second form of the Kleene recursion theorem). Let $f(x)$ be a recursive function. Then there exists a number n such that $\psi_n(x) = \psi_{f(n)}(x)$.

Proof. Let us assume that $f(x)$ is recursive. Then $\varphi(z, x) =_{\text{def}} \psi_{f(z)}(x)$ is partial recursive. By Theorem 11.30A there is a number, n, such that

$$\varphi(n, x) = \psi_n(x) = \psi_{f(n)}(x).$$

Recalling that $\psi_{u_0}(z, x) = \psi_z(x)$, where u_0 is the code number of the particular universal Turing machine, U_0, of Tm. 11.27A, p. 390, we conclude that for each partial recursive function $g(y, x)$ there exists a number e such that $\psi_{u_0}(e, x) = \psi_e(x) = g(e, x)$. With τ for $\langle e, x \rangle$ we have that $\psi_{u_0}(\tau) = g(\tau)$. We shall see that this result holds for any universal Turing machine.

Theorem 11.32A. With each universal Turing machine U and each partial recursive function $g(x)$ there exists an argument τ such that $\psi_u(\tau) = g(\tau)$.

Proof. Let U be a universal Turing machine. By Def. 11.15A, p. 389, we have, for all z and x, $\psi_z(x) = \psi_u(\varphi(z, x))$. Let $g(x)$ be an arbitrary partial recursive function. Since $\varphi(z, x)$ is recursive, the composition $h(z, x) = g(\varphi(z, x))$ will be partial recursive. By Theorem 11.30A, there exists an e such that $h(e, x) = \psi_e(x)$. Hence $g(\varphi(e, x)) = h(e, x) = \psi_e(x) = \psi_u(\varphi(e, x))$. Hence there exists an argument $\tau = \varphi(e, x)$ such that $f(\tau) = \psi_u(\tau)$.

If the function $g(x)$ of Theorem 11.32A is partial recursive and not recursive, it can happen that the argument which is guaranteed by the theorem does not belong to the domain of g and hence is not in the domain of ψ_u.

Theorem 11.33A. With each universal Turing machine U, and each recursive function $g(x)$, there exists an argument τ in the domain of ψ_u such that $\psi_u(\tau) = g(\tau)$.

Proof. Since $g(x)$ is recursive, it is defined for all integers and hence for the particular argument τ which is guaranteed by Theorem 11.32A. Therefore, by Theorem 11.32A, this τ must also be in the domain of the partial recursive function ψ_u.

Let us indicate the power of Theorem 11.33A by applying it twice in the following example, which constitutes a sharper version of Tm. 11.13A, p. 382.

Example 11.6A. Let $f_e(x)$ be an extension to the domain N of a partial recursive function $f(x)$. Hence $f_e(x) = f(x)$ whenever $f(x)$ is defined; $f_e(x)$ may be defined in an arbitrary way for those x which are not in the domain of $f(x)$. Then there is a partial recursive function $f(x)$, with a nonfinite domain, such that no extension $f_e(x)$ is recursive.

Let us see that $\psi_u(x)$, where u is the code number of a universal Turing machine, qualifies as a function $f(x)$ of the example. First of all, we can conclude from Theorem 11.33A that the domain of $\psi_u(x)$ is nonfinite. There are an infinite number of recursive constant functions, $g_i(x) = i$. Hence, Theorem 11.33A assures that for each i there is a τ such that $\psi_u(\tau) = g_i(\tau) = i$, that is, the range of ψ_u is N. Hence $\psi_u(x)$ must also have a nonfinite domain. It follows that $g(x) = _\mathrm{def}\, \psi_u(x) + 1$ is a partial recursive function with an infinite domain. No extension $g_e(x)$ of $g(x)$ can be recursive, however, for otherwise Theorem 11.33A would imply that there exists a τ in the domain of $\psi_u(x)$, that is, also in the domain of $g(x)$, such that $\psi_u(\tau) = g_e(\tau) = \psi_u(\tau) + 1$, which is false.

I.15. STRONG RECURSIVE REDUCIBILITY

A nonrecursive set may have very specific nonrecursive properties, which would not be revealed if we simply classified it as nonrecursive. Ways of classifying nonrecursive sets are offered by the various reducibility relations (cf. Davis [5] and Rogers [20]). Here it will suffice to present only one of these, namely, strong recursive reducibility.

Definition 11.16A. Let A and B be sets. Then A is *strongly recursively reducible* (or *strongly reducible*) to B, denoted $A \ll B$, if there exists a recursive function $f(x)$, the *reducibility function*, such that $[x \in A] \equiv [f(x) \in B]$.

Theorem 11.34A. The strong reducibility relation is reflexive, that is, $A \ll A$.

Proof. Obviously there exists a recursive function $f(x)$ such that $[x \in A] \equiv [f(x) \in A]$; $f(x) = _\mathrm{def} x$ is such a function.

Theorem 11.35A. The strong reducibility relation is transitive, that is, $A \ll B$ and $B \ll C$ implies that $A \ll C$.

Proof. Assume that $A \ll B$ and $B \ll C$. Then there are recursive functions $f(x)$ and $g(x)$ such that $[x \in A] \equiv [f(x) \in B]$ and $[x \in B] \equiv [g(x) \in C]$. Hence $[x \in A] \equiv [f(x) \in B] \equiv [g(f(x)) \in C]$, that is, there exists a recursive function $h(x) = _\mathrm{def} g(f(x))$ such that $[x \in A] \equiv [h(x) \in C]$.

Theorem 11.36A. $A \ll B$ if and only if $\bar{A} \ll \bar{B}$.

Proof. $A \ll B$ if and only if there is a recursive function $f(x)$ such that $[x \in A] \equiv [f(x) \in B]$, which can be written equivalently as $[x \notin A] \equiv [f(x) \notin B]$, that is, $[x \in \bar{A}] \equiv [f(x) \in \bar{B}]$. Hence $[A \ll B] \equiv [\bar{A} \ll \bar{B}]$.

Theorem 11.37A. A set S is recursive if and only if $S \ll \{1\}$.

Proof. Assume that S is recursive, that is, its characteristic function $C_S(x)$ is recursive and $[x \in S] \equiv [C_S(x) \in \{1\}]$. Hence $S \ll \{1\}$. Conversely, assume that $S \ll \{1\}$, that is, there is a recursive function $f(x)$ such that $[x \in S] \equiv [f(x) \in \{1\}] \equiv [f(x) = 1]$. Hence S is the extension of the computable predicate $P(x) =_{\text{def}} [f(x) = 1]$, that is, S is recursive.

Theorem 11.38A. A set S is recursively enumerable if and only if $S \ll K$, where $K = \{x: \exists y\, T(x, x, y)\}$ (cf. Tm. 11.20A, p. 387).

Proof. Assume $S \ll K$, that is, there is a recursive function $f(x)$ such that $S = \{x: f(x) \in K\} = \{x: \exists y\, T(f(x), f(x), y)\}$. Then, by Tm. 11.10A, p. 381, S is recursively enumerable. Conversely, assume that S is recursively enumerable. Then, by Tm. 11.9A, p. 381, there is a recursive predicate, $R(x, y)$ such that $S = \{x: \exists y\, R(x, y)\}$. By Tm. 11.16A, p. 385, there is a recursive function $g(x)$ such that

$$\exists y[R(x, y)\ \&\ (z = z)] \equiv \exists y\, T(g(x), z, y).$$

Hence

$$S = \{x: \exists y\, R(x, y)\} = \{x: \exists y[R(x, y)\ \&\ (g(x) = g(x))]\}$$
$$= \{x: \exists y\, T(g(x), g(x), y)\} = \{x: g(x) \in K\}.$$

Hence $S \ll K$.

Theorem 11.39A. Let A and B be two recursive sets such that each of the sets A, B, $\bar{A}$, $\bar{B}$ is nonempty. Then $A \ll B$ and $B \ll A$.

Proof. Since A and B are recursive, their characteristic functions, $C_A(x)$ and $C_B(x)$, are recursive. Hence $f(x) =_{\text{def}} \mu y[C_A(x) = C_B(y)]$ is partial recursive. Furthermore, since both B and $\bar{B}$ are nonempty, $C_B(y)$ is a mapping onto $\{0, 1\}$, that is, $f(x)$ is a total function. Hence, $f(x)$ is a recursive function such that $[x \in A] \equiv [f(x) \in B]$, that is, $A \ll B$. In the same way $B \ll A$.

Theorem 11.40A. Let A and B be any two given sets of integers such that A, but not B, is recursive. (Since B is nonrecursive, it has to be given in a less constructive way than A, for which we know every element and every non-element. Let us assume that B is given in such a way that we know at least one of its elements, a, and at least one of its nonelements, b.) Then $A \ll B$ and $B \ll A$.

Proof. Since A is recursive, its characteristic function, $C_A(x)$, is recursive. Hence $f(x) =_{\text{def}} a \cdot C_A(x) + b \cdot (1 \dot- C_A(x))$ is recursive and $[x \in A] \equiv [f(x) \in B]$, that is, $A \ll B$. Suppose that $B \ll A$. Then there is a recursive function $g(x)$ such that $C_B(x) = C_A(f(x))$, contradicting the fact that $C_B(x)$ is nonrecursive. Hence $B \ll A$.

Theorems 11.39A and 11.40A show that the recursive sets are strongly reducible to the nonrecursive sets, and that the converse does not hold.

Although every pair of recursive sets (except $\emptyset$ and N) are strongly reducible to each other, the same does not hold for the recursively enumerable, nonrecursive sets. It is true that there is one recursively enumerable set, the creative set K, such that every recursively enumerable set is strongly reducible to K (Theorem 11.38A). There are, however, nonrecursive sets S such that $K \ll S$ and $S \ll K$ (cf. Davis [5]). Hence, the strong reducibility relation reveals distinct nonrecursive properties already existing among the recursively enumerable sets, that is, among the theoremhoods of r-formal theories (cf. Explanation Hypothesis II, p. 350).

APPENDIX II. FORMAL SYSTEMS

A basic thought behind the formalization of a theory is to make it susceptible to exact mathematical study. For example, in certain formal theories (r-formal theories), there need never arise a dispute about whether or not an alleged proof really is a proof. The proof may be difficult to follow, but it can always be effectively checked by a Turing machine.

II.1. FORMAL THEORIES

In a formal theory, the so-called well-formed formulas (abbreviated wffs) are the basic objects. A wff is a finite string of symbols from a certain alphabet. Invariably, the wffs are specified in such a way that an effective procedure exists for deciding whether or not a given string is well formed.

Although the English language is not a formal system, we may compare a grammatical English sentence with a wff of a formal theory. For example, "Snoopy is a dog and" is not a grammatical sentence [cf. "D(s) &" which is not a wff of a predicate calculus]. The sentence "Snoopy is a dog and Anne is a man" is grammatically correct [cf. "D(s) & M(a)," which is a wff], although it is false. By the rules of English grammar, we can effectively decide whether or not a given sentence is grammatically correct (cf. well formed). However, we may not be able to decide effectively whether a given grammatical sentence is true or false.

In order to apply the concepts of computability and recursivity, we must encode the formulas of a formal theory to integers. In this way a set S of wffs is called recursive if the set S^* of code numbers of the elements of S is recursive. Also a function $f(X)$, which maps wffs on wffs, is called recursive

if the function $f^*(x)$ is partial recursive, where $f^*(x)$ has as its domain the set of all code numbers of the wffs and $f^*(c(X)) = c(f(X))$, where $c(X)$ is the code number of the wff X.

Definition 11.17A. A *formal theory* $\mathscr{S}$ is a set A of wffs, called the *axioms* of $\mathscr{S}$, together with a set of predicates, the *rules of inference* of $\mathscr{S}$. When $R(Y, X_1, X_2, \ldots, X_n)$ is a rule of inference of $\mathscr{S}$, the wff Y is said to be a *consequence of the wffs* $X_1, X_2, \ldots, X_n$ in $\mathscr{S}$ by R.

Although many authors define a formal theory according to Definition 11.18A below, some others do not want to impose such restrictions on the general concept of a formal theory. Hence we feel it necessary to introduce a special concept r-formal (recursively formal).

Definition 11.18A. A formal theory $\mathscr{S}$ is *r-formal* if $\mathscr{S}$ has a recursive set of axioms and a finite set of recursive rules of inference.

Definition 11.19A. A finite sequence of wffs $X_1, X_2, \ldots, X_n$ is called a *proof in a formal theory* $\mathscr{S}$ if, for each i, $1 \leq i \leq n$, either $X_i \in A$ or there exists $j_1, j_2, \ldots, j_k < i$ such that X_i is a consequence of $X_{j_1}, X_{j_2}, \ldots, X_{j_k}$ by one of the rules of inference of $\mathscr{S}$. Each of the X_i, $i = 1, 2, \ldots, n$, is called a *step of the proof*.

Definition 11.20A. A wff W is a *theorem* of $\mathscr{S}$, or, equivalently, a wff W is *provable in* $\mathscr{S}$, if there exists a proof in $\mathscr{S}$ whose final step is W. Also, $\vdash_{\mathscr{S}} W$ symbolizes that W is a theorem* of $\mathscr{S}$. The set of all theorems of $\mathscr{S}$, the *theoremhood* of $\mathscr{S}$, is denoted $T_{\mathscr{S}}$. Occasionally, as should be clear from the context, $T_{\mathscr{S}}$ may also denote the set of code numbers at all the theorems of $\mathscr{S}$.

Theorem 11.41A. The theoremhood of an r-formal theory is recursively enumerable.

Proof. Let $P_{\mathscr{S}}$ be the set of all code numbers of proofs in an r-formal theory $\mathscr{S}$. Let us first argue that $P_{\mathscr{S}}$ is recursive. For any given code number, x, of a proof, we can effectively decode x to obtain a string of wffs, $X_1, X_2, \ldots, X_n$. Since the set of axioms is recursive, we can effectively decide whether or not X_1 is an axiom. If $X_1 \notin A$, we conclude from Definition 11.19A that $x \notin P_{\mathscr{S}}$. Next, we can effectively determine whether $X_2 \in A$ or whether X_2 is a consequence of X_1 by a rule of inference in $\mathscr{S}$. The reason is that the rules of inference are finite in number, and each is a recursive predicate. Since the string $X_1, X_2, \ldots, X_n$ is finite, we can in this way effectively decide whether it is a proof in $\mathscr{S}$, that is, $P_{\mathscr{S}}$ is recursive. Returning to $T_{\mathscr{S}}$, we have, by Definition 11.20A, $T_{\mathscr{S}} = \{x: \exists z[z \in P_{\mathscr{S}}$ and x is the code number of the last

_____
* ⊢ is a turn-styled T.

element of the sequence whose code number is z]}. Since $P_{\mathscr{S}}$ is recursive, the whole predicate [$z \in P_{\mathscr{S}}$ and x is the code number of the last element of the sequence whose code number is z] is recursive. Hence $T_{\mathscr{S}}$ is recursively enumerable (cf. Tm. 11.17A, p. 385, and Tm. 11.10A, p. 381.)

Definition 11.21A. A formal theory $\mathscr{S}$ is called *consistent* if not every wff is a theorem of $\mathscr{S}$. If not consistent, $\mathscr{S}$ is *inconsistent*.

Definition 11.22A. A formal theory $\mathscr{S}$, with the negation symbol $\neg$, is called ($\neg$)-*consistent* if, for no wff W, we have $\vdash_{\mathscr{S}} W$ and $\vdash_{\mathscr{S}} \neg W$. If $\mathscr{S}$ is not ($\neg$)-consistent, it is called ($\neg$)-*inconsistent*.

Example 11.7A. An r-formal theory $\mathscr{L}$ for the propositional logic may be defined as follows. The alphabet of $\mathscr{L}$ consists of the symbols $\Rightarrow \neg$ ( ) and the propositional variables $p_1, p_2, p_3, \ldots$. Also, W is a wff of $\mathscr{L}$ if there exists a sequence $W_1, W_2, \ldots, W_n$ of formulas such that W_n is W and for each i, $1 \leq i \leq n$, either W_i is a propositional variable, W_i is $(W_j \Rightarrow W_k)$, where $j < i$ and $k < i$, or W_i is $\neg W_j$, where $j < i$. Moreover, $\mathscr{L}$ has only one rule of inference, namely, $R(Y, X_1, X_2)$, which is true if and only if X_2 is $X_1 \Rightarrow Y$. This rule of inference, which is called *modus ponens*, says, in other words, that Y is a consequence of X_1 and of $X_1 \Rightarrow Y$. The axioms of $\mathscr{L}$ are as follows:

A1. $(A \Rightarrow (B \Rightarrow A))$,
A2. $((A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C)))$,
A3. $((\neg A \Rightarrow \neg B) \Rightarrow (B \Rightarrow A))$,

where A, B, and C are any wffs of $\mathscr{L}$.

Example 11.8A. Let us demonstrate that, for any wff W of $\mathscr{L}$, $\vdash_{\mathscr{L}} (W \Rightarrow W)$. The following is a proof of $(W \Rightarrow W)$ in $\mathscr{L}$:

(1) $\vdash_{\mathscr{L}} ((W \Rightarrow ((W \Rightarrow W) \Rightarrow W)) \Rightarrow ((W \Rightarrow (W \Rightarrow W)) \Rightarrow (W \Rightarrow W)))$
(axiom A2 with the wffs W, $(W \Rightarrow W)$, W for A, B, C, respectively);
(2) $\vdash_{\mathscr{L}} (W \Rightarrow ((W \Rightarrow W) \Rightarrow W))$
(axiom A1 with the wffs W, $(W \Rightarrow W)$ for A, B, respectively),
(3) $\vdash_{\mathscr{L}} ((W \Rightarrow (W \Rightarrow W)) \Rightarrow (W \Rightarrow W))$
(a consequence of (2) and (1) by *modus ponens*);
(4) $\vdash_{\mathscr{L}} (W \Rightarrow (W \Rightarrow W))$
(axiom A1 with the wffs W, W for A, B, respectively);
(5) $\vdash_{\mathscr{L}} (W \Rightarrow W)$
(a consequence of (4) and (3) by *modus ponens*).

Theorem 11.42A. Let $\mathscr{S}$ be a formal theory which contains A1 and A3 among its axioms and *modus ponens* among its rules of inference. Then $\mathscr{S}$ is consistent if and only if $\mathscr{S}$ is ($\neg$)-consistent.

Proof. Assume that $\mathscr{S}$ is $(\neg)$-consistent. Then, if A is a theorem of $\mathscr{S}$, $\neg A$ is a wff which is not a theorem of $\mathscr{S}$. Hence $\mathscr{S}$ is consistent. Conversely, assume that $\mathscr{S}$ is consistent and at the same time $(\neg)$-inconsistent. Then there is a wff B such that both $\vdash_{\mathscr{S}} B$ and $\vdash_{\mathscr{S}} \neg B$. Let W be an arbitrary wff of $\mathscr{S}$. By axiom A1 we have $\vdash_{\mathscr{S}} (\neg B \Rightarrow (\neg W \Rightarrow \neg B))$. Hence, by *modus ponens*, $\vdash_{\mathscr{S}} (\neg W \Rightarrow \neg B)$. Furthermore, $\vdash_{\mathscr{S}} ((\neg W \Rightarrow \neg B) \Rightarrow (B \Rightarrow W))$ by axiom A3. Hence, applying *modus ponens* twice, we get $\vdash_{\mathscr{S}} W$. This means that every wff W of $\mathscr{S}$ is a theorem of $\mathscr{S}$, contradicting the assumption that $\mathscr{S}$ is consistent. Hence, if $\mathscr{S}$ is consistent, it cannot be $(\neg)$-inconsistent but must be $(\neg)$-consistent.

Further examples of theories for which Theorem 11.42A applies are the so-called first-order theories. A first-order theory contains among its symbols the quantifiers $\forall$ and $\exists$, operating on individual variables. Among these theories is the pure first-order predicate calculus, which has logical but no proper axioms. The importance of this theory stems from the fact that every r-formal theory can be translated into it (see [5], [9], [16]).

II.2. THE DECISION PROBLEM FOR A FORMAL THEORY

Although the set of proofs of an r-formal theory is always recursive, its set of theorems, although always recursively enumerable, need not be recursive.

Definition 11.23A. The *decision problem for a formal theory* $\mathscr{S}$ is the problem of determining, for a given wff, whether or not it is a theorem in $\mathscr{S}$. The decision problem is *recursively solvable* if $T_{\mathscr{S}}$ is recursive; otherwise, it is *recursively unsolvable*.

Let us give an example showing that there are r-formal theories with recursively unsolvable decision problems.

Example 11.9A. As we know (Tm. 11.20A, p. 387, and Tm. 11.24A, p. 388), the set $K = \{x : \exists y\, T(x, x, y)\}$ is recursively enumerable but not recursive. Let us try to use K for constructing an r-formal, undecidable theory by associating the proof sequences with the computation sequences y of the recursive Turing machine predicate $T(z, x, y)$ [cf. Tm. 11.6A, p. 378; $T(z, x, y)$ is true when Turing machine Z, started on the argument x, performs a computation sequence $Y_1, Y_2, \ldots, Y_n$ whose code number is y]. To interpret $T(z, x, y)$ as "y is (the code number of) a proof of $q_0 \bar{x}$ in the theory z" would be unconventional in the sense that the computation sequence y begins, not ends, with x. Hence it is more natural to consider the reverse computation sequence as a proof of x from the axiom Y_n. In order to simplify the axioms, let us consider

the associate Z^* to an arbitrary Turing machine Z; Z^* is a Turing machine which behaves precisely as Z, except that Z^* halts in state q_h, scanning the number 0 on an otherwise blank tape, if and only if Z halts. It is left as an exercise to prove that the code number z^* of Z^* can be effectively obtained from z, that is, $z^* = f(z)$, where f is recursive. We can now interpret the recursive predicate $T(f(z), x, y)$ as "y is a proof of $q_0 \bar{x}$ from the axiom $q_h \bar{0}$ in a system whose rules are the inverse quadruples of Turing machine Z^*." Let F_z denote the formal system which is defined as follows. The set of wffs is the recursive set of the Turing machine tape configurations. There is a single axiom, the tape configuration $q_h \bar{0}$. There are a finite number of recursive rules of inference, $R_i(Y, X)$, such that Y is a consequence of X according to R_i if and only if tape configuration Y forces Turing machine Z^* over into tape configuration X (by the ith quadruple defining Z^*). According to Def. 11.18A, p. 396, the formal system F_z is r-formal for every code number z. Furthermore, $q_0 \bar{x}$ is a theorem of F_z if and only if there is a proof of $q_0 \bar{x}$, that is, there is a computation sequence from $q_0 \bar{x}$ to $q_h \bar{0}$, that is, Turing machine Z halts [i.e., $\exists y\, T(z, x, y)$]. Finally, by Theorem 11.9A, p. 387, there is a z_0 such that $\exists y\, T(x, x, y) \equiv \exists y\, T(z_0, x, y)$. Therefore

$$\{x: \exists y\, T(z_0, x, y)\}$$

is nonrecursive, and hence the decision problem for the r-formal theory F_{z_0} (does $q_0 \bar{x}$ belong to the theoremhood of F_{z_0}?) is recursively unsolvable.

II.3. TRANSLATABILITY BETWEEN FORMAL THEORIES

Let $\mathscr{L}_1$ and $\mathscr{L}_2$ be two natural languages. It may happen that $\mathscr{L}_1$ is richer that $\mathscr{L}_2$ so that an attempt to translate from $\mathscr{L}_1$ into $\mathscr{L}_2$ will not fully mirror the fine details or convey the complete contents of a sentence in $\mathscr{L}_1$. Again, it may be impossible to retranslate a sentence from $\mathscr{L}_2$ back into the original $\mathscr{L}_1$-sentence without having the "translator" add to the contents.

In the same way, translatability (Davis [5]) between formal theories should be neutral with respect to powers of effectivity. It should neither add to the deductive power of a theory nor destroy it.

Definition 11.24A. Let $\mathscr{S}$ and $\mathscr{S}'$ be formal theories. Then $\mathscr{S}$ is *translatable* into $\mathscr{S}'$ if there exists a recursive function $f(X)$ such that $\vdash_{\mathscr{S}} X$ if and only if $\vdash_{\mathscr{S}'} f(X)$; moreover, if $X \neq Y$, we also have $f(X) \neq f(Y)$.

Theorem 11.43A. $\mathscr{S}$ is translatable into $\mathscr{S}'$ if and only if $T_{\mathscr{S}}$ is strongly recursively reducible to $T_{\mathscr{S}'}$ and the reducibility function is 1-1.

Proof. Let x be the code number of the wff X. Then $[\vdash_{\mathscr{S}} X] \equiv [x \in T_{\mathscr{S}}]$ and $[\vdash_{\mathscr{S}} f(X)] \equiv [f^*(x) \in T_{\mathscr{S}'}]$, where $f^*(x)$ is a recursive function with the code number of $f(X)$ as value. Hence, $\mathscr{S}$ is translatable into $\mathscr{S}'$ if and only if $[x \in T_{\mathscr{S}}] \equiv [f^*(x) \in T_{\mathscr{S}'}]$, where the recursive function $f^*(x)$ is 1-1. This means that $T_{\mathscr{S}} \ll T_{\mathscr{S}'}$ with a reducibility function which is 1-1 (cf. Def. 11.16A, p. 393).

Theorem 11.44A. If $\mathscr{S}$ is translatable into $\mathscr{S}'$ and the decision problem for $\mathscr{S}'$ is recursively solvable, then the decision problem for $\mathscr{S}$ is recursively solvable. If the decision problem for $\mathscr{S}$ is recursively unsolvable, then so is it for $\mathscr{S}'$.

Proof. Suppose that $\mathscr{S}$ is translatable into $\mathscr{S}'$ and that $T_{\mathscr{S}'}$ is recursive. Then, by Theorems 11.43A, 11.19A, p. 386, and 11.38A, p. 394, $T_{\mathscr{S}} \ll T_{\mathscr{S}'} \ll K$, that is, $T_{\mathscr{S}} \ll K$ by Theorem 11.35A. Furthermore, $\overline{T}_{\mathscr{S}'}$ is recursive. Hence, by Theorem 11.36A, $\overline{T}_{\mathscr{S}} \ll \overline{T}_{\mathscr{S}'} \ll K$, that is, $\overline{T}_{\mathscr{S}} \ll K$. Thus, by Theorems 11.37A, p. 393, and 11.19A, p. 386, $T_{\mathscr{S}}$ is recursive, that is, the decision problem for $\mathscr{S}$ is recursively solvable. On the other hand, if we assume that $T_{\mathscr{S}}$ is not recursive, and that $T_{\mathscr{S}} \ll T_{\mathscr{S}'}$, then $T_{\mathscr{S}'}$ must also be nonrecursive. For, if $T_{\mathscr{S}'}$ were recursive, we would have both $T_{\mathscr{S}} \ll T_{\mathscr{S}'} \ll K$ and $\overline{T}_{\mathscr{S}} \ll \overline{T}_{\mathscr{S}'} \ll K$, that is, $T_{\mathscr{S}}$ would be recursive, which contradicts our assumption (cf. Tm. 11.40A, p. 394, for another proof).

II.4. INCOMPLETENESS AND UNSOLVABILITY IN FORMAL THEORIES

To provide a measure of the deductive power of a formal theory, the following definitions (Davis [5]) of completeness are appropriate. This concept is very strongly associated with the concepts of strong recursive reducibility (Def. 11.16A, p. 393), and translatability (Definition 11.24A).

Definition 11.25A. A formal theory $\mathscr{S}$ is said to be *semicomplete with respect to a set of integers* Q if there exists a recursive function f such that $Q = \{n: f(n) \in T_{\mathscr{S}}\}$. Also, $\mathscr{S}$ is said to be *complete with respect to* Q if it is semicomplete with respect to both Q and $\overline{Q}$.

Theorem 11.45A. A formal theory $\mathscr{S}$ is semicomplete with respect to a set Q if and only if $Q \ll T_{\mathscr{S}}$.

Proof. The statement follows immediately from Def. 11.16A, p. 393, and Def. 11.25A.

Theorem 11.46A. $\mathscr{S}$ is semicomplete with respect to a nonrecursive set if and only if $\mathscr{S}$ has a recursively unsolvable decision problem.

Proof. Assume that $\mathscr{S}$ is semicomplete with respect to a nonrecursive set Q, that is, $Q \ll T_{\mathscr{S}}$ and either $Q \ll K$ or $\bar{Q} \ll K$. Then $T_{\mathscr{S}}$ must be nonrecursive, for otherwise $T_{\mathscr{S}} \ll K$ and $\bar{T}_{\mathscr{S}} \ll K$, that is, both $Q \ll K$ and $\bar{Q} \ll K$, which contradicts the assumption. Conversely, assume that $\mathscr{S}$ has a recursively unsolvable decision problem, that is, $\bar{T}_{\mathscr{S}}$ is nonrecursive. Then, because of Tm. 11.37A, p. 393, $\mathscr{S}$ is semicomplete with respect to the nonrecursive set $\bar{T}_{\mathscr{S}}$ itself.

Theorem 11.47A. If $\mathscr{S}$ is semicomplete with respect to every recursively enumerable set, then $\mathscr{S}$ has a recursively unsolvable decision problem.

Proof. Assume that $\mathscr{S}$ is semicomplete with respect to every recursively enumerable set. Then $\mathscr{S}$ is semicomplete with respect to the set K, which is nonrecursive (Tm. 11.20A, p. 387). Hence, by Theorem 11.46A, $\mathscr{S}$ has a recursively unsolvable decision problem.

(A converse of Theorem 11.47A, namely, that if $T_{\mathscr{S}}$ is nonrecursive, $\mathscr{S}$ is also semicomplete with respect to every recursively enumerable set, is false.)

Theorem 11.48A. Let $\mathscr{S}$ be an r-formal theory which is semicomplete with respect to a set Q. Then Q is recursively enumerable.

Proof. Let $\mathscr{S}$ be as in the statement of the theorem. Then, by Tm. 11.41A, p. 396, $Q \ll T_{\mathscr{S}} \ll K$, that is, $Q \ll K$ and Q is recursively enumerable by Tm. 11.38A, p. 394.

Theorem 11.48A has the following obvious corollaries.

Corollary 11.1A. If Q is not recursively enumerable, then no r-formal theory is semicomplete with respect to Q.

Corollary 11.2A. If an r-formal theory $\mathscr{S}$ is complete with respect to a set Q, then Q is recursive.

Corollary 11.3A. If Q is recursively enumerable but not recursive, then no r-formal theory is semicomplete with respect to $\bar{Q}$.

Corollary 11.4A. If Q is recursively enumerable but not recursive, then no r-formal theory is complete with respect to Q.

Corollary 11.1A can be regarded as an abstract form of Gödel's famous incompleteness theorem (see [6]). An interpretation of this theorem is given in the following example.

Example 11.10A. In any r-formal theory which has a certain minimal complexity and for which the notion of " true " wff can be defined in a certain natural way, the set of " true " wffs is productive (cf. Rogers [20]). By Tm. 11.22A, p. 388, there is no r-formal theory which is semicomplete with respect

to this set of "true" wffs. Hence there is no r-formal theory with an explicatory power large enough to explain all these "true" wffs (cf. Explanation Hypothesis II, p. 350). The productivity of a set of "true" wffs may be exemplified as follows.

Consider an r-formal theory that is flexible enough to make assertions about Turing machines as arithmetical statements about their code numbers. Statements like "0 is the code number of a recursive function," and "1 is the code number of a recursive function," will be expressible within the theory. From Tm. 11.26A, p. 388, we know that the set $\{z : \psi_z$ is recursive$\}$ is productive. Hence the set of true statements of the form "n is the code number of a recursive function" is productive. Then, as we know from Corollary 11.1A, there is no r-formal theory which will produce *all true* statements and *no false* statements of this form as theorems. Indeed, an r-formal theory that produces only true statements of the above form can be effectively used to produce a new true statement (cf. the production function of Def. 11.14A, p. 387). Thus all these produced true statements can be produced by a single r-formal theory (see the proof of Tm. 11.25A, p. 388), which again has to be incomplete. No effective continuation of this process will yield an r-formal theory that produces all true, but no false, statements.

As another example, there is no r-formal theory of arithmetic: the Peano axiomatization of arithmetic can only capture a fragment of all arithmetical truths.

PROBLEMS

11.1 (Section 11.2). Construct a formal theory $\mathscr{S}$ with a nonfinite set of recursive rules of inference and a recursive set of axioms such that Tm. 11.1, p. 344, is violated. In other words, the problem of deciding whether an alleged proof really is a proof in $\mathscr{S}$ should be recursively unsolvable.

Hint: Consider an r-formal theory with a nonrecursive theoremhood (cf. Ex. 11.9A, p. 398). Use the recursive function which enumerates the theoremhood of trade each theorem, but a single axiom, for a new recursive rule of inference, to obtain a new theory with the desired property.

11.2 (Section 11.3). Let W_1 and W_2 be two wffs in the formal theory $\mathscr{L}$ for the propositional logic (Ex. 11.7A, p. 397), such that $\vdash_{\mathscr{L}} (W_1 \Rightarrow W_2)$. Prove that $I(W_2, \mathscr{L}) \subset I(W_1, \mathscr{L})$ (cf. Def. 11.2, p. 348).

11.3 (Section 11.3). Let W be a theorem of $\mathscr{L}$ (Ex. 11.7A, p. 397). Prove that $I(W, \mathscr{L}) = \phi$ and $I(\neg W, \mathscr{L}) = S_{\text{wff}} - T_{\mathscr{L}}$, where S_{wff} is the set of all wffs of $\mathscr{L}$. Hence, if the *amount* of syntactic information is measured by card $I(W, \mathscr{L})$, the wffs W and $\neg W$ do not in general carry the same amount of syntactic information in $\mathscr{L}$.

11.4 (Section 11.3). Let $\mathscr{S}$ be an r-formal theory with a decidable logical basis $\mathscr{B}$ (cf. Def. 11.23A, p. 398). Prove that the syntactic information of $\mathscr{S}$ with respect to $\mathscr{B}$ is recursively enumerable.

11.5 (Section 11.4). Let A and B be two given finite, nonempty sets of integers such that card $A <$ card B. Prove that $A \ll B$ and $B \ll A$ by exhibiting corresponding reducibility functions (cf. Def. 11.16A, p. 393). Furthermore, prove that A is translatable into B and that the converse does not hold (cf. Def. 11.24A, p. 399, and Tm. 11.43A, p. 399.)

Use the results to compare the explicatory powers of two theories with finite syntactic information.

11.6 (Section 11.4). Let $\mathscr{S}$ be an r-formal theory with a logical basis $\mathscr{B}$ such that, for any wff W, $W \Rightarrow W$ is a logical theorem (a theorem of $\mathscr{B}$). Furthermore, let $\mathscr{S}$ have the following deduction property, D: Let $\mathscr{S}(A)$ denote the theory obtained from $\mathscr{S}$ by appending the wff A as a proper axiom; then W is a theorem of $\mathscr{S}(A)$ if and only if $A \Rightarrow W$ is a theorem of $\mathscr{S}$.

Prove that $\mathscr{S}(A)$ can be translated into $\mathscr{S}$ with the reducibility function $f(X) =_{\text{def}} (A \Rightarrow X)$. Does this particular translation imply that $\mathscr{S}$ has greater explicatory power than $\mathscr{S}(A)$ with respect to $\mathscr{B}$?

Hint: What will be the logical status of the translation of the proper axiom A?

11.7 (Section 11.4). Prove that the translatability relation (Def. 11.24A, p. 399) is transitive, that is, if $\mathscr{S}_1$ is translatable into $\mathscr{S}_2$ and $\mathscr{S}_2$ is translatable into $\mathscr{S}_3$, then $\mathscr{S}_1$ is translatable into $\mathscr{S}_3$. Hence, by Explanation Hypothesis II, p. 350, the property of having a greater explicatory power also defines a transitive relation.

11.8 (Section 11.5). Use of Explanation Hypothesis III, p. 352, raises the question of how to find a shortest set of axioms for a theory. Prove that there are r-formal theories for which the problem of deciding whether or not an axiom can be deleted, without changing the theoremhood, is recursively unsolvable.

Hint: Assume that there are no such r-formal theories and derive a contradiction to the fact that there are r-formal theories with nonrecursive theoremhoods (Ex. 11.9A, p. 398).

11.9 (Section 11.6). Let the set S be defined by $S =_{\text{def}} \{x : \exists z\, s(z, u) = x\}$, that is S is the set of shortest descriptions which make the universal Turing machine U compute the integers. Prove that S is not recursively enumerable.

Hint: Assume that S is enumerated as a recursive function $f(n)$, and obtain a contradiction from Tm. 11.33A, p. 392, by applying it to the function $\psi_u(f(\mu n(f(n) > x)))$.

11.10 (Section 11.6). Let us say of two strings of symbols, z_1 and z_2, that z_2 is the more *complex* (to describe) in relation to U if and only if $s(z_1, u) < s(z_2, u)$. (See the order-randomness hypothesis (p. 354), from which we can

conclude that z_2 is more randomized than z_1, only if we know that z_1 and z_2 are of equal lengths.) Let $c(z)$ be a complexity operation such that

$$s(z, u) < s(c(z), u).$$

Thus c transforms a string z into a more complex string $c(z)$ relative to U.

Prove that the complexity operation $c(z)$ is nonrecursive for every choice of universal Turing machine U.

Hint: Assume that $c(z)$ is recursive. Then the yth composition $c^y(z)$ is recursive, and

$$s(z, u) < s(c(z), u) < s(c(c(z)), u) < s(c^3(z), u) < \cdots < s(c^y(z), u) < \cdots.$$

Prove that this unbounded complexity increase contradicts Tm. 11.3, p. 353.

11.11 (Section 11.7). Although no genuinely universal learning machines exist, there are machines which can learn limited classes of regular behaviors. For example, let Σ be the class of output sequences s on the alphabet $\{0, 1\}$, generated by autonomous finite-state machines. Discuss a learning machine M, with a learning mechanism in the form of a Turing machine, which can extract the rules of any sequence s in Σ and, given an unlimited learning time, will be able to predict s with certainty.

11.12 (Section 11.8). Define an S_{FM}-surrounding along the lines of Def. 11.9, p. 360, for S_{TM}-surroundings, but with Turing machines replaced by finite-state machines. Give examples of simple S_{FM}-surroundings.

11.13 (Section 11.9). Construct a partially self-describing sentence of the form: "This sentence contains precisely   letters" by filling the empty spaces with an English word designating a number which makes the sentence true.

11.14 (Section 11.9). According to Def. 11.14, p. 387, the recursively enumerable sets might be completely described by their n-indices of the $S_{[n]}$-forms. Hence we could say of a recursively enumerable set $S_{[m]}$, such that $S_{[m]} = \{m\}$, that it is self-describing. Prove that there are such self-describing sets.

Hint: A set $\{x\}$ is recursively enumerable, and hence there is an $n = f(x)$ such that $S_{[f(x)]} = \{x\}$. Use Tm. 11.16A, p. 385, to prove that $f(x)$ is recursive. Next, apply Tm. 11.31A, p. 392, or see Rogers [20] for a similar proof.

11.15 (Section 11.9). The proof of Tm. 11.5, p. 361, is constructive in the sense that it indicates how a Turing machine can be constructed so that it will be self-producing in an S_{TM}-surrounding. Determine in this way a Turing machine which is self-productive in an S_{TM}-surrounding with the code function $c(z) = z$ (cf. Def. 11.9, p. 360).

11.16 (Section 11.10). Use Tm. 11.10, p. 365, for a comparative discussion of the evolution of normal behavior and the evolution of sturucture.

GLOSSARY OF SYMBOLS USED

Symbol	Definition	Page	Name or Explanation
$\Rightarrow$			Implication symbol.
$\neg$	11.22A	397	Negation symbol.
$\vdash_{\mathscr{S}}$	11.20A	396	$\vdash_{\mathscr{S}} W$ means that the wff W is a theorem of the formal theory $\mathscr{S}$.
$\vDash$		347	$\vDash W$ means that the wff W is logically true.
$\overline{}$			$\overline{R}$ is the complement of the set R with respect to the set of all n-tuples, N^n, if R is a set of n-tuples.
$\dot{-}$	11.5A	373	Proper subtraction.
$\ll$	11.16A	393	$A \ll B$ means that A is strongly recursively reducible to B.
$A \to d \xrightarrow[S]{} B$	11.6	358	The object A is productive in the surrounding S.
$\lambda(x)$	11.12A	383	Left pair-decoding function.
$\rho(x)$	11.12A	383	Right pair-decoding function.
$\tau(x, y)$	11.12A	383	Pair-encoding function.
$\mu y[f(x, y) = 0]$	11.4A	372	Minimalization.
$\psi_z(x_1, \ldots, x_n)$	11.1A	370	The partially computable function computed by a Turing machine with code number z.
$\mathscr{B}$	11.3	349	The logical basis of a formal theory.
$C_S(x)$	11.6A	376	The characteristic function of the set S.
$c(z)$	11.9	360	Code function.
$I(W, \mathscr{S})$	11.2	348	Syntactic information of the wff W in the formal theory $\mathscr{S}$.
$\mathrm{In}(\mathscr{S}, \mathscr{B})$	11.3	349	Syntactic information of the formal theory $\mathscr{S}$ relative to its logical basis $\mathscr{B}$.
$\mathrm{Inf}(W)$		348	Semantic information of a wff W.
$\mathscr{L}$			The propositional logic.
p-explanation	11.1	344	Proof sequence explanation.
r-formal	11.18A	396	Recursively formal.
$\mathscr{S}$	11.17A	396	A formal system (theory).
$S_{[n]}$	11.14A	387	The nth recursively enumerable set $\{x : \exists y\, T(n, x, y)\}$.

GLOSSARY (*Continued*)

Symbol	Definition	Page	Name or Explanation
S_{TM}	11.9	360	Surrounding complete with respect to the class of Turing machines.
$S(x)$	11.5A	373	The successor function.
$s(z, u)$	11.4	353	The shortest form function.
$T_{\mathscr{S}}$	11.20A	396	The theoremhood of the formal theory $\mathscr{S}$.
$T(z, x, y)$	11.9A	378	The Turing machine predicate.
U	11.15A	389	A Universal Turing machine.
$U(y)$	11.10A	379	Final-element function.
$U_i^{\,n}(x_1, \ldots, x_n)$	11.5A	373	Projection function.
wff			Well-formed formula.

REFERENCES

1. **Bar-Hillel, Y.**, *Language and Information*. Reading, Mass.: Addison-Wesley, 1964.
2. **Church, A.**, "An Unsolvable Problem of Elementary Number Theory." *American Journal of Mathematics*, Vol. 58, No. 2, pp. 345–363, April 1936.
3. **Davis, M. D.**, "A Note on Universal Turing Machines." In *Automata Studies*, C. Shannon and J. McCarthy (Editors), pp. 167–175. Princeton, N.J.: Princeton University Press, 1956.
4. **Davis, M. D.**, "The Definition of Universal Turing Machine." *Proceedings of the American Mathematical Society*, Vol. 8, No. 6, pp. 1125–1126, December 1957.
5. **Davis, M. D.**, *Computability and Unsolvability*. New York: McGraw-Hill, 1958.
6. **Davis, M. D.**, (Editor), *The Undecidable*. New York: Raven Press, 1965.
7. **Dobzhansky, T.**, *Evolution, Genetics, and Man*. New York: Wiley, 1961.
8. **Henderson, I. F.**, et al., *A Dictionary of Biological Terms*, 8th ed. Edinburgh: Oliver and Boyd, 1963.
9. **Kleene, S.**, *Introduction to Metamathematics*. Princeton, N.J.: Van Nostrand, 1952.
10. **Klir, G. J.**, *An Approach to General Systems Theory*, New York: Van Nostrand Reinhold, 1969.
11. **Lee, C. Y.**, "A Turing Machine Which Prints Its Own Code Script." In *Proceedings of the Symposium on Mathematical Theory of Automata*, J. Fox (Editor), pp. 155–163. Brooklyn: Polytechnic Press, 1963.
12. **Löfgren, L.**, "Recognition of Order and Evolutionary Systems." In *Computer and Information Sciences*, Vol. II, J. Tou (Editor), pp. 165–175. New York: Academic Press, 1967.
13. **Löfgren, L.**, "An Axiomatic Explanation of Complete Self-Reproduction." *Bulletin of Mathematical Biophysics*, Vol. 30, No. 3, pp. 415–425, September 1968.

14. **Löfgren, L.,** "Relative Recursiveness of Randomization and Law-Recognition." *Notices of the American Mathematical Society*, Vol. 16, No. 4, p. 685, June 1969.

15. **Mayr, E.,** *Animal, Species and Evolution.* Cambridge: Harvard University Press, 1963.

16. **Mendelson, E.,** *Introduction to Mathematical Logic.* Princeton, N.J.: Princeton University Press, 1964.

17. **Myhill, J.,** "The Abstract Theory of Self-Reproduction." In *Views on General Systems Theory*, M. Mesarovic (Editor), pp. 106–118. New York: Wiley, 1964.

18. **Popper, K.,** *The Logic of Scientific Discovery.* New York: Basic Books, 1961.

19. **Popper, K.,** "A Realist View of Logic, Physics, and History." In *Physics, Logic, and History*, W. Yourgrau and A. Breck (Editors), pp. 1–30. New York: Plenum Press, 1970.

20. **Rogers, H., Jr.,** *Theory of Recursive Functions and Effective Computability.* New York: McGraw-Hill, 1967.

21. **Turing, A.,** "On Computable Numbers with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society*, Vol. 42, No. 2144, pp. 230–265, November 1936; Vol. 43, No. 2198, pp. 544–546, May 1937.

22. **Von Neumann, J.,** and **O. Morgenstern,** *Theory of Games and Economic Behavior.* Princeton, N.J.: Princeton University Press, 1953.

23. **Von Neumann, J.,** *Theory of Self-Reproducing Automata.* Edited and completed by A. Burks. Urbana: University of Illinois Press, 1966.

24. **Woodger, J. H.,** "Formalization in Biology." *Logique et Analyse*, Vol. 1, No. 1, pp. 97–104, January 1958.

12. Mathematics and Systems Theory

PRESTON C. HAMMER*

Department of Computer Science, The Pennsylvania State University, University Park

12.1. Introduction 409
12.2. Prelude 410
12.3. Extensions of Topological Spaces 411
12.4. Connectedness of Sets 414
12.5. Continuity 416
12.6. Functions 417
12.7. Remarks on Function Theory 419
12.8. Transitivity 420
12.9. Perspectives 421
Appendix. Extended Topology: The Continuity Concept 423
Problems 429
References 432

EDITOR'S COMMENTS

Although Chapter 12 may well be easy reading for a person familiar with the concepts discussed, it may become quite difficult for readers without such a background. To obtain the full benefit from this chapter, the reader should at least be familiar with the fundamentals of set theory, calculus, abstract algebra, and topology. The references recommended in Editor's Comments to Chapter 8–10 seem to represent sufficient preliminaries for this chapter.

Some readers may be interested to read an original paper:

Appert, Antoine, "On the Equivalence of Different Definitions of Continuity in General Topological Spaces, and the Axiomatization of such Spaces" (in French). *Rendiconti del circolo Matematico*, Vol. 10, Series 2, pp. 333–346, 1961.

* This work was partially supported by the NSF Grant GJ-797.

For the definition and a discussion of the Lipschitzian function, see the following book:

Cech, E., *Topological Spaces*. New York: Wiley, 1966.

12.1. INTRODUCTION

My own work in what may be called a systems-theoretic direction has been slanted toward the interpretation of mathematics. The scientist with an appreciation of mathematical methods and the mathematician who is delving into the technical ramifications of his special subarea tend to accept most of the concepts and structures defined by others and to be unconcerned with interpretations. Yet, after some years of work in a rather few areas, I have discovered that rarely is an important concept given proper interpretation; axiom systems have been chosen without careful thought, and, worse, attitudes which are detrimental to the development of mathematics have arisen.

For example, the cavalier treatment of nomenclature indicates a lack of discipline which is far from the idealistic image of scientists portrayed to the public. I have repeatedly shown how certain concepts have been defined in contexts too limited for their characters. Only rarely, however, have I detected any appreciation of the need for clarity.

Now, I have been asked to record some of my conclusions and the reasons for them rather than to choose one of the various areas on which I have worked and to expand on it. My philosophy is simple: It is important to understand every major concept in the simplest of terms possible. Although I accept a concept on which mathematicians have devoted much time as important, I cannot usually accept their definitions or interpretations as being the best which can now be made. Moreover, once I have reached what I feel to be a basically better interpretation or definition, I do not claim to have achieved the ultimate. At one time, I might have felt that some terminal goal had been reached, but experience has indicated that new insights are not only possible but also likely.

Anything of real importance (certainly this includes mathematics) should be made available to the young with as much clarity as possible. Confusion produced by misinterpretations, by poor nomenclature, and by inadequate instruction contributes to the mystery but does not help in resolving the fundamental questions which remain after we have done our best.

System theorists may find that this has little to do with their work. However, since every person who has come through the educational system has been subjected to a number of misinterpretations in regard to mathematics, and since few have had the acumen to question these interpretations, it is clear

that systems theorists, like everyone else, do not have an adequate grasp of the tools they are using. Let me add here that I myself do not have an adequate grasp of mathematics. I differ mainly in the nature of the attempt to improve the situation.

So far I have left out specific clues to the matters that I have tackled. These will come in the body of this chapter. Let me first cover in a rather abrupt fashion topics that I have treated in more detail in a number of papers and manuscripts.

12.2. PRELUDE

Among the terms which seem appropriate in relating mathematics to the objectives of people, I find *accessibility*, *structure*, and *information* to constitute a suitable starting place. The structures of phenomena, that is, the real world or universe, are not directly available to us. Yet man tries to increase the accessibility of these structures by acquiring information concerning them. The means of increasing access are numerous. We have means and modes of travel and of communication; we have books, schools, and teachers to increase accessibility. We have tools, such as telescopes, microscopes, and electronic computers; we have languages, theories and algorithms. Mathematics makes part of its contributions through increasing information via abstract models.

Now, although increasing accessibility is an objective, there are also activities which restrict accessibility. Thus we have buildings, prisons, and locks; we have legal and natural restraints. Guilds, such as special interest organizations, have restricted access to their techniques and guarded their prerogatives, and they continue to do so. Schools not only provide access but also limit access. Technical jargons as well as natural languages provide barriers. Mathematics is no exception despite the fact that pupils are required to study this discipline for a protracted period of years. Perhaps this is the real reason for the failure to introduce general principles in mathematics education, although it seems also true that mathematicians themselves have not really sorted out the general principles.

Thus the structure of mathematics remains a mystery to mathematicians, but unless every teacher of mathematics has some concept of its structure, how can he or she teach it well? I have produced a chart of elemental mathematics which, when suitably modified and explained, could be used to display some of the relationships among mathematical areas. However, nothing comparable has appeared in teacher-preparation courses.

As a general observation, I should say that algebras tend to be simpler than geometries, including analyses and topologies. The reason that such

algebraic constructs as groupoids, semigroups, and groups are comparatively simple is that their definitions can be quite well managed in the language. However, most geometric concepts (e.g., angle) cannot be satisfactorily explained through language.

The reason that arithmetic is taught early in school is not that it is easy but that it is adjudged important. Arithmetic, especially if taught without algebra, is difficult. It should be obvious that every important aspect of mathematics should be taught before graduation from high school. Yet this is not done.

12.3. EXTENSIONS OF TOPOLOGICAL SPACES

Among the branches of mathematics, *topology* is an example of the more active ones of comparatively recent vintage. In keeping with my theme, I ask, What is the quality of information provided by topology? To a marked degree, topology deals with conceptualizations rather than procedures and algorithms. The flavor of topology, accordingly, tends to be qualitative rather than quantitative. In a given context such terms as "closed," "open," "compact," "connected," "perfect," "dense," "separable," and "continuous," as used in topology, indicate ways of describing situations which are assumed, but which are ordinarily beyond the reach of computation.

Let me mention one way in which topologists have provided information. The number of surfaces which we perceive as boundaries of real objects comprise a veritable morass. To attempt to classify the surfaces of all mathematical solids in three-dimensional space or just the surfaces in terms of the locations of their points would, as far as can now be seen, be impossible and fruitless. However, by ignoring most differences and defining the topological equivalence of surfaces, the topologists have provided a classification of such surfaces. This is but one of the successes topologists have achieved.

The commonly accepted definition of topological spaces is an astonishingly permissive one. To generate a topological space, choose a set, say E, and choose *any* class of subsets of E which covers E; you now have a subbase for the open sets of a topology in E. To get all the open sets, include all intersections of finite subclasses of the class of sets chosen; the union sets of arbitrary subclasses of the resulting class constitute a class of open sets which determines a topology.

Now, even in finite sets E there can be a large number of different topologies, and if E is infinite the number is indeed large. Surely, you may then say, this is going too far—who has use for all these topologies? The answer is that the number of explicit topological spaces used is trivial (a small finite number, in a sense) compared to the possibilities. Many mathematicians, including a

number of topologists, would agree. Topological spaces are, they say, *too general*.

The advantages of a general definition are that the conceivable scope of the field is greater and the reasons why certain results hold can be made clearer. The disadvantage is that without specializations the general system may not yield the results needed. However a generalization of a given system always embraces that system and, usually, others as well. A generalization, then, may be effective in providing a basis for comparing concepts, results, and structures in areas previously considered to be separate.

I have given certain generalizations of topological spaces some attention. I shall mention first one generalization of a moderate sort with pleasant consequences. In topology, the intersection of two open sets is an open set, by definition. I drop that requirement. Start with a set E, choose any class of subsets of E, and extend it to include the union sets of every subclass of the class you chose. I will call the resulting class of sets the *open* sets of an *Appert* space (after A. Appert). Note that it is simpler, conceptually, to form an Appert space than a topological space since I do not require that the class you chose cover E, nor that the intersection of each pair of open sets be an open set. However, since these possibilities are not excluded, Appert spaces embrace topological spaces. Now let $x \in E$; then the class $V_0(x)$ of all open sets which contain x as an element is an open *neighborhood base* for x. The complement on E of an open set is a *closed* set by definition, and hence the class of closed sets contains the intersection set of each of its subclasses. The *closure* $f(A)$ of a set A, is the intersection set of all closed sets which contains it. The interior $g(A)$ of a set A is the union of all open subsets of A.

It is now readily verified that f and g are *dual* functions, that is, $g = cfc$ (and $f = cgc$), where c is the complement function. A topologist might object that I have distorted the meanings of "closedness" and "openness." However algebraists also use the term "closed" and my definition embraces all algebraic closures. Again, algebraists might object: What do they care about open sets?

I consider it an obligation, when possible, to meet objections and to establish relevance. If I can show that I have not merely engaged in a generalization for the sake of generalization (as some would say), then I can rest my case. First let me admit that the neighborhood structure is often quite complicated when algebraic closures are involved and, while providing insight, does not seem to lead to nice algorithms. However, consider the case of convexity as a property of sets in some real linear vector space, E. The intersection set of any class of convex sets being convex, I choose to interpret convex as "closed" and the complement of convex as "open." Thus convexity gives examples of Appert spaces which are not topological spaces.

The convex hull operation is a closure operation. It is a closure with respect

to a binary operation, say $f\{x, y\}$, which has the line segment xy as the product, or $\{x\}$ if $x = y$. Since the convex hull of A, say $f(A)$, is a closure with respect to a finitary operation, there exists for each $x \in E$ a neighborhood base of open sets which is minimal, that is, a class of sets each having no proper subset which is a neighborhood of x and no proper subclass of which is a base. It is *necessary and sufficient*, for x to be in the convex hull, $f(B)$, of a subset B of E, that B intersect every one of the neighborhoods of x (in convexity). This result of mine is the first such condition proved. It amounts to reinterpreting the topological result and applying it where it had not been applied before. I have, in fact, established a series of results concerning convexity and generalizations of convexity which were not proved before [16, 18].

Now let me ask; Are Appert spaces too general? They are the least general spaces which include all closures. As applied to convexity, they are quite specialized, having special properties not holding in general Appert spaces. The more general framework was necessary to contain convexity.

I have given an indication that Appert spaces are useful, and it might be thought that they comprise an ultimate in useful generalizations of topology. This is not the case. I shall mention a few other generalizations, using the neighborhood concept. A *Fréchet* (or *reflexive*) *space* is given in E by assigning to each point x in E a class of subsets of E, each of which contains x as an element. This class of sets is called a neighborhood base for x. Then a *convergent* of x is any set which intersects each of the neighborhoods in the base. The class of all neighborhoods of x consists of the ancestral closure of its assigned neighborhood base, that is, of each set which is a superset of some base neighborhood. As a convergent of x is a sample of all neighborhoods of x, so a neighborhood of x is a sample of all convergents of x. Thus there is a basic duality between convergents and neighborhoods. The equivalent set-valued functions f and g replacing closure and interior now are not idempotent in general, but still $cfc = g$. Fréchet spaces are the most general neighborhood spaces which give a satisfactory definition of connectedness of sets by direct extension of the formalism of the topological definition of separation of pairs of sets.

What is a distinctive difference between Appert spaces and Fréchet spaces which are not Appert spaces? In an Appert space the base of open sets may be assigned without reference to points. In Fréchet space each point must be assigned its neighborhoods, since there is generally no base of open sets.

The reflexivity required in all Fréchet spaces (including Appert spaces and topological spaces) is that each $x \in E$ is close to itself, that is, x is an element of all its neighborhoods and thus $\{x\}$ is a convergent of x. In what I call *primitive spaces*, each x has a neighborhood base *no* set of which contains x. These spaces are *irreflexive*. They appear in topology when one deals with the

limit points of sets. In an *isotonic* space, each element is assigned a class of subsets of E as a neighborhood base with no other restrictions [17]. These spaces embrace Fréchet spaces and primitive spaces.

The most general sort of neighborhood spaces that I have considered admit two sets, say E and E_1. To each $x \in E$ there is assigned a class of E_1 which is a neighborhood base for x. A subset of E_1 is a convergent of x, provided it intersects all neighborhoods. Now of what use is such a structure? The answer is that this context is the natural one for approximations and modeling. It is appropriate, for example, to language translation.

All these extensions of topological spaces make use of the neighborhood concept. There are applications (of specific character) of each which are not embraced in a less general space structure. Now, it might be thought, an ultimate has been reached. This is not true. Many concepts associated with topology cannot be best described in any neighborhood space structure, however general, of the kinds I have mentioned. Among these are the concepts *connectedness* of sets and *continuity* of functions.

The gains made in the generalizations I have mentioned involve greater applicability, better understanding, and superior notation. Intuitively, what are neighborhoods? Suppose that you have an objective: for example, getting a Ph.D. degree or a Cadillac. The neighborhoods of the objective are the conditions you must *meet* to achieve it. Neighborhoods *protect* objectives from effortless achievement. When you have met every condition—penetrated every neighborhood—you have *converged*; you have reached the goal. This interpretation, available in examples to children, provides a better insight into neighborhoods than all the topology treatises and textbooks I have seen.

12.4. CONNECTEDNESS OF SETS

The concepts of connections and connectedness in the common language are very broad indeed, being essentially the same as relationships. In the earlier part of this century appreciable effort was spent in finding a definition of connectedness for sets of real numbers. Three men, Riesz, Lennes, and Hausdorff, seemingly in that order, provided the same definition, which was accepted and generalized to higher-dimensional spaces and to abstract spaces. It is reported that G. Cantor earlier suggested a definition which was more general but included, say, the set of rational numbers as a connected set.

Connectedness provides a most interesting example of inertia. Accepting the formal definition, many researchers published studies on its consequences, whereas practically no thought seems to have been given to the sense of the definition in relationship to applications. In 1959, I defined connectedness of sets relative to certain binary relations among sets which I named *Wallace*

separations after A. D. Wallace [28]. This work was published in 1964 [7], following a paper on binary relations among sets [4].

Let R be a binary relation in the power set of E. Then R is a *separation* provided $(X, Y) \in R$ and $X \supset X_1$, $Y \supset Y_1$ implies $(X_1, Y_1) \in R$, that is, I define a separation as hereditary. If $(X, Y) \in R$ implies that X and Y are disjoint, then R is *exclusive or disjunctive*. A symmetric disjunctive separation is a *Wallace separation*.

Let R be a Wallace separation. Then a subset A of E is R-connected, provided A is not the union of a pair of nonempty R-separated sets. This definition of connectedness is a major improvement on the topological definition, which represents a special case. The advantages gained are simplicity, increased applicability, deeper results, and greater insight.

Simplicity is achieved because, in order to specify a Wallace separation, one need merely specify an exclusive binary relation in the power set of a set. From this, by using symmetry and the hereditary property, a unique Wallace separation is determined. Alternatively, from one of the results of my theory, you may specify which sets you wish to have connected. Then the binary relation, comprised of *every pair* of disjoint sets which separate none of the sets you specified, is a Wallace separation—the *maximum Wallace separation* which gives you all the connected sets you specified and as few others as the theory here can allow.

Connectedness-preserving functions are of interest in topology. I have given a *necessary and sufficient* condition for a function to preserve connectedness, a result which applies in topology. In this case, having the proper framework for connectedness of sets led easily to a result which had eluded prior research efforts. Although topologists deal with certain forms of connectedness, the topological definition is not adequate in topology. For example, if anyone had defined connectedness of sets in Euclidean spaces, he might have chosen arc-wise connected sets. Yet, as is shown in [31], arc-wise connectedness is not achievable in any neighborhood space in Euclidean spaces of two or more dimensions. Although the topological definition *preserves the form of definition* for the real line, it *does not preserve the sense of connectedness*. My definition admits the class of arc-wise connected sets in the plane as a complete class of connected sets.

What can be proved about the more general form of connectedness? An amazing number of theorems have been proved. For example, I have given a *necessary and sufficient* condition for the union set of a class of connected sets to be connected. Since the union of any class of connected sets which have a common element is connected, component decompositions of sets are readily established. In fact, my definition provides a much better structure for the discussion of connectedness than the topological definitions since connectedness is not, basically, a topological property. For example, if we

agree that each pair of consecutive integers is a connected set, then the consecutive strings of integers form connected sets. This connectedness is beyond the topological definition, but it is of the kind important in numerical analysis, number theory, logic, and the theory of machines.

Again, it might be thought adequate to let connectedness rest with the Wallace separations. However, in my 1964 paper [7] I pointed out that directed forms of connectedness (e.g., causality) and forms of connectedness resulting from higher-order relations among sets have not as yet been satisfactorily discussed. Where in mathematics education should connectedness be discussed? It is possible to introduce the ideas very early, but at present there is no curriculum which does connectedness even partial justice.

12.5. CONTINUITY

Like connectedness, continuity of functions is normally considered to be in the domain of topology or analysis, and, to a remarkably similar extent, the fundamental notion of continuity is too big for either of these branches of mathematics. The best way I have discovered of considering continuity is to regard it as *dual to invariance*. A function is continuous with respect to whatever properties or relations it preserves.

Recently I was amused to hear an objection to my interpretation of continuity on the grounds that "continuity is continuity, and you shouldn't try to change its meaning." My interpretation is actually compatible with the common language interpretations. Let me assume that I have a function and that I am restricted to topological spaces. Is the function continuous? If it is not a constant function, then the answers are yes *and* no! All constant functions are continuous in topological spaces, as in any reflexive neighborhood space. *Every* function is continuous by proper choice of topologies. Since continuity in topological space contexts does not separate one function from another, it has no *intrinsic* meaning in that regard.

My search for a better understanding of continuity took several years. It took this long because the extensions of neighborhood spaces that I was making could not embrace algebraic homomorphisms as continuous functions. Once I recognized that continuity cannot be well described if restricted to neighborhood spaces, I arrived at the interpretation just given. As for the interpretation itself, I feel that it makes sense and can be presented early in mathematics education, having no need for infinite sets. It is important in systems theories since systems theories try to construct *approximately continuous mappings* from real systems to models and from models to real systems.

There has been an enormous waste of mental energy because of the misinterpretation of "continuity." In thousands of places in the literature it is

proved that a continuous function of a continuous function is continuous. This result is best proved once, early in mathematics. Although having the general idea of continuity does not mean technical excellence in one of its ramifications, it is obvious that mere technical excellence has not produced the general idea! A man may specialize in a field much more effectively if he does so in a context which embraces his specialty. Specialization itself seems to destroy the ability to see broad patterns and relevance.

Let me give an example. The translation of an article from French to English is an approximation to the original. If you say that the translation is good, you have said that the translating was *approximately continuous*—because it saved well enough the important stuff. You say that it is good; another says that it is not. One man's continuity is another man's discontinuity! My article "Extended Topology: The continuity Concept," reprinted from Mathematics Magazine [9], is included as an appendix of this chapter. For a more specialized presentation reference [21] will counter any arguments concerning the mathematical prospects of the new continuity.

Although preserving information or structure is important, it may be expected that many types of functions are not to be so classified. Thus I think that neither Lipschitzian functions nor differentiable real functions can be characterized as types of continuous functions, although I have not proved this. A number of problems arise because of the new interpretations, for example, the comparisons made between algebraic homomorphisms and topological homomorphisms. In topology itself, it seems that limit-point preserving functions have been neglected. These, in Euclidean spaces, are weaker than homeomorphisms but much more restrictive than homomorphisms.

Among other examples of interpretations of concepts normally limited to topology, but not best seen therein, are filters [25], compactness, dimension, perfectness (of sets) [6], and, in fact, the entire area of functional analysis of set-valued set functions, parts of which I have developed in a series of papers.

12.6. FUNCTIONS

Along with the concept of sets, functions and relations can be used to provide an organizational basis for the structure of mathematics in terms of its subareas. Other concepts such as connectedness, continuity, and filters are not confined to a subarea, but as yet are not treated well in formal education. The basic nature of the concept of function is agreed upon by all mathematicians of my acquaintance. Nevertheless, it is treated as a side issue only, there being to my knowledge no textbook or treatise which does the subject partial justice. It is my opinion that every such general concept should be carried along in the educational system, and its applications and ramifications pointed out.

There is a superstitition among mathematicians that functions should not have sets as values; as a result, simplifications which might have been made have not been achieved. Moreover, numbers of interpretations of functions are otherwise neglected. Concerning the term "function" itself, I have been told, in a letter from Professor K. O. May, that it got into mathematics through a misinterpretation of a proper usage of the word by Leibnitz!

For some years it has been clear to me that the suggested synonyms "transformation" and "mapping" should be replaced by "transformer" and "mapper." It was not until I had given a lecture on computer science in March, 1970, that it suddenly occurred to me that a *function is a verb* of a special kind, Thus, if $y = f(x)$, I may diagram the sentence

$$\underline{x \,|\, f \,|\, y} \quad \text{or} \quad \underline{x \,|\, f} \quad ,$$
$$\searrow \underline{y}$$

where x is the subject, f is the verb, and y is the object, direct or indirect.

In speaking I may variously use "x goes into y," "x determines y," "x is mapped into y," "x is labeled y," "x is represented by y," or "x implies y" to indicate the action of the function. Note that all these forms are in the present tense. A computing machine may also be considered as a function. A program is put in, and the computer then produces the output.

A theory of computers, then, should display details as to how the computer does this processing. Until I stumbled onto the verb interpretation I was at a loss as to how to treat the control aspect of the computer. It now becomes clear that here the imperative and conditional imperative modes of verbs supply the clue. A control function times the application of other functions, telling them when to act and on what. The program always forms part of the control. The sequencing of statements, the imperatives LET, DO, and GO TO, and the conditional imperative IF . . . THEN explicitly display the controls which are indeed the essential "thinking" capacity of computers.

I think, then, that by using the concept of functions a theory of computers may be developed in which control functions at various levels are the essential feature. The learning aspect of computers amounts to altering controls during processing.

Let us now return to the verb interpretation. Binary relations are often written xRy, which is already in sentence diagram format. Although a ternary relation may be written as $xyRz$ or $xRyz$, this format would not usually be used. Perhaps $xyzR$, to be read as "(x, y, z) are related," or $yBxz$, "y is between x and z," is, for some uses, better. In any event, it is seen that higher-order relations do not fit the verb format well if an object is required.

Once one sees that the verb interpretation has merit, several new interpretations arise: "x will become y (at a later time)," "x did determine y,"

"x should become y," "x might go into y," "x ought to go into y," "x shall have gone into y," "probably x will go into y," and so on. Thus there is a range of possibilities which have real-life counterparts and some of which have mathematical models. Inverse functions correspond to inverted sentence structure. Thus "y is determined by x_1 and x_2" diagrams to

$$\underline{y \,|\, f^{-1} \,|\, x_1, x_2}\,.$$

Here, I may point out, the inverse function is normally set-valued and, because of present conventions, is not often a function. However, it is a function in the strict sense since the formal definitions do not preclude sets as values.

Why has this interpretation been missed? I find this question especially intriguing since, unlike many of my other interpretations, it has been welcomed by mathematicians who, after all, "know" what a verb is and what a function is. It is alarming to realize how effectively training and specialization inhibit thinking. The separation of mathematics from other aspects of our activities has been too successful. It hampers education and inhibits understanding.

12.7. REMARKS ON FUNCTION THEORY

From an early school age pupils learn about complicated numerical-valued functions. In fact, they start with functions of two or more variables in arithmetic; they learn basically complicated formulas concerning areas and angles, which are peculiar functions of great importance. The trigonometric functions, logarithms, and exponential functions are of a still higher level of difficulty. But when will pupils learn about functions as such? The answer seems to be nowhere except in complicated situations.

Can a theory of functions be presented which does not involve the difficult problems of computation? The answer is yes! Functions mapping finite sets into finite sets can be grasped, and research problems yet unsolved can be posed in this context. For example, let E be a finite set, with n elements, and let F be the family of all functions mapping E into itself. Then each function f in F has associated with it an integer which is the order of the smallest composition semigroup which contains f (already, we have a *new function* associating an integer with each f). One problem is to determine for each n the maximum such integer.

For $n = 10$, the answer is 30. That is, any function f which maps a set of 10 elements into itself generates a semigroup $\{f, f^2, \ldots, f^k\}$, where $k \leq 30$. Any permutation f_0, which is, let us say, a cycle on parts of a partition of E

into 2, 3, 5 element sets, achieves the maximum, 30. The permutations form a subgroup of F.

It might be assumed that a permutation will always provide a maximal-length subsemigroup of F. This does not seem, however, to be the case. For $n = 21$, my calculations show there is a function which generates a semigroup of order 421, whereas the maximum for permutations is 420. Moreover, on the shaky assumption that my calculations are correct, 21 is the smallest integer for which permutations do not provide a maximum!

Functions on finite sets can serve to illustrate the properties of functions better than the intricate ones presently used in the schools. Proceeding to advanced considerations, let me briefly mention aspects of a theory of periodicity I have generated.

Let F be the family of functions mapping a set E into itself. Let T be the family of functions mapping E into a set E_1; I will assume that E_1 has cardinal equal to that of E to achieve maximum scope. Now a function f in F is a *period function* of t in T provided $tf(x) \equiv t(x)$ for $x \in E$. Symbolically $tf = t$, or f is a *right identity* of t, if f is a period function of t.

Let $u(f, x) = \{x, f(x), \ldots, f^n(x), \ldots\}$ be the *orbit* of x under f. Let $\bar{u}(f, x)$ be comprised of all elements y of E, such that $u(f, y)$ intersects $u(f, x)$. Then $\bar{u}(f, x)$ is the *orbital tree* of x under f. Note that $y \in \bar{u}(f, x)$ implies $\bar{u}(f, x) = \bar{u}(f, y)$. Hence the sets $\bar{u}(f, x)$ decompose E into minimal sets invariant under both f and its inverse, f^{-1}.

Theorem 12.1. A necessary and sufficient condition that f be a period function of t is that $t^{-1}t(x) \supseteq \bar{u}(f, x)$. Specifically, t must be constant on $\bar{u}(f, x)$ for each x. Letting e be the identity function in F, the only period function of t is e, if and only if t is biunique. A constant function t has all functions in F as period functions. The set F_t of all period functions of t comprises a subsemigroup of F.

This theorem, which I do not prove here, is basic to periodicity considera-ations, and many more results can be established without introducing extraneous properties of functions. Since I have seen a treatment of this kind nowhere, and since it is within reach of many, the question arises as why it is not presented. The answer is that functions are considered only for their applications; attempts to grasp their properties have been comparatively negligible.

12.8. TRANSITIVITY

Relations, in particular binary relations, are half-heartedly recognized as important in mathematics. In regard to order relations, there have been some grievous misinterpretations and confusing terminology. For example, one

textbook on algebra contains a statement to the effect that lattices are too general to be considered as mathematical objects. Another book dealing with topology warns the reader not to consider order relations apart from their applications. This is arrant nonsense. Binary relations that are not transitive occur everywhere, such as in the graphs of functions. Transitive relations that are not antisymmetric are as common as comparative adjectives! Strict order relations by the mumbo jumbo of mathematics, are not order relations! This is not true—any transitive relation may be called an order relation.

I now illustrate the uses of transitive relations in defining *measure* and *approximation spaces*. Let E be a set, and let f map E into a value space V. Let R be an order relation in V. Then, if $(fx, fy) \in R$, I say " x has f-measure less than or equal to that of y." This is the simplest and best definition of measure I have been able to generate. It covers most examples of measure in mathematics which the so-called measure theory does not.

Again let E be a set, and to each $x \in E$ let $v(x)$ be a set of elements which are admissible approximants of x. Let $V = \cup \{v(x): x \in E\}$. Then an *approximation list* or *dictionary* is the binary relation in $E \times V$ defined by

$$L = \{(x, y): x \in E, \quad y \in v(x)\}.$$

To convert this list into an approximation space, let $T(x)$ be a reflexive order (i.e., transitive) relation in $v(x)$ for each $x \in E$. The corresponding approximation space is the ternary relation in $E \times V \times V$:

$$S = \{x, y_1, y_2): x \in E, (y_1, y_2) \in T(x)\}.$$

Here $(x, y_1, y_2) \in S$ is interpreted as " y_1 is at least as good an approximant of x as y_2 is." I have generated a rudimentary theory of approximation based on this very simple structure, which, incidentally, embraces all topologies.

Approximation and measure are truly big concepts, important to every systems theorist, for example. There is every reason not to accept as suitable the trivialization of measure in measure theory or the confinement of approximation to normed linear spaces.

12.9. PERSPECTIVES

Over the years I have found myself forced to alter my views repeatedly as new evidence comes to my attention concerning mathematics. For example, I was once taught that mathematicians, although not free from errors in making proofs, tended to exhibit excellent taste in the selection of terms. This I find to be far from the case. True, there are many instances of appropriate choices of terms, for example, those derived from Greek and not

otherwise used. However, the nomenclature of binary operations and of binary relations, particularly order relations, shows evidence of utter carelessness and lack of language sense. "For all" is used when "for each" is superior. "There exists" is the rather esoteric equivalent of "there is." The term "measure" has been so defined in measure theory that it is impossible to use it properly therein. Anyway, the concept of "measure" is too big for measure theory!

The adoption of certain terms, possibly on some valid basis, has often left out the reason, thus depriving the younger generations of an insight. Such is the case with "continuity" and "filter" in analysis and topology. Of course, such choices as "ring" and "field" were poorly motivated.

Disregard of language and irrelevance are shown in the repetition of the statement "there is a *unique* empty set." In my estimate, it is important that nomenclature does not violate good taste, and it should, when possible, enable learning to take place rather than consume time with idiosyncrasies. The structure of mathematics as a system and as a sublanguage is what is missing. A nomenclature system cannot be well devised in the absence of a structure displaying the interrelationships of the parts.

Several features of mathematical activity should be raised to the conscious level. The uses of generalization and skill in it can be taught. The search for atoms is expressed in mathematics by the search for bases, for generators of various kinds. Here we always look for the smallest set which, in the presence of certain operators, will generate the entire set. If no minimum can be achieved, then other reduction criteria can be used. Every system should be discussed in terms of how to generate all possibilities. How, for example, can we generate all continuous functions, all differentiable functions, all norms, or all topologies? Wherever possible, the general relevance of concepts and results should be discussed. Techniques of proofs and of discovery should be presented. Moreover, the nonapplicability of mathematics should be made explicitly clear. There is no more deadly aspect of mathematics education than the continued emphasis on its values and its successes while its shortcomings are ignored.

In mathematics education, there is an inversion. General principles and concepts are brushed over, if presented at all, while myriads of technical details which form no pattern are learned. Yet it is far easier to grasp the general concept of continuity, as I have presented it, than to be proficient in any one of the special types. Furthermore, grasp of a pattern should enable better specialization.

Projective geometry, now relegated to undergraduate or graduate programs, where it is considered passé, is a good area for the schools since it sheds light on visual perception. Euclidean geometry is the geometry of the tactile sense, which is accorded greater reality than the visual in view of optical illusions.

A stick held in water appears to be bent; our fingers tell us it is straight. We believe out tactile sense.

Logicians have failed to grasp that there are many interpretations of implication systems. Thus a function is an implication system—assume x, $f(x)$ is implied. A differential equation is a continuous logic or implication system. How can we get any idea of what mathematics is about if we miss obvious interpretations?

The emphasis on axiomatics in recent years has been not only helpful, but harmful as well. In many cases it would be far better to formulate simple definitions, rather than to freeze a concept in an unwise choice of axioms. Examples in point are norms, measures, metrics, and topological spaces. There are important distances, for example, which not only do not satisfy *any* of the Fréchet axioms but are not even real-valued.

It seems clear, to me at least, that system theorists should not assume that current mathematics is adequate. Systems theories cannot be general except in relationship to an earlier, more specialized theory. Why? Consider, for example, the almost total absence of higher-order operators and relations in mathematics. Some of these will one day be recognized as important, but no one can at present forsee which ones. If we claim to embrace them all, we will be dealing with unoperational generalities. If we exclude them all, sooner or later some will come to be recognized as very important. There is no reason to believe that any mathematical structure will enable us to deal with all future contingencies. Therefore the goal of generality should never be claimed as accomplished.

*APPENDIX. EXTENDED TOPOLOGY: THE CONTINUITY CONCEPT**

INTRODUCTION

The concept of continuity as defined in topology and analysis has been extremely useful in discussing invariance of certain set-properties. However, the forms of the definitions are such as to make it seem that continuity is fruitfully confined to transformations or functions on spaces with an infinity of points. We have been developing a system of extended topology which includes concepts relevant to both finite and infinite spaces.

Applicability to finite spaces is necessary for a system of this kind, to be basic since we want it to apply to numerical analysis, statisics, computing, and logic, as well as to language theory in general. Nearly ten years ago we had decided that any reasonable definition of continuity would necessarily display the homomorphisms of algebras as particular forms of continuity. However, we have only recently found a simple and intuitive way of defining continuity which includes topological and algebraic homomorphisms in the same framework.

Our feeling about the essential feature of continuity is that continuous mappings are those which preserve information or orderliness of some kind. However, this notion is too broad to implement now in some form of theory if we keep certain theorems from topology. Hence, in this paper we give a very general definition by standards of the current literature but do not attempt to carry out even a rudimentary classification of these kinds of continuity.

It is noticable that, while a function or transformation defined to be continuous maps elements into elements, the invariant properties almost always refer to sets in the topological treatments. Thus connectedness, compactness, closedness, or openness are all set properties. It is not necessary or useful always in dealing with invariance of set properties to consider set-valued transformations generated by element-valued ones. However, we will, in order not to deviate too far from custom, consider only such induced set-valued transformations here.

A Simple Extension of Topological Continuity

Let M be the space with null set N and let $\mathcal{M}$ be the class of all subsets of M. Then a function u which associates with each subset of M a subset of M is called *expansive* provided $uX \supseteq X$ and $u(X_1 \cup X_2) \supseteq uX_1 \cup uX_2$ always. These expansive functions form a generalization of the closure functions of topology since an expansive function u in order to be a Kuratowski closure function must also be *idempotent* $(u(uX) \equiv uX)$, *additive* $u(X_1 \cup X_2) \equiv uX_1 \cup uX_2$ and satisfy $uN = N$. We have shown that much of the basic structure of topology is better considered for expansive functions than for the Kuratowski closure functions.

Now suppose M_1 is another space and v is an expansive function in M_1. Then if $t: M \to M_1$ is a function mapping M into M_1 this mapping induces a mapping from $\mathcal{M}$ into $\mathcal{M}_1$, the class of all subsets of M_1, by defining $tX = \{tp : p \in X\}$. We follow custom and use the same symbol t for the set-to-set transformation as for the original element-to-element one. Now if u is an expansive function in M, then t is said to be (u, v)-*continuous* provided $t(uX) \subseteq v(tX)$ for all X in $\mathcal{M}$ or, more shortly, provided $tu \subseteq vt$. In ordinary terms, if u and v are closure functions, then t is (u, v)-continuous provided

the transform of the closure of X is contained in the closure of the transform of X. The reader may verify that this definition coincides with that of topology and hence of real or complex functions.

An ordered pair (X_1, X_2) of subsets of M is said to be *separated* with respect to an expansive function u in M provided $uX_1 \cap X_2 = N$; i.e., provided uX_1 and X_2 have no elements in common. This may be better read " X_2 is separated from X_1."

Theorem 12.1A. A necessary and sufficient condition that $t: M \to M_1$ be (u, v)-continuous, where u and v are expansive functions, is that if $X_1, X_2 \in \mathcal{M}$ and (X_1, X_2) is not separated with respect to u then (tX_1, tX_2) is not separated with respect to v.

Remark. This theorem is proved in [3]. However, its proof follows from the definitions given and the reader may have the pleasure of verifying the conclusion himself. This is a *basic* theorem on several counts. It gives an image of a continuous mapping as one which does *not separate* ordered pairs of sets not separated in its domain. The theorem would *not* hold if we had blindly followed custom and required (X_1, X_2) to be considered separated provided $uX_1 \cap X_2 = N = X_1 \cap uX_2$ and it also would not then hold for general topology. Moreover, since separations of ordered set pairs are commonly in use which do not depend *in this fashion* on expansive or closure functions, we now have a springboard for generalization.

What is a Separation

Let us consider pairs (X_1, X_2) of sets from M. What sort of consideration might lead us to say X_2 is separated from X_1? *Certainly* we will not require that X_2 separated from X_1 will imply X_1 is separated from X_2, since many applications may be expected to be asymmetric. Shall we require $X_1 \cap X_2 = N$? No! For, X_1 and X_2 might have a common part considered trivial from some standpoint. By this reasoning and examples we were led to conclude that the only property we would use was that if (X_1, X_2) are separated and $X_1 \supseteq X_3$, $X_2 \supseteq X_4$ then (X_3, X_4) are separated. Thus we have a tentative picture of a separation as a *hereditary* binary relation in $\mathcal{M}$; i.e., a separation is a set of ordered pairs of sets satisfying the above hereditary property.

But, wait! Could not there be something desirable added if we did not consider simply *pairs* of sets but also triples, quadruples, or sequences of sets as separated? It became clear that this also was reasonable and that for certain applications pairs of sets would not do. Hence we now may think of a separation as a hereditary subset of a sort of vector space in which each vector has sets as components. To be explicit let us assume S is a set of ordered ennuples, $\{(X_1, X_2, \ldots, X_n)\}$ of subsets of M which is hereditary. Then the set of

ennuples which are not separated form an *ancestral* subset of the space of all ordered ennuples and this we call as *associated* or the *association dual to S*.

Let A be an ennary association for M, and B an ennary association for M_1. Then t is (A, B)-*continuous* provided $tA \subseteq B$; i.e., $(X_1, \ldots, X_n) \in A$ implies $(tX_1, \ldots, tX_n) \in B$. *A continuous mapping is association preserving!*

To see that this definition is not foolish we now indicate its application to the characterization of homomorphisms. Let t map M into M_1 and suppose there is a binary operation $\cdot$ defined in M and a binary operation $\circ$ defined in M_1. Then t is a *homomorphism* (multiplication-preserving) provided $t(p_1 \cdot p_2) = (tp_1) \circ (tp_2)$ for all $p_1, p_2 \in M$.

We let A be the minimal ternary association in M including all triples of the form $(p_1, p_2, p_1 \cdot p_2)$, where we use $p_1, p_2, p_1 \cdot p_2$ as one-point sets. Note that the triples mentioned form the *graph* of the binary operation. Now for purposes of our interpretation $(X_1, X_2, X_3) \in A$ provided there exists $p_1, p_2, p_3 \in M$ such that $p_i \in X_i$, $i = 1, 2, 3$ and $p_3 = p_1 \cdot p_2$. Similarly let B be the minimal ternary association in M_1 such that B contains all triples of the form $(q_1, q_2, q_1 \circ q_2)$ for all $q_1. q_2 \in M_1$.

Theorem 12.2A. With A and B as defined above, $t: M \to M_1$ is a homomorphism if and only if t is (A, B)-continuous.

Proof. First suppose t is a homomorphism and $(X_1, X_2, X_3) \in A$; then there is $(p_1, p_2, p_3) \in A$, where $p_3 = p_1 \cdot p_2$ and $p_i \in X_i$, $i = 1, 2, 3$. Then $(tp_1, tp_2, tp_3) = (tp_1, tp_2, tp_1 \circ tp_2) \in B$ and hence $tA \subseteq B$ if t is a homomorphism.

Next, suppose t is (A, B)-continuous. Then $(p_1, p_2, p_1 \cdot p_2) \in A$ for each $p_1, p_2 \in M$ and hence $(tp_1, tp_2, t(p_1 \cdot p_2)) \in B$. But tp_1, tp_2, and $t(p_1 \cdot p_2)$ are elements in M_1 and hence necessarily $t(p_1 \cdot p_2) = tp_1 \circ tp_2$ by definition of B. Hence t is a homomorphism. Q.E.D.

The point of this theorem is not so much that the result is obtained as it is that we have now demonstrated a common framework including all the form preserving maps called homomorphisms in mathematics. Why should the simple binary relation-preserving map require triples, whereas topology with its infinities requires ordered pairs? It is because we assume that the binary operation is not necessarily *commutative* that we must use triples. If the binary operations $\cdot$ and $\circ$ had been commutative, then we could have defined A as the binary association generated by all pairs $(\{p_1, p_2\}, p_1 \cdot p_2)$ and B correspondingly. However, it is more convenient in general to use the triples. In order to preserve several functions in one mapping we may put all their graphs in ordered ennuples of sets. Thus, if $\cdot$ and $+$ are two binary operations we may generate an association by using quadruples $(p_1, p_2, p_1 + p_2, p_1 \cdot p_2)$ as a base: we may include ordinary topological continuity by considering, say, $(p_1, p_2, p_1 + p_2, p_1 \cdot p_2, X, p)$ as a base for an association where $p \in uX$ for u a topological closure function.

Suppose M is the set of real numbers and $p_1, p_2, \ldots, p_n, \ldots$ is a convergent sequence. Then we may use $(p_0, p_1, \ldots, p_n, \ldots)$ as a base for an association where p_0 is either one of the p_i for $i \geqq 1$ or p_0 is a limit point of $\{p_n\}$. However, this is equivalent to basing a binary association on (p_0, S), where S is the set of numbers in a convergent sequence and $p_0 \in uS$ where u is the closure function. Hence, the topological definitions of continuity require only ordered *pairs* of sets. Here $\{(p_0, S)\}$ may be considered as the *graph* of the real-number topology.

Association-preserving maps, however, cover a much wider range of applications than is indicated by using graphs or multiple graphs. We may generate associations from any relations among sets involving a fixed number of sets (finite or infinite). Thus for the real numbers we might define $(X_1, X_2) \in A$ provided $uX_1 \cap uX_2 \neq N$ where u is the closure function. Mappings of the real numbers into themselves which are (A, A)-continuous are of interest to study. One property they have is that every set X dense in a connected subset of the real numbers maps into a set dense in a connected subset.

Properties of Continuous Transformations

This new form of the definition of continuity is so recent that we have not begun the necessary classifications of separations or their dual associations. However, one important property of continuous mappings is mentioned as stated in the following theorem.

Theorem 12.3A. Let A, B, C, respectively, be ennary associations for spaces M, M_1, and M_2. Then if $t_1: M \to M_1$ is (A, B)-continuous and $t_2: M_1 \to M_2$ is (B, C)-continuous, the composite transformation $t_2 t_1: M \to M_2$ is (A, C)-continuous. In particular, if $t: M \to M$ is an (A, A)-continuous mapping, then $t^k: M \to M$ is (A, A)-continuous for each integer $k \geqq 2$.

Proof. We have $t_1 A \subseteqq B$, $t_2 B \subseteqq C$ whence $t_2(t_1 A) \subseteqq t_2 B \subseteqq C$ and $t_2 t_1$ is (A, C)-continuous. Note that t_2 is inclusion-preserving and hence $t_1 A \subseteqq B$ implies $t_2 t_1 \subseteqq t_2 B$. Q.E.D.

It will be noted that the proof consumes very little space although the theorem applies for all the usual homomorphisms, algebraic or topological. Among the problems of classification which we mentioned are those related to descriptions of minimal bases for associations. When, for example, can continuity *at a point* be defined so that a transformation continuous at each point is continuous? We have shown that certain binary separations are characterized by means of the Wallace functions which, in a appropriate circumstances, become topological closure functions. There are now the problems of extending this work to functions of several set-variables. Concepts of connectedness of sets depending on triples of sets and so on are now also

seen to be useful to consider. We have completely characterized connectedness-preserving transformations as a form of association-preserving continuity for topological spaces.

CONCLUSION

The definition of continuity we have given here has the satisfactory feeling of being more intuitive than the topological definition—a continuous mapping separates no associated sets. That the definition applies to finite spaces is quite clear and very important. Thus we may say that two people are not separated provided they have at least one grandparent in common. Then it is feasible and sensible to speak of continuity in people-to-people transformations.

The insight into the notion of continuity arising from our definitions should not be considered ultimate. We are developing the concepts which go with uniform continuity, Lipschitz conditions, etc., which have a different flavor. Moreover, as we have suggested, there seems no necessity for restricting the transformations involved to those induced by element-to-element mappings and this restriction will have to be dropped to characterize closedness-preserving transformations. An ordinary continuous transformation has the property that the inverse image of closed sets is closed. Now the inverse of a transformation is normally *set-valued* and it is separation-preserving when the transformation is association-preserving. Perhaps we shall want to consider separation-preserving maps under some other heading than continuity, but it is clear that they are dual to association-preserving ones.

The papers most relevant to an understanding of this aspect of extended topology at this time are [1], [2], [3]. These contain references to other works. Reports containing these and the yet-unpublished [3] are available at the Numerical Analysis Department, University of Wisconsin.

Finally it now seems possible to extend the Erlanger Program of Felix Klein to many more systems than the geometries he discussed. Invariance and continuity are essentially dual concepts, and we have taken a step here in the direction of demonstrating it.

REFERENCES

1. **Hammer, P. C.,** Extended Topology: Set-Valued Set-Functions. *Nieuw Archief voor Wiskunde*, Vol. 10, 1962.
2. **Hammer, P. C.,** Extended Topology: The Wallace Functions of a Separation. *Nieuw Archief voor Wiskunde*, Vol. 9, 1961.
3. **Hammer, P. C.,** Extended Topology: Continuity I, to appear.

PROBLEMS

Let F be the family of all functions f mapping the power set 2^E of a set E into itself. F inherits from the set algebra and inclusion relation in 2^E a corresponding algebra and order relation. Thus, $f \cup g$ and $f \subseteqq g$ mean, by definition, the function $(f \cup g)(A) \equiv f(A) \cup f(A)$ and the inclusion $f(A) \subseteqq g(A)$ for all subsets A of E. Thus F is a composition semigroup with an identity element e and with the complement function c.

A function f is called a *closure function*, provided that it is *enlarging* [$f(A) \supseteqq A$ always], *inclusion-preserving* or *isotonic* [$A \subseteqq B$ implies $f(A) \subseteqq f(B)$], and *idempotent* [$f(f(A)) \equiv f(A)$]. A function g is an *interior function* provided it is *shrinking* [$g(A) \subseteqq A$ always], isotonic, and idempotent.   (Hammer [5]).

12.1. Prove that the maximum order of a composition semigroup generated by a closure function f and an interior function g is 6.   (Hammer [4])

12.2. The *dual* of a function f in F is cfc. If f is a closure function, $g = cfc$ is an interior function. Prove that the smallest semigroup under composition containing c and a closure function f has at most 14 distinct elements.

12.3. Generalize the concepts of closure and interior to functions mapping a partially ordered set E into itself. Show that the theorem in Problem 12.1 holds in this case. What would be analogous to the complement function here? Apply to real-valued functions defined on the real line.   (Hammer [4])

12.4. A function $u \in F$ is a *primitive* function, provided that

$$u(A) \equiv \cup \{(f \cap c)A : B \subseteqq A\}.$$

If v is cuc, where u is a primitive function, prove that

$$v(A) \equiv \cap \{(f \cup c)B : B \supseteqq A\}.$$

Prove that the dual of a primitive function is a primitive function.

12.5. Let (f, g) be a dual pair of isotonic functions in F. A set A is called a *convergent* of x, provided $x \in f(A)$. A set B is called a *neighborhood* of x, provided $x \in g(B)$. Note that the class of all convergents of a point x is ancestrally closed, and so is the class of all neighborhoods of x. Prove that A is a convergent of x if and only if $A \cap Y \neq \varnothing$ for every neighborhood Y of x, and that B is a neighborhood of x if and only if $X \cap B \neq \varnothing$ for every convergent X of x.   (Hammer and Gastl [17])

12.6. A *base* class of an ancestrally closed class $\mathscr{C}$ of subsets of E is a subclass of $\mathscr{C}$ which has $\mathscr{C}$ as its ancestral closure.

To each $x \in E$ let $V_0(x)$ be a class of subsets of E, *no one* of which contains x. Let $V_0(x)$ be a neighborhood base for x. Let $\mathscr{C}(x)$ be the class of all convergents of x, that is, $A \in \mathscr{C}(x)$ if and only if $A \cap Y \neq \varnothing$ for each $Y \in V_0(x)$.

Let $V(x)$ be the class of all neighborhoods of X, that is, $V(x)$ is the ancestral closure of $V_0(x)$. Let $f, g \in F$ be defined by $x \in f(A)$, provided A is a convergent of x and $x \in g(B)$, and provided B is a neighborhood of x.

Prove that f and g form a dual pair of primitive functions. *Hint*: Prove that there exists a convergent base $\mathscr{C}_0(x)$ for each x, no set of which contains x. Then prove that an isotonic function u is primitive if and only if $x \in u(A)$ implies $x \in u(A - \{x\})$.

12.7. Hammer first defined primitive functions to provide the "natural" first order, or primary limit points, relative to an arbitrary function u in F. Let $u \in F$. Prove that the function u_0, defined by

$$u_0(A) \equiv \cup (u \cap c)B : B \subseteq A),$$

is a primitive function.

12.8. Let F_1 be the subfamily of F composed of all closure functions in F. Prove that the intersection function of each subfamily of F_1 is a function in F_1. Hence prove that, if u is any function in F, there exists a unique minimal closure function containing u. Dually, using interior functions, prove that there exists a unique maximal interior function contained in u.

12.9. A function $f \in F$ is *domain finite* provided

$$f(A) \equiv \cup \{f(B): B \subseteq A, B \text{ a finite set}\}.$$

Note that every domain finite function is isotonic. Prove that a function $f \in F$ is domain finite if and only if $f \cup_\alpha (A_\alpha) = \cup_\alpha f(A_\alpha)$ for every class of sets $\{A_\alpha\}$ linearly ordered by inclusion. Ascending chain conditions are naturally assumed!

12.10. The domain finite function f generated by an arbitrary function u for F is defined by

$$f(A) \equiv \cup \{u(B): B \subseteq A, B \text{ a finite set}\}.$$

If u is a closure function, prove that f is a closure function.

12.11. If f is a domain finite function and $g = cfc$, prove that there exists for each $x \in E$ a unique minimal convergent base and a unique minimal neighborhood base. Note that the dual g of a domain finite function t is not domain finite in general but satisfies the identity

$$g(A) \equiv \cap \{g(B): B \supseteq A, cB \text{ is finite}\}.$$

All closures under finitary operations of algebras are domain finite functions.

12.12. A union base class of sets for an isotonic function $f \in F$ is a class $\mathscr{A}$ of subsets of E, such that

$$f(A) \equiv \cup \{f(B): B \subseteq A, B \in \mathscr{A}\}.$$

Prove that each domain finite function f has a unique minimal base class, composed of finite sets.

12.13. Define two operators, Σ and Δ, mapping F into itself as follows:

$$(\Sigma f)A = \cup \{f(B): B \subseteq A\};$$
$$(\Delta f)A = fA \cap c \cup \{fB: B \subset A\} \text{ (note } proper \text{ inclusion).}$$

Prove that Σf is the minimum isotonic superfunction of f and that Σ is a universally additive closure operator, that is, $\Sigma f \supseteq f$ always, $\Sigma(\cup f_i) \equiv \cup (f_i)$, $\Sigma(\Sigma f) = \Sigma(f)$.

Prove that a necessary and sufficient condition that $\Sigma\Delta f = f$ (i.e., that Σ is inverse to Δ) is that $x \in fA$ implies the existence of a minimal subset B of A such that $x \in fB$. In particular, for every domain finite function f, $\Sigma\Delta f = f$.

(Hammer [15])

12.14. Let E be a real linear vector space, and let $f(A)$ be the minimum convex set containing A. Then f is a closure function as described above. Show that f is domain finite. What is the order of the semigroup under composition generated by complement c and f? (One of my students reported he could prove that 10 was the number for any space of finite dimensions $n \geq 1$. Note that 14, the maximum possible number, may not be achieved, but that the number is necessarily even.) A number of similar problems can be generated: closure under multiplication in the positive integers, closure under the greatest common divisor or lowest common multiplier operations in the positive integers, closure under subtraction in the set of all integers, and so on.

12.15. Let e be the identity function, let f and g, respectively, be closure and interior functions, let r be a contractive (isotonic, shrinking) function, and let s be an expansive function (isotonic, enlarging) in F. Prove that $e \cup rf$ is a closure function, and that $e \cap sg$ is an interior function. Consequently, $e \cup gf$ is a closure function, and $e \cap fg$ is an interior function.

(Hammer [4]).

12.16. Let f be the Kuratowski closure function for a metric space, and let $g = cfc$. A pair (A, B) of sets is *separated* provided $f(A) \cap B = \varnothing = A \cap f(B)$. Prove that the maximum function u mapping 2^E into itself such that

$$\{(A, B): fA \cap B = \varnothing = A \cap f(B)\} = \{(A, B): fA \cap u(B) = \varnothing = u(A) \cap fB\}$$

is $u = e \cup gf$. (Note that u is shown in Problem 12.15 to be a closure function, but *not* a Kuratowski closure function.)

(Hammer [7])

12.17. A Wallace separation in E is a symmetric binary relation R which is disjunctive [$A \cap B = \varnothing$, if $(A, B) \in R$] and hereditary [$[A, B) \in R$ and $A \supseteq A_0$, $B \supseteq B_0$ implies $(A_0, B_0) \in R$]. For a given Wallace separation R, a set C is R-connected provided that $C \subseteq A \cup B$ and $(A, B) \in R$ implies $C \subseteq A$ or $C \subseteq B$. Show that there exists a unique maximal Wallace separation R^*

which has the same class of connected sets as R. Hence give a *necessary* and sufficient condition for a function on E to another set to preserve connectedness.   (Hammer [7])

12.18. Does there exist a finest topology on the real line (i.e., with a maximal collection of open sets) which has the usual connected sets as connected? (Unsolved.)   (Hammer and Singletary [27])

REFERENCES

1. Hammer, P. C., "Maximal Convex Sets." *Duke Mathematical Journal*, Vol. 22, No. 1, pp. 103–106, March 1955.

2. Hammer, P. C., "General Topology Symmetry, and Convexity." *Transactions of the Wisconsin Academy of Sciences, Arts, and Letters*, Vol. 44, pp. 221–255, 1955.

3. Hammer, P. C., "Kuratowski's Closure Theorem." *Nieuw Archief voor Wiskunde*, Vol. 8, Ser. 3; pp. 74–80, 1960.

4. Hammer, P. C., "Extended Topology: the Wallace Functions of a Separation." *Nieuw Archief voor Wiskunde*, Vol. 9, Ser. 3, pp. 74–86, 1961.

5. Hammer, P. C., "Extended Topology: Set-Valued Set-Functions." *Nieuw Archief voor Wiskunde*, Vol. 10, Ser. 3, pp. 55–77, 1962.

6. Hammer, P. C., "Extended Topology: Perfect Sets." *Portugaliae Mathematika*, Vol. 23, Fasc. 1, pp. 27–34, 1964.

7. Hammer, P. C., "Extended Topology: Connected Sets and Wallace Separations." *Portugaliae Mathematika*, Vol. 22, Fasc. 4, 1963.

8. Hammer, P. C., "Extended Topology: Structure of Isotonic Functions." *Journal fur die Reine, und Angewandte Mathematisch*, Vol. 213, No. 3, pp. 184–186, 1964.

9. Hammer, P. C., "Extended Topology: the Continuity Concept." *The Mathematics Magazine*, Vol. 36, No. 2, pp. 101–105, March 1963.

10. Hammer, P. C., "Extended Topology: Domain Finiteness." *Proceedings of the Amsterdam Academy*, Vol. LXVI, Ser. A, No. 2, pp. 200–212, 1963.

11. Hammer, P. C., "Extended Topology: Correction." *Proceedings of the Amsterdam Academy*, Vol. LXVI, Ser. A, No. 2, 1963.

12. Hammer, P. C., "Extended Topology: Appendix." *Proceedings of the Amsterdam Academy*, Vol. LXVI, Ser. A, No. 3, pp. 384–386, 1963.

13. Hammer, P. C., "Extended Topology: Additive and Subadditive Subfunctions of a Function." *Rendiconti del Circolo Matematico, di Palermo*, Vol. XI, No. 3, pp. 262–270, 1962.

14. Hammer, P. C., "Semispaces and the Topology of Convexity." In *Proceedings of Symposia in Pure Mathematics Convexity*, American Mathematics Society, Vol. VII, pp. 305–316, 1963.

15. Hammer, P. C., "Topologies of Approximation." *Journal of the Society for Industrial and Applied Mathematics*, Ser. B, Vol. 1, pp. 69–75, 1964.

16. Hammer, P. C., "Extended Topology: Caratheodory's Theorem on Convex Hulls." *Rendiconti del Circolo Matematico di Palermo*, Vol. XIV, Ser. II, 1965.

17. **Hammer, P. C.,** and **G. Gastl,** "Extended Topology: Neighborhoods and Convergents." In *Proceedings of the Colloquium on Convexity*, pp. 104–116, 1965 (1967).

18. **Hammer, P. C.,** "Isotonic Spaces in Convexity." In *Proceedings of the Colloquium on Convexity, in Copenhagen, Denmark, 1965*, pp. 132–141, 1967.

19. **Hammer, P. C.,** "Extended Topology and Systems Theory." *Mathematical Systems Theory*, Vol. 1, No. 2, pp. 135–142, 1967.

20. **Hammer, P. C.,** "Extended Topology: Continuity I." *Portugalie Mathematika*, Vol. 24, 1965.

21. **Hammer, P. C.,** "Extended Topology: Continuity II." *Proceedings of the First International Symposium on Extension Theory of Topological Structures and Its Applications, in Berlin, 1967*, pp. 91–100, 1969.

22. **Hammer, P. C.,** "Language, Approximation and Topologies." In *Approaches in Linguistic Methodology*, pp. 33–45. Madison: University of Wisconsin Press, 1967.

23. **Hammer, P. C.** (Editor), *Advances in Mathematical Systems Theory.* University Park, Pa.: Pennsylvania State University Press, 1969.

24. **Hammer, P. C.,** "The Role and Nature of Mathematics." *The Mathematics Teacher*, pp. 514–521, 1964.

25. **Hammer, P. C.,** and **Sister Gregory Michaud,** "Extended Topology: Transitive Relations." *Nieuw Archief voor Wiskunde*, Vol. 15, Ser. 3, pp. 38–48, 1967.

26. **Hammer, P. C.,** "Generalizations in Geometry." *CUPM Report II*, Vol. 18, pp. 115–132, 1967.

27. **Hammer, P. C.,** and **W. E. Singletary,** "Connectedness Equivalent Spaces on the Line." *Rendiconti del Circolo Matematico di Palermo*, Vol. XVII, Ser. II, pp. 343–355, 1968.

28. **Wallace, A. D.,** "Separation Spaces." *Annals of Mathematics*, Vol. 42, Ser. 2, pp. 687–697, 1941.

Epilogue.   The Past's Future:
Estimating Trends by Systems Theory

C. WEST CHURCHMAN

Space Sciences Laboratory and Schools of Business Administration, University of California at Berkeley

This epilogue is an attempt to apply a type of systems theory to inquiry and, specifically, to inquiring systems [1] whose purpose is to predict the future or, in the terminology of this volume, to estimate trends. Hence this epilogue contains a philosophical discussion of the methodology of trend estimation, of which trends in general systems theory constitute a special case.

We can begin with a general system question: What is the relationship between a component of an inquiring system which describes the past and a component of one that forecasts the future? Common sense provides a rather ready answer to the question of the role of the future in attempts to describe the past; the future has no role. That is to say, the past describer need not concern himself with forecasting. Perhaps one might want to go so far as to say that the past describer must necessarily keep himself free of forecasts, lest they bias his descriptions.

Built into this common-sense reply to the question, however, is a preconception. Many of us have come to recognize how treacherous common-sense preconceptions can become, especially as they burrow like ticks into the living flesh of a scientific discipline. A student of mine has been conducting what he calls "black-box experiments." The subject has a black box whose "theory" he is supposed to describe. He gets his information by putting four numbers into the box and then observing the four-digit output. In one of the black boxes the output is the time of day. It takes many of the subjects quite a bit of effort to realize that there is no relationship between what they are putting in and what is coming out, because "time of day" is not one of their preconceptions for such a black box. This situation is illustrative of the kind of fix into which our preconceptions can get us. Professors often tell their students to "write down all of their preconceptions," but this piece of advice may be of little value, because if one could write down his inmost preconceptions

then they would not be "inmost." In a way, it is the broad task of philosophy to shatter the old tablets, so to speak. As Nietzsche said, "All the secrets of your foundation must come to light; when you are uprooted and broken in the sun, your lie will be separable from your truth" [5].

Suppose we begin with the common-sense preconception just mentioned: the preconception of a mind "bound to the past." For such a mind the past is sure; it is a "fact," a firm foundation, value free. The future, however, is unknown, uncertain, vague, treacherous, threatening, and, if you wish, value-loaded. In the past-bound view, we all know how we have lived. But what can we know of life in the future or life after death?

Two historical examples will suffice. David Hume in his famous *Treatise* [4] argues that the future is now known in the sense that direct experience is known. Indeed, from Hume's point of view, the kind of knowledge that arises from experience and memory is totally different from the kind of knowledge that is entailed in forecasting. Hume believes that it is natural for people to try to forecast. Anyone having seen a flash expects that the noise of an explosion will occur, or having seen the heat on the stove anticipates that it will cause a sensation of warmth. But this, says Hume, is expectation based on habit, and is totally different from the kind of knowledge which we acquire from observation. If we were to plot a chart in which the ordinate shows certainty and the abscissa time, then up to the moment of the time of the experience there is no certainty at all. At the time of the experience, there is a sense impression, and if it is intense enough, considerable certainty is attached to it. After this point in time, says Hume, there will be a decay of certainty as memory enters in and begins to distort what has been directly observed.

A second example comes from the story of historical method in the nineteenth century, when von Ranke made the distinction between "official" records in which one can obtain objectivity, and the "subjective" accounts of eye witnesses and other individuals. Von Ranke was arguing that the historian's job is to sift out the subjective accounts that have no real objectivity and to devote his time to assimilating and accurately recording historical events as they are written down in various kinds of records. The similarity between von Ranke's philosophy and the one that many accountants hold seems notable. The operating statement and the balance sheet are frequently regarded as the results of the official records of the company, carefully examined by the accountant; they are not based on subjective impressions of managers and other individuals.

In order to look carefully at the common-sense preconception that the future plays no role in the past, suppose that we write out four propositions for consideration. In order to do this we need to say something about systems and especially their components. In system science, a system is conceived as

a set of components which play the role of serving the basic purposes of the whole system. In designing such systems, the systems scientist has to pay due regard to the way in which the effectiveness of one component is related to that of another.

In the simplest case, we say that one component, A, is "separable" from another component, B, if the effectiveness of A does not depend in any way on the effectiveness of B. If we could write down the relationship in mathematical terms, we would say that A's effectiveness is measured by variables which are causally independent of the activities occurring in B. The concept of separability is often expressed by saying that the separability of the total system can be represented in a linear form, that is, as a linear function of the effectiveness of each of the components. In this regard it should be noted that one could not arrive at such a judgment of linearity without having taken a look at the larger system and made some judgment about it. Hence even in the case where the systems scientist arrives at a linear function some non-linearities have probably crept into his considerations. For example, if two workers are engaged in digging a ditch, it may happen that the effectiveness of one worker is largely independent of the effectiveness of the other. Even in this simple case, however, one might suspect that pure separability does not occur. Indeed, it is safe to say that pure separability never occurs in social systems.

Now let us look at a system the purpose of which is to tell as nearly as possible the accurate story of what has happened, as well as what will happen. In such a system we can identify two components, one of which devotes itself primarily to telling as accurately as possible what has happened (or is happening), and the other to telling what will happen.

The four propositions that we will consider are the following:

1. *The activity of estimating what has happened in the past is separable from the activity of estimating what will happen in the future.* An abbreviated form of this proposition might be "Past reckoning is separable from future reckoning."

2. *Future reckoning is separable from past reckoning.*

3. *Any specific activity of estimating what has happened in the past can be evaluated along an effectiveness scale ranging from zero or a negative number to some maximum positive number.* In other words, this proposition states that it is possible to describe what has happened in the past and to do so with more or less effectiveness. The proposition does *not* state that one can describe the past with complete accuracy; it states only that there are a worse method and a better method of describing the past. A brief version of this statement would be "Knowledge of the past is possible."

4. *Knowledge of the future is possible.* Here, as in proposition 2, I have used the abbreviated form.

Now we can bring in a logician to consider our four propositions; he will tell us that, if these are meaningful statements, then each can be either accepted or denied, and that the result of such acceptances and denials are sixteen possible positions. Thus one can accept all four of the propositions, or accept the first three and deny the fourth, etc. However, there is a consideration which reduces the list of possible opinions which these four propositions express. Suppose, for example, that you believe proposition 4 to be false, that is, you do not believe that knowledge of the future is possible. In the way in which I have expressed the meaning of proposition 4, your denial amounts to saying that any activity engaged in trying to study the future will be absolutely ineffective. Hence you believe that no effectiveness measure is associated with such an activity. If now we look at proposition 2, which in its complete form says that the activity of estimating what will happen in the future is separable from the activity of estimating what will happen in the past, we see that the proposition is largely meaningless if one has already accepted the idea that knowledge of the future is not possible. What the logician suggests at this point is a "vacuous" stipulation regarding the concept of separability, that is, a kind of arbitrary decision as to what is to be done when an activity has no effectiveness measure associated with it. The arbitrary decision made here will be that, if one argues that an activity has no effectiveness with respect to the total system, then one arbitrarily states that such an activity is nonseparable from all other activities. The situation is very much like the one pertaining to the so-called null class in Boolean algebra, where the logician has to decide whether a class that has no members belongs or does not belong to other classes. In extensional logic, it has been customary to say that the null class belongs to all classes; this rule produces certain conveniences in the calculus.

If we make our arbitrary stipulation, it therefore follows that if one denies proposition 4 he will also deny proposition 2. In other words, if he accepts proposition 2 he is committed to accepting proposition 4. This means that one cannot under the arbitrary stipulation consistently accept proposition 2 and deny proposition 4. Similarly, one cannot accept proposition 1 and deny 3.

One final minor point rules out two other possibilities, a position which asserts that knowledge of the past is possible (accepts proposition 3) but is nonseparable from knowledge of the future (denies 1), and goes on to say that knowledge of the future is impossible (denies 4), would be a ridiculous position to take. A similar remark can be made for the "dual" of this, in which past and future are interchanged.

What remains are seven consistent proposals as follows (we use the convention that a prime after the number represents the denial of the proposition):

1, 2, 3, 4:   "Separated past and future."
1, 2', 3, 4:   "Forecasting from the past."
1', 2, 3, 4:   "Past reckoning from the future."
1, 2', 3, 4':   "Past but no future reckoning."
1', 2, 3', 4:   "Future but no past reckoning."
1', 2', 3, 4:   "Integrated past and future."
1', 2', 3', 4': "Skepticism."

With appropriate apologies for this logical exercise, suppose that we now examine these seven consistent statements, or rather all of them except the last. I assume that for general systems theorists there can be no real interest in skepticism, because if one were to adopt it, the whole activity of systems theory would become a kind of sardonic joke—and we are not joking!

In this examination, as I hinted at the beginning, I should like to take both an epistemological and a strategic look at the propositions. By a "strategic" look, I mean that a practitioner might agree, for example, that the future can be predicted, but assert that it is none of his business to predict it.

At the outset I mentioned what I thought would be a common preconception, namely, that one can tell the past but cannot tell the future, or, strategically, it is none of his business to tell the future. This is expressed in the fourth of the list of positions, which I have dubbed "Past but no future reckoning." It is a series of propositions that has often been accepted by strong positivists, or by individuals in disciplines like history who have felt that man can know what his past has been like but is completely incapable of predicting the future even approximately. We will see as we progress in the discussion that this particular piece of common sense has many shades of meaning.

The opposite of the common-sense position is the one I have called "Future but no past reckoning." This says that one can tell very well what is going to happen but cannot tell what did happen. For example, a man whose wife has just told him that she is going to divorce him and marry the iceman believes that he can predict what will happen, but does not have any idea what did happen. However, no discipline of science that I know of would accept this combination of assertions and denials. The past has always been such a fundamental part of scientific inquiry that to deny the possibility of saying anything sensible about it would seem to aim at the very heart of the scientific method itself.

The position that I want to argue most strongly for, and which is the "deadly enemy" of the common-sense preconception, however, is the one

called "Integrated past and future." This position, too, has many different shades of meaning, depending on how the future enters into the determination of the past. I want to give its strongest possible meaning, and for this purpose I shall turn to operations research—specifically, inventory control.

An accountant with a strong empiricist bias will make a distinction between a "report" and "physical fact." He illustrates this in the case of inventory by saying that the report contains the items described by numbers, whereas the physical facts are the items actually in inventory that can be observed. From this illustration one might infer, as did Hume in the discussion above, that the direct observation of the physical condition of inventory is more reliable than the report, the report representing Hume's "decay in memory."

But the question that faces the operations researcher is the meaning of "reliable." His task is to assist the decision maker in controlling inventory; he will do this by trying to decide on the optimal amounts to be ordered into inventory at various points of time.

Now what are the appropriate data that the operations researcher should use in making his study in order to assist the decision maker? An obvious reply to this question, a reply that is contained in many operations research textbooks, is that the operations researcher should examine past invoices. The student is told to make a frequency chart, using certain intervals of time, such as, a day, a week, or a month. This provides the basis of his inferring the probability distribution of demand on inventory. He is also cautioned to observe trends in time, for example, seasonal fluctuations, or gradually rising or falling sales demand, and to extrapolate into the future on the basis of these trends.

These recommendations to the operations research student in fact are based on what I have labeled "Forecasting from the past," that is, they are based on the assumption that past reckoning is independent of future reckoning but not vice versa. A moment's reflection, however, shows the weakness of this position. Suppose, for example, that there is a seasonal fluctuation of demand. Then it may be very sensible during the off-season to reduce prices and increase advertising in order to smooth the demand curve. If this were done, then obviously the use of very careful statistical analysis of past data and an extrapolation of seasonal fluctuations into the future would be largely irrelevant because a new kind of demand system would have been created. In the language of systems theory, it is quite obvious that the demand system is not separable from the inventory system. If one does use past demand and carries out the kinds of extrapolations mentioned above, he is making a very strong systemic judgment, namely, that nothing can be changed about the demand system, for example, because the managers are reluctant to make such changes or else because the customers are fixed in their patterns of purchasing.

The same remarks apply to the determination of cost by operations researchers. Obviously in the case of inventory it is necessary to determine the cost of holding items in inventory. This cost is an opportunity cost. It is an inference as to how a dollar released from inventory could best be spent in some other activity of the firm. Opportunity costs are what some philosophers of science call "counterfactual conditionals" [3]. The counterfactual conditional has the form, "If X were to occur, then Y would occur." In the case of the cost of holding inventory, for example, the counterfactual conditional is "If inventory were to be reduced by such and such an amount, then the released funds could optimally be used to yield P percent return." It is to be noted that the demand on inventory is also an "opportunity demand"; that is, it is based on a counterfactual conditional of the form, "If such and such were to be done to the demand system, then the demand function would be so and so."

What is it that the operations researcher observes in order to provide information for decision-making purposes? In other words, what does one observe in order to verify a counterfactual conditional? At first glance, the problem seems impossible to solve; how can I observe anything in order to judge what would happen (but never does)? This is why Goodman calls these conditionals "counterfactual." Their premises never "in fact" occur in nature. Hence it begins to appear as though operations researchers must be spinning their wheels.

But the situation is not hopeless. If one were willing to make a judgment about the future of the whole system, then on the basis of this judgment he would be justified in using a certain kind of data. Suppose, for example, that one makes a judgment that nothing can be changed about the demand system. Then, on the basis of this judgment and the additional judgment that the system will exist in essentially the same environment as it has in the past, one would be justified in taking past invoices and performing the exercise specified above, that is, extrapolating into the future and using these extrapolations as the basis for calculating optimal inventory policy. In other words, if a strong judgment about the future is made, a certain kind of data bank based on past observation can be said to be "authorized." If no judgment about the future seems sensible to make, then the operations researcher must regard the problem as intractable.

We see that information for decision making is really a compound of at least two kinds of activities: the one concerned with *authorizing* a certain set of data for use on the analysis, and the other with the collection of the data itself. But the authorization procedure is essentially a forecast about the future, because it makes a judgment about the characteristics that a system will or would have. It is in fact much more than a simple forecast, because it must

be a model which permits one to say what would happen if certain things were to occur. In this regard the systemic judgment is much more like a set of differential equations in physics, where the boundary conditions can be changed and one can infer which events would occur under these changes.

It is clear that the authorization of a data bank is "future reckoning." We can now understand how past reckoning is inseparable from future reckoning, because we need to make very strong and effective judgments about the future in order to be able to use the past effectively. I might add that the reverse is also clear; that is to say, effective reckoning of the past is essential, because effective judgments about the future of the system must somehow draw on past experience. Hence future reckoning is nonseparable from past reckoning, and vice versa. From these remarks we can conclude that the operations researcher must adopt the position that I have labeled "Integrated past and future."

What relevance has all of this discussion for the system theorist? We have all seen a great deal of interest in developing systems which forceast technologies, wars, population, etc. Some of these systems merely try to predict without relevance to the utility of the predictions made. Others, however, are obviously designed to aid the decision maker. I would say that the distinction between the two positions is essentially the strategic question of whether or not the forecaster should be involved in what I have called authorization of data banks, that is, whether he should be involved in the very difficult problem of making adequate systemic judgments. One might adopt the position that the forecaster essentially gathers and analyzes the data, and the authorization is made by the managers or by the legal system. This position would argue for a separability of the information system from the decision-making system, whereby the forecaster does one kind of job and the managers or lawyers do the other kind. I think that position is undoubtedly weak in terms of system design. But the real issue depends, so to speak, on the ambition of "futurists." Do they wish to become involved in authorizing data banks and hence in making strong systemic judgments?

I have argued elsewhere [2] that information becomes measurement if the information is widely usable in a variety of contexts. I should guess that most people are engaged in forecasting their results in terms of the user and his characteristics, and are seeking to make forecasting a measurement process. If so, then I would infer that they are strongly involved in considerations of the authorization of data banks based on strong systemic judgments.

In concluding, I should like to make several remarks of a general nature about the "Integrated past and future" position. We are going through an age in which we are reconsidering many of our traditional human values. From the point of view of science of the last century, precision, rigor, and

clarity were desiderata. The scientist, it was believed, should become clear and precise about his position, and his position should be an essentially consistent one. These values led scientists to regard descriptions of the past in terms of the " quality of the reports." Reports should be specific, concrete, and unobjectionable. We note, however, in terms of our earlier discussion, that the quality of being clear and precise may be at variance with the quality of best serving the user. What does the reader think when he reads about a certain trend in general systems theory? If he is sensible he will wonder, " What other trends might there have been?" Perhaps he will sense something sinister in the glittering facades of a new trend. He is indeed raising the counterfactual question again. And the answer to his question must be based on a strong systemic judgment which, I believe, will inevitably be ambiguous, not clear and precise, and certainly not unobjectionable. We live in a world in which we have to make strong systemic judgments in order to reach our decisions, but if we are honest we will see that we will forever fail to find an unobjectionable basis for these systemic judgments that authorize the use of certain data banks.

The quality of a report, therefore, has changed in terms of a new set of values. On the positive side, this new set of values represents a willingness to be as honest as possible about the basis of our decision making. Along with this willingness goes, by necessity, the need to accept ambiguity, vagueness, and incomplete consensus as essential qualities of our " good " reports.

I should like to close with a very general philosophical opinion about which I hope there will be considerable debate, for debate is the essence of everything that I have discussed in terms of systemic judgments. I realize that we have been developing a culture which pays more and more respect to the future—to what will be or should be in 1984, 2000, or 10,000. But in this epilogue I have really been putting in a plea for our respect to the past, to what it was and might have been. It is quite disrespectful for us to assume that the past was simple and easy to describe. What was it like to be alive in the year 1800? No number of historical data could possibly probe the depth and complexity of such a question. The past is as deep an uncertainty and ambiguity as is the future.

Although I appreciate the demand for forecasting to limit and define its task, I also appreciate the need for it to expand its horizons in a systemic manner. Forecasting needs to ally itself with those who are devoting their lives to the worship of the past—the historians, anthropologists, novelists, poets, and the like. There was a time when basic science regarded itself as one form of the adoration of God. The ritual of this form of worshipping God by worshipping the past entails also the enormous and heroic task of telling the future.

PROBLEMS

1. Which would be a more difficult problem to solve:
(a) what it was like to be alive in the year 1900 in New York City, or
(b) what it will be like to be alive in the year 2000 in New York City.
Explain the methods you would use in each case and justify them.

2. It is sometimes said that general systems theory is an "emerging force" in our society. Explain how you would formulate a hypothesis in this statement and how you would test its validity.

3. The word "general" in "general systems theory" could describe either "systems" or "theory." In other words, "general systems theory" could mean a theory of general systems or a general theory of systems. Explain what the difference is in these two interpretations.

4. Does general systems theory include a theory of morality? A theory of theology? If so, what is the evidence for these theories? If not, how can general systems theory claim to be general?

REFERENCES

1. Churchman, C. W., *The Design of Inquiring Systems*. New York: Basic Books, 1971.
2. Churchman, C. W., *Prediction and Optimal Decision: Philosophical Issues of a Science of Values*. Englewood Cliffs, N.J.: Prentice-Hall, 1961.
3. Goodman, Nelson, *Fact, Fiction and Forecast*. Indianapolis, Ind.: Bobbs-Merrill, 1965.
4. Hume, David, *Treatise on Human Nature*. Oxford, 1739.
5. Nietzsche, F., *Thus Spake Zarathustra* (1883). Gateway Paperback Edition, p. 93, 1960.

Appendix. A Guide to the Literature

The most important periodicals in the area of general systems theory are the **General Systems Journal** (a quarterly) and the **General Systems Yearbook**. Both of them are published by the Society for General Systems Research, 2100 Pennsylvania Avenue, Washington, D.C. 20006. Original contributions, as well as tutorial articles, are published in the *Journal*; the *Yearbook* is devoted primarily to publishing reprints of the most important papers in general systems theory appearing elsewhere. The *Yearbook* started publication in 1956; the *Journal*, in 1972.

Mathematical and metamathematical aspects of general systems theory are well covered in **Mathematical Systems Theory**. This journal has been published quarterly since 1967 by Springer-Verlag, 175 Fifth Avenue, New York, N.Y. 10010.

For engineering problems associated with systems, the reader is referred to **IEEE Transactions on Systems, Man, and Cybernetics**, formerly entitled **IEEE Transactions on Systems Science and Cybernetics**. This journal, published quarterly by The Institute of Electrical and Electronics Engineers, 345 East 47th Street, New York, N.Y. 10017, began publication in 1965.

Several other journals should also be brought to the attention of the reader:

Cybernetica: published by the International Association of Cybernetics, Palais des Expositions—Place Andre Rijckmans, Namur, Belgium; since 1958; quarterly.

IBM Systems Journal: published by IBM Corporation, Armonk, N.Y. 10504; since 1962; quarterly.

Information and Control: published by Academic Press, 111 Fifth Avenue, New York, N.Y. 10003; since 1957; monthly.

Information Sciences: published by American Elsevier, 52 Vanderbilt Avenue, New York, N.Y. 10017; since 1969; quarterly.

International Journal of Systems Science: Published by Taylor Francis, 10–14 Macklin Street, London WC2B 5NF; since 1970; quarterly.

Journal of Computer and System Science: published by Academic Press, 111 Fifth Avenue, New York, N.Y. 10003; since 1967: bimonthly.

Journal of Cybernetics: published by Scripta Publishing, 1511 K Street, N.W., Washington, D.C. 20005, since 1971: quarterly.

In addition to journals, proceedings from the following, regularly held

conferences represent a good source of information for general systems theory enthusiasts:

Proceedings of IEEE Systems Science and Cybernetics Conferences.
Proceedings of the International Cybernetic Conferences in Namur, Belgium.
Proceedings of the Annual Symposia of the American Association for Cybernetics.

Finally, here is a list of the most important books directly involved in general systems theory. If a book is referred to in this volume, the specific chapter (or chapters) in which the citations occur is indicated at the end of the listing.

1. **Ashby, W. R.,** *Design for a Brain.* London: Chapman and Hall, 1952. Preview and Chapter 3.
2. **Ashby, W. R.,** *An Introduction to Cybernetics.* New York: Wiley, 1956. Preview and Chapters 1, 4, and 5.
3. **Berrien, F. K.,** *General and Social Systems.* New Brunswick, N.J.: Rutgers University Press, 1968.
4. **Bertalanffy, L. von,** *General Systems Theory.* New York: George Braziller, 1968. Chapters 1 and 8.
5. **Blauberg, I. V., V. N. Sadovskii,** and **E. G. Yudin** (Editors), *Problems of the Methodology of Systems Research* (in Russian). Moscow: Izd. Mysl, 1970.
6. **Blesser, W. B.,** *A Systems Approach to Biomedicine.* New York: McGraw-Hill, 1969.
7. **Boulding, K. E.,** *Beyond Economics.* Ann Arbor: University of Michigan Press, 1968.
8. **Buckley, W.,** *Sociology and Modern Systems Theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1967. Chapters 1 and 6.
9. **Buckley, W.** (Editor), *Modern Systems Research for the Behavioral Scientist.* Chicago: Aldine Publishing Co., 1968. Chapters 1 and 6.
10. **Carzo, R.,** and **J. N. Yanouzas,** *Formal Organization: A Systems Approach.* Homewood, Ill.: Irwin and Dorsey Press, 1967.
11. **Chestnut, H.,** *Systems Engineering Tools.* New York: Wiley, 1966.
12. **Chestnut, H.,** *Systems Engineering Methods.* New York: Wiley, 1967.
13. **Churchman, C. W.,** *The Systems Approach.* New York: Dell Publishing Co., 1968. Preview.
14. **Eckman, D. P.,** and **M. D. Mesarovic** (Editors), *Systems: Research and Design.* New York: Wiley, 1961. Preview.
15. **Ellis, D. O.,** and **F. J. Ludwig,** *Systems Philosophy.* Englewood Cliffs, N.J.: Prentice-Hall, 1962.
16. **Emery, F. E.** (Editor), *Systems Thinking.* Harmondsworth, England: Penguin Books, 1969.
17. **Fox, J.** (Editor), *System Theory.* New York: Polytechnic Press of the Polytechnic Institute of Brooklyn, 1965.
18. **Freeman, H.,** *Discrete Systems.* New York: Wiley, 1965. Chapter 7.
19. **Gray, W., F. J. Duhl,** and **N. D. Rizzo** (Editors), *General Systems Theory and Psychiatry.* Boston, Mass: Little, Brown and Company, 1969. Chapter 1.
20. **Gray, W.,** and **N. Rizzo** (Editors), *Unity Through Diversity* (4 Vols.). New York: Gordon and Breach, 1971. Chapter 1.

21. **Greniewski, H.,** *Cybernetics Without Mathematics.* Oxford: Pergamon Press, 1960.

22. **Hall, A. D.,** *A Methodology for Systems Engineering.* Princeton, N.J.: Van Nostrand, 1962. Chapter 9.

23. **Hammer, P. C.,** (Editor), *Advances in Mathematical Systems Theory.* University Park, Pa.: Pennsylvania State University Press, 1969. Chapters 10 and 12.

24. **Hare, V. C.,** *Systems Analysis: A Diagnostic Approach.* New York: Harcourt, Brace & World, 1967.

25. **Kalman, R. E., P. L. Falb,** and **M. A. Arbib,** *Topics in Mathematical System Theory.* New York: McGraw-Hill, 1969. Preview and Chapters 8 and 10.

26. **Kast, F. E.,** and **J. E. Rosenzweig,** *Organization and Management: A Systems Approach.* New York: McGraw-Hill, 1970.

27. **Klir, G. J.,** and **M. Valach,** *Cybernetic Modelling.* London: ILIFFE Books, 1967; Princeton, N.J.: Van Nostrand, 1967. Preview.

28. **Klir, G. J.,** *An Approach to General Systems Theory.* New York: Van Nostrand Reinhold, 1969. Preview and Chapters 1, 7, 9, 10, and 11.

29. **Kuhn, H. W.,** and **G. P. Szego** (Editors), *Mathematical Systems Theory and Economics.* New York: Springer-Verlag, 1969. Chapter 8.

30. **Laszlo, E.,** *Introduction to Systems Philosophy.* New York: Gordon and Breach, 1971. Chapter 1.

31. **Mesarovic, M. D.,** *The Control of Multivariable Systems.* New York: Wiley, 1964.

32. **Mesarovic, M. D.** (Editor), *Views on General Systems Theory.* New York: Wiley, 1964. Preview and Chapters 1, 8, and 11.

33. **Mesarovic, M. D.** (Editor), *Systems Theory and Biology.* New York: Springer-Verlag, 1968. Chapters 5 and 8.

34. **Mesarovic, M. D., D. Macko,** and **Y. Takahara,** *Theory of Hierarchical Multilevel Systems.* New York: Academic Press, 1970. Preview and Chapters 5 and 8.

35. **Milsum, J. H.,** *Biological Control Systems Analysis.* New York: McGraw-Hill, 1966. Chapter 5.

36. **Milsum, J. H.** (Editor), *Positive Feedback.* New York: Pergamon Press, 1968. Chapter 5.

37. **Rose, J.** (Editor), *Progress of Cybernetics* (3 vols.). New York: Gordon and Breach, 1970. Preview and Chapter 3.

38. **Sheldon, A., F. Baker,** and **C. P. McLaughlin** (Editors), *Systems and Medical Care.* Boston, Mass.: M.I.T. Press, 1970.

39. **Shinner, S. M.,** *Techniques of System Engineering.* New York: McGraw-Hill, 1967.

40. **Whyte, L. L.,** et al (Editors), *Hierarchical Structures.* New York: American Elsevier, 1969. Preview and Chapter 5.

41. **Wilson, I. G.,** and **M. E. Wilson,** *Information, Computers, and System Design.* New York: Wiley, 1965.

42. **Windeknecht, T. G.,** *General Dynamical Processes: A Mathematical Introduction.* New York: Academic Press, 1971.

43. **Wymore, A. W.,** *A Mathematical Theory of Systems Engineering: The Elements.* New York: Wiley, 1967. Preview and Chapters 8, 9, and 10.

44. **Zadeh, L. A.,** and **C. A. Desoer,** *Linear System Theory.* New York: McGraw-Hill, 1963. Preview and Chapter 9.

45. **Zadeh, L. A.,** and **E. Polak** (Editors), *System Theory.* New York: McGraw-Hill, 1969. Preview and Chapter 8.

Index