W. Ross Ashby (1903-1972)

ROGER CONANT

MECHANISMS OF INTELLIGENCE:

Ashby's Writings on Cybernetics

INTERSYSTEMS PUBLICATIONS

Published in the United States of America by Intersystems Publications.
PRINTED IN U.S.A.

ISBN 1-127-19770-3

THE SYSTEMS INQUIRY SERIES

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

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

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

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

Series Editors

BELA H. BANATHY and GEORGE KLIR

Int. J. General Systems
Vol. 1, 1974, pp. 4–7

Professor W. ROSS ASHBY, was internationally recognized as a pioneer and authority of cybernetics. Trained in medicine and psychiatry, he served as a research pathologist, as Director of Research at Barnwood House Hospital, Gloucester, as Director of the Burden Neurological Institute in Bristol, as a Professor in the Biological Computer Laboratory at the University of Illinois, and upon retirement as an Honorary Professorial Fellow at the University of Wales. During the twelve years at Barnwood House (a mental hospital) he produced his famous Homeostat (put together of old RAF parts on Mrs. Ashby's kitchen table) and two books, *Design for a Brain* (1952) and *Introduction to Cybernetics* (1956), composed in Dr. Ashby's private padded cell and since translated into many languages. The decade spent in the United States resulted in a host of publications and was in his own estimation the most fruitful period of his career.

Dr. Ashby's central interest was in mechanistic explanations of brain-like activity. Consistent with his conviction that the brain operates on mechanistic principles, he greatly enjoyed debunking various myths about the magical powers of the brain ("For 2000 years psychology was a simple discussion of Man's highest faculties—most of which he does not possess") and devising mechanical models of behavior, the most famous of which was the Homeostat, deliberately constructed of unreliable components to emphasize that intelligence resides not in clever, high-quality components but in the structure of the whole. Although he constantly searched for simple explanations for behavior, he embraced complexity wholeheartedly and was chiefly interested in nonlinear, richly interconnected systems in which the complex relations constitute the chief object of interest. As a symbol of his interest in relations he carried a chain constructed of three simpler chains interlocked in parallel; he enjoyed watching microscopic ecosystems (captured with fishpole and bottle from Boneyard Creek in Urbana) for the richness of interaction they displayed, and he built a semi-random electronic contraption with 100 double

triodes and watched it for two years before admitting defeat in the face of its incomprehensibly complex behavior. Perhaps it was the triode network which aroused his interest in information theory as a tool for dealing with complexity and for measuring the strength of interactions between variables; his consequent Law of Requisite Variety and development of multivariable information theory are major contributions to the understanding of complex systems.

Professor Ashby had a great gift for making apparently complex ideas seem simple and for illustrating abstract concepts with homely examples ("a certain centipede . . .") He had a gift for seeing significances where others see only trivialities, and principles where others see only facts. He was incessantly enthusiastic and creative; even after retirement he mastered the clarinet, then set about redesigning its man/machine interface for better information transfer. His grave and somewhat forbidding demeanor gave way, when he was engaged in a conversation or lecture, to an animated style in which his unique keen wit and knack for viewing the commonplace from unusual perspectives would soon turn the discourse into a startling stream of surprises. His enthusiasm would then quite overcome his normal reserve, as when he once fired an imaginary six-shooter and the "kick", to everyone's surprise including his own, sent him reeling across the room and to the floor.

Besides being an honest and meticulous scholar, he was a warm-hearted, thoughtful, and generous person, eager to pass to his students the credit for ideas he had germinated himself; he was in addition a modest man who when asked what he wished done with his voluminous unpublished research notes responded characteristically with "Destroy it all" (to give the next generation a chance for rediscovery.)

Those who knew Ross Ashby personally will remember him as a good and unforgettable person, and those who knew him by his works will remember him as a genius and giant of systems science.

ROGER CONANT

PREFACE

The American Society for Cybernetics, at a meeting in the spring of 1980, decided to promote the publication of a series of books containing seminal works of key workers in the area of cybernetics and systems theory. This book, and another centered on Heinz Von Foerster, are the first two such works to emerge.

It was my good fortune to have been at Von Foerster's Biological Computer Laboratory at the University of Illinois during its remarkable golden age in the 1960's when W. Ross Ashby was in residence. As a doctoral student under Ashby I was endlessly amazed at his creativity and energy and inspired by his broad views and by the freedom he exhibited from the mindsets of more ordinary folk. This freedom often showed up as an ability to see things in shockingly novel ways – to give just one example, in his offhand remark that it would be just as difficult to lose in the stock market as to gain. I am grateful to the ASC and to the publisher for the impetus and opportunity to repay, in part, my debt to Ross Ashby by putting together this book.

Ashby's two books, ***Design for a Brain*** and ***An Introduction to Cybernetics,*** are classics in the literature of cybernetics and systems theory, have been reprinted in many languages, and are presumably well known to the reader. However, much of Ashby's work is scattered here and there in journals, conference proceedings, and books not commonly available, and some pieces have appeared only as handouts given to his students at Urbana. It is the purpose of this book to collect many of these into one volume and so to make the work of this remarkable man more readily accessible.

The principles which guided selection of the papers to be included here were:

1. If a work was substantially incorporated into Ashby's two earlier books, it was not included here.

One exception was made in the case of "Requisite Variety and Its Implications . . .", included here because it is such a clear portrayal of the Law of Requisite Variety, one of Ashby's most famous results.

2. Select the minimum number of papers to maximally cover the ideas contained in all works.

There is considerable overlap between Ashby's papers in that some of his favorite themes are mentioned in many separate works, and I have attempted to simultaneously minimize the overlap, the size of this book, and the "losses" – ideas in papers not included here. This task cannot be done to anyone's complete satisfaction, not even mine, and some excellent papers have been left out as a result of this criterion of parsimony. However I believe that the

collection here reflects nearly all of Ashby's intellectual contributions to cybernetics, beyond those included in his two books.

I could not discover any authoritative listing of Ashby's publications. Therefore the listing of his writings which is given at the end of the book has been put together from documents of the Biological Computer Laboratory, from references within Ashby's publications, and from correspondence with his students and friends. Nevertheless, it may be incomplete in some respects and cannot be taken as final.

I should like to thank Stuart Umpleby, president of the American Society for Cybernetics, who provided motivation and a big head start on this project, Heinz Von Foerster who provided material assistance in the effort, the staff of the Library of the University of Illinois at Chicago Circle who tracked down many obscure Ashby articles for me, Henrietta Cokes who did the typing, George Klir who contributed an overview of Ashby's work and suggestions for papers to be included, my good wife, Shirley, who cheerfully tolerated and even supported my preoccupation with the project, and finally Rosebud Ashby, Ross's wife, who sent the photograph and the quote below which she found among his writings.

Roger Conant

"I am something of an artist, not with pencil or paint, for I have no skills there, but with a deep appreciation of the perfect. My taste is wide, for I can appreciate anything superbly done, whether a chapter by Churchill, a statue by Epstein, or even a suggestion by Max Miller. I have an ambition someday to produce something faultless."

"Work on the brain, of the type recorded in my notebooks, was to me merely a delightful amusement, a hobby I could retreat to, a world where I could weave complex and delightful patterns of pure thought, untroubled by social, financial and other distractions . . ."

W. Ross Ashby

CONTENTS

FOREWORD

By GEORGE J. KLIR

I would like to take this opportunity to express my personal feelings about the richness and significance of the intellectual treasury left to us by the late W. Ross Ashby in his writings. Above all, I would like to show that many of the systems principles or ideas he described in his books and papers were well ahead of the intellectual climate at the time of their publication to be properly understood, appreciated and developed. Some of these principles and ideas have only recently become subject of considerable attention and are currently being analyzed and further developed; others are still left undeveloped and represent a rich intellectual resource for future developments in systems research.

For most people, Ross Ashby is known only through his second book, "An Introduction to Cybernetics," republished after its original publication in 1956 [b2] in numerous editions and fourteen different languages. Little less known is his first book, "Design for a Brain," published in 1952 [b1], and the least known seem to be the many papers he wrote in the period 1930 - 1972; they are scattered in a tremendous variety of publications, many of them being hidden in conference proceedings or edited collections of papers.

I have always been amazed by Ashby's remarkable book "An Introduction to Cybernetics." The book, written with superb mastery, is as good an introductory text to basic systems concepts and principles now as it was 25 years ago. It is still a great source of inspiration as well as underdeveloped ideas. During my first meeting with Ross in Namur in 1970, I complimented him for writing such an excellent and lasting book. I told him that it seemed that writing was easy for him. "No," he replied, "I had to learn it the hard way." He then told me that he had suddenly realized at some point during his work on the book that his knowledge of English was not adequate. He put the partially completed manuscript aside and started to study English. After more than two years of rather intensive study of the language, he finally returned to his work on the book. He did not use any part of the previously written manuscript but started completely from scratch; the whole book was completed in a few months. Thus, this is the way in which this remarkable book was made; no wonder it is such a masterpiece.

Ashby's interests in various aspects of systems research can be traced back to the 1940's. This decade is characterized by the emergence and initial development of a number of novel and promising ideas relevant to systems research. These ideas led to the development of new areas such as operations research, control theory, information and communication theory, automata theory, cybernetics, general systems theory, and computer technology. Ashby's development of the concepts of adaptive and self-organizing systems were his main contributions during this decade. Since the 1940's, he had increasingly become involved in and greatly contributed to the area of systems research until his death in 1972.

Since the first time I became familiar with some of Ashby's writings (the late 1950's), I have always felt his great influence upon my own work in systems research [107,109,110]. It seems that many other people involved in systems research feel the same way. For example, a recent survey conducted by Roger Cavallo [109] clearly demonstrates that Ashby was by far the most influential person in the area of systems research. According to this survey, he influenced almost twice as many systems researchers as the second most influential person, Ludwig Von Bertalanffy, who is usually considered the "father of general systems theory."

One of Ashby's great insights was his clear distinction between an object, loosely understood as a part of the world in which someone is interested, and a system defined on the object. He says [b2,p.39]:

> "At this point we must be clear about how a 'system' is to be defined. Our first impulse is to point at the pendulum and to say 'the system is that thing there.' This method, however, has a fundamental disadvantage: every material object contains no less than an infinity of variables and therefore of possible systems. The real pendulum, for instance, has not only length and position; it has also mass, temperature, electric conductivity, crystalline structure, chemical impurities, some radio-activity, velocity, reflecting power, tensile strength, a surface film of moisture, bacterial contamination, an optical absorption, elasticity, shape, specific gravity, and so on and on. Any suggestion that we should study 'all' the facts is unrealistic, and actually the attempt is never made. What is necessary is that we should pick out and study the facts that are relevant to some main interest that is already given. ...The system now means, not a thing, but a list of variables."

It is rather surprising and, in my opinion, unfortunate that the fundamental difference between these two concepts, those of the object and system, is still foreign to many systems researchers on the current scene. Yet, it is a difference which is at the very heart of systems research. Confusions arise when it is not recognized and, as some critics suggested, systems research becomes then the study of everything (every object) and is thus logically empty.

Ashby's clear distinction between objects and systems defined on objects allowed him to recognize that the same object can be viewed (modelled) in many different ways; each is based on such attributes and associated resolution levels which are pragmatically most relevant to the purpose for which the object is investigated. It also allowed him to restrict the notion of complexity to systems and reject it for objects. Let me quote from one of his late papers [99,p.1]:

> ".. although all would agree that the brain is complex and a bicycle simple, one has also to remember that to a butcher the brain of a sheep is simple while a bicycle, if studied exhaustively (as the only clue to a crime) may present a very quantity of significant detail. is, in my opinion, the only workable way of measuring complexity."

One of the unique characteristics of Ashby's work is that the various systems concepts and principles he developed are highly general in the sense

that they are not limited to systems which are based on variables with some particular mathematical structure. He expressed his views on this issue as follows [95,p.103]:

> "The worker who has some training in mathematics can only too easily fall into the habit (or trap) of thinking that a 'variable' must mean a numerical scale with an additive metric. This is quite unnecessarily restrictive, sometimes fatally so. The meteorologist has long worked with his five 'types of cloud,' the veterinarian with the various 'parasites of the pig,' the hemotologist with the four basic types of 'blood-groups.' Modern mathematics, using the method of set theory, is quite able to handle such variables, which are often unavoidable in the behavioral sciences."

The independence of many Ashby's ideas on the scale of the variables involved make these ideas perfectly suited for the so-called soft sciences. Ironically the myth that variables must be "quantitative," which still heavily dominates the soft sciences, seems to be the main inhibitor in this respect.

It is fair to say that the hierarchy of epistemological levels of systems [107,109,115,116], which is increasingly recognized as the necessary "skeleton" of any meaningful framework for systems problem solving, is implicit in Ashby's writings. Although he does not explicitly formulate such a hierarchy, his writings cover at least four of the epistemological levels incorporated in my formulation of the hierarchy [109].

The lowest epistemological level (source system or level 0), defined in my formulation of the hierarchy as a set of variables (partitioned into basic and supporting variables) and a resolution level defined for each variable [109], is clearly just a more precise and complete elaboration of Ashby's definition of a system as a set of variables. His concept of a protocol (or an activity) regarding the chosen variables corresponds then directly to the concept of a data system (level 1) in my epistemological hierarchy.

As far as level 2 in my epistemological hierarchy is concerned (generative or behavioral systems), it is represented in Ashby's early writing by the concept of the state-determined system (machine) [b1], but also by his more general concept of transformation [b2]. In the early fifties, these were quite novel ideas from which some systems theories were developed such as the theory of finite-state machines (automata) and, later, the theory of dynamical systems.

Structure systems (defined as sets of coupled subsystems), which represent level 3 in my hierarchy, are covered in Ashby's writings quite extensively. His primary interest seems to be in investigating the relationship between various properties of parts (subsystems) and the corresponding properties of wholes (structure systems). For instance, for one such property - the state of equilibrium - he derives a simple but important relationship between wholes and parts: "The whole is at a state of equilibrium if and only if each part is at state of equilibrium in the conditions provided by the other parts." [b2,p.83]. The famous homeostat, which he himself designed and built was also motivated by his strong interests in the study

of whole-part relationships.

While most systems researchers focus on problems involving only one or, at most, two epistemological levels, Ashby was probably the first contributor to the emerging area of systems research who managed to integrate in one conceptual framework a considerably larger spectrum of epistemological levels. It is clear from his writings that, contrary to many current systems researchers, he considered the experimental end of the epistemological spectrum as important as the cognitive end. Moreover, his work is an excellent demonstration of a balanced use of the discovery (inductive) and postulational (deductive) approaches to the investigation of systems.

One of Ashby's greatest contributions, the law of requisite variety, is also one of his earliest ideas [b2,43]. Its simplest but most general formulation - "only variety can destroy variety" - opens a number of directions in which it may be developed with potentially profound implications. Although the interest in the law of requisite variety has lately been increasing, it is surprising that some relevant areas, most notably control theory, are still totally ignorant of the whole idea. This unfortunate situation is well characterized by Brian Porter, one of a few control theorists with broader views, in a paper published several years ago [113,p.227]:

> "...it seems ironical that whilst for example, the theory of optimal control has been developed almost ad nauseam in a detailed but rather uncritical way, control scientists seem to have been largely unaware of the existence and importance of Ashby's work. This situation is particularly singular in regard to Ashby's law of requisite variety which has the same crucial significance for regulation and control as has the second law of thermodynamics for physics. Thus, the law of requisite variety - which, incidentally, can be proved very simply by elementary reasoning - imposes strict bounds on the achievable behaviour of regulators regardless of their structure or design: for Ashby's law states quite soundly that the capacity of any physical device as a regulator cannot exceed its capacity as a channel of communication. The intimate interconnection between control and communication expressed by this law surely indicates that this domain could well be a most exciting, rewarding, and important area for future research by systems and control scientists. It is certainly the case that Ashby himself felt that he had only just begun his work in this field at the end of a long and productive scientific life. ...Ashby's law of requisite variety...shows that there are many non-trivial problems waiting to be solved in the systems and control sciences and that - applying the law anthropomorphically - the best methods of tackling these problems are likely to be found by maximising the capacities of systems and control scientists as channels for variety."

Regulation is one concept to which Ashby paid a lot of attention in his writings. He studied both feedback and feedforward regulations and developed general principles for analyzing and designing regulators. Most of his results are described in set-theoretic terms and are applicable to variables of any scale. This might be one reason why his work on regulation has not been utilized in control theory as yet. Control theory has

been largely developed for continuous variables or discrete representations of continuous variables. It might be difficult to integrate differential or difference equations, which have been the basic mathematical tools in control theory, with the general set-theoretic formulation offered by Ashby. Yet, such integration will tremendously enhance the capabilities of control theory and, particularly, its relevance to the "soft sciences."

The phrase "The whole is more than the sum of its parts," which characterizes the central issue of systems research, is frequently considered mysterious by some and trivial by others; it is rarely understood in all its implications. For Ross Ashby, the phrase was neither mysterious nor trivial; he understood it well ad tried to develop methodological tools by which the whole-part relationship could be rigorously analyzed.

In the early 1960's, Ashby published an algorithm by which it can be determined whether or not an n-dimensional relation can be reconstructed from all its (n-k)-dimensional projections (k = 1,2,...,n-1) [67]. He showed that a relation can be reconstructed from the appropriate projections if and only if the set intersection of cylindrical extensions of the respective projections is the same as the given relation. Although the algorithm deals only with a small portion of the whole-part relationship problem, it is significant as the first attempt to clarify this issue. Yet, it was almost unnoticed by the professional community. Even now, more than fifteen years after its publication, the situation is not much better. Indeed, systems models which have recently been developed in many different areas are almost invariably constructed from subsystems. While the subsystems, each associated with a subset of the set of variables of the overall system, are often well validated models of the phenomena involved, the question of the ability to reconstruct the overall system from the given subsystems is almost never raised. It seems that there has been a tendency among many systems modellers to take the reconstructability for granted. It is clear that without an analysis by which the reconstruction ability of a systems model is determined, the model is likely to be fundamentally incorrect and might be vastly misleading.

I was quite impressed by Ashby's insight into the reconstruction problem when I read his paper [67] in 1964, but it took me more than ten years to properly comprehend its significance and become sufficiently motivated to pursue further research in the direction initiated by Ashby. My first paper dealing with the reconstruction problem was published in 1976 [106]. After its publication, a number of researchers joined in the effort to further investigate various aspects of the problem and develop a new methodological area referred to as reconstructability analysis [103,117].

Some contributors to the reconstructability analysis, most notably Gerrit Broekstra, Roger Conant and Klaus Krippendorff, have investigated various aspects of the reconstructability problem in terms of information-theoretic concepts [117]. This is a direction which was also initiated by Ashby in the mid 1960's [71], and it was apparently one of his main interests shortly before he died in 1972 [92,96]. It is fair to say that his demonstration of the relevance of information theory to systems research is one of Ashby's main contributions.

Ashby's strong interest in the reconstruction properties of systems was just one aspect of his larger interest - a permanent search for methods of simplification. The following quote from his remarks at a panel discussion in 1964 [69,pp.166,168,169] describes his views in this respect rather well:

> "...system theory (is) the attempt to develop scientific principles to aid us in our struggles with dynamic systems with highly interacting parts, possibly exceeding 10^{100} who faces problems and processes that go vastly beyond this size. What is he to do? At this point, it seems to me, he must make up his mind whether to accept this limit or not. If he does not, let him attack it and attempt to find a way of defeating it. If he does accept it, let him accept it wholeheartedly and consistently. My own opinion is that this limit is much less likely to yield than, say, the law of conservation of energy. The energy law is essentially empirical, and may vanish overnight, as the law of conservation of mass did, but the restriction that prevents a man with resources of 10^{100} from carrying out a process that genuinely calls for more than this quantity rests on our basic ways of thinking about cause and effect, and is entirely independent of the particular material on which it shows itself. If this view is right, systems theory must become based on methods of simplification, and will be founded, essentially, on the science of simplification. ...The systems theorist of the future, I suggest, must be an expert in how to simplify."

Although Ashby was not a computer scientist, he had unusual talents for using the computer. He demonstrated that it is perfectly meaningful to view the computer as a laboratory of the systems scientist and computer-simulation as one of his most important laboratory tools. He conducted one of the most exemplary computer-based experimental studies in systems science; its objective was to determine the effect of the size of a system (the number of variables involved) and its connectance (the percentage of dependencies among the variables) on the probability of stability in a particular class of systems [90]. The study was restricted to linear dynamical systems. Among other results, it led to the discovery of a critical value of connectance (13%); it is critical in the sense that for a sufficiently large number of variables (10 or more) almost all systems whose connectance is smaller than the critical value are stable, while almost all systems whose connectance is greater than the critical value are unstable. In another study, the class of systems built up from functionally identical finite state machines was experimentally investigated on the computer. The aim of the study was the determination of the dependence of the cycle length and other behavioral characteristics on the size of the system for various types of finite state machines [75]. These, as well as some computer-based experimental studies of systems in which Ashby was involved, made a clear demonstration of the role of the computer as the systems science laboratory.

One of Ashby's unique contributions is his idea of extending the well understood principle of power amplification from the domain of energy systems into the domain of information systems. This idea of information amplification is discussed in both of his books and many of his papers in a number of different contexts such as regulation amplification, adaptation

amplification or design amplification. Its most general form is embedded in the concept of an intelligence amplifier [41]. In this context, Ashby takes the position that intelligence implies the ability of solving problems which, in turn, implies the ability of making proper selections from the totality of possibilities. Hence, he views the intelligence amplifier basically as a selection amplifier. When the problem is regulation, the selection amplifier would take a special form of a regulation amplifier, when the problem is to design systems with given properties, it would take a form of a design amplifier, etc.

The idea of the various forms of information-based amplifiers has profound philosophical as well as practical implications. If such amplifiers are possible, then it is also possible, at least in principle, to build man-made systems capable of solving problems which are beyond the intellectual powers of their designers. Notwithstanding this great theoretical and practical potential, the idea of information-based amplifiers has not been elaborated beyond the conceptual level developed by Ashby himself. It is apparently one of his great ideas which are still underestimated and, consequently, underdeveloped.

Various aspects of systems design and, particularly questions of meta-design, are often discussed in Ashby's writings, especially in his late papers [96,98]. He looked at systems design as a process of regulation and employed some concepts from information theory to develop a number of meta-design principles which are particularly significant for the design of extremely complex systems. The inclusion of these principles in every textbook of systems design is long overdue.

I was able to discuss only some of Ashby's ideas, particularly those which have had some influence upon my own research work. There are many more ideas in his writings regarding topics such as adaptive, self-organizing and self-reproducing systems, systems modelling, induction, prediction, ultrastability, biological computers and others. Some of his ideas have influenced current scientific views and have been further developed (e.g., the concept of state-determined machine or the role of information theory in systems research), some have only recently become subjects of considerable interest (e.g., the reconstructability analysis), but there are still many rich ideas in his writings which have been largely unnoticed or, at least, have not been developed beyond Ashby's own presentation (e.g., the information-based amplification or his metadesign ideas). It is reasonable to expect that the publication of Ashby's main papers in this volume will renew interest in this intellectual treasury and will lead to further development of the many ideas it contains.

THE LAWS OF MECHANISM

INTRODUCTION

How can adaptiveness and intelligence arise from the operation of mechanical laws? What can be discovered about organization, reproduction, information, and the like by looking into the laws of mechanism? How must brains work? Questions such as these led Ashby to probe the laws of mechanism in an attempt to make precise many formerly vague concepts, such as organization.

Ashby's first book, Design for a Brain, is largely concerned with questions of mechanism, and in it are collected the result of many earlier papers on the topic. On the assumption that the reader is familiar with that marvellous book only one such early publication is included in this chapter: "Effect of Controls on Stability," a brief letter to Nature in which Ashby points out that artificially fixing one variable in a complex system (e.g. prices in an economic system) may render the entire system unstable. Such an observation is typical of his work; it is profound while being easy, even trivial, to demonstrate. It is also somewhat typical in being unknown to or forgotten by those concerned with real-world complex systems.

"The Place of the Brain in the Natural World" is an essay on the application of the laws of mechanism to the explanation of neurological and psychological processes: reflexes, instinct, adaptation, learning, memory, and the like. It ends with his deduction, from these laws, that evolution can only take place successfully in an environment which is in some sense weakly connected, a point made famous by Simon [114] in his classic essay, "The Architecture of Complexity."

In "The Set Theory of Mechanism and Homeostasis" Ashby collected the primary definitions of set theory, going deep enough to allow rigorous discussion of relations and of relations between relations. His objective in casting system concepts into set-theoretic terms was not to inject formalism but to promote clarity; indeed, his papers avoid formalism wherever possible, as can be seen in his very informal "proofs." Set theory was for Ashby a device for describing phenomena unambiguously, and in his papers it is used often as a descriptive tool but almost never as a device for mathematical deduction or proofs. (The only major exception is the paper coauthored with Richard Madden, and that is largely Madden's work.) In "Set Theory..." Ashby describes many of the basic concepts of systems theory and cybernetics in the language of set theory; directive correlation, machine-like behavior, feedback, the effect of one variable on another, simplification, equilibrium, and the like. The use of set theory in this way not only promotes clarity and has an aesthetic appeal but also casts cybernetic investigations into a framework to which information theory may be applied.

"Principles of the Self-Organizing System" is a gem, showing Ashby at his best, vanquishing imprecision and vagueness with the sharp language of set theory. The paper was delivered at a Symposium on Self-Organizing Systems, and we can imagine the glee with which he must have delivered its message, which is basically that (1) if taken at face value, self-organization is an illogical concept, and (2) if interpreted reasonably,

self-organization is commonplace and nearly trivial!

In the paper Ashby first makes the point that "organization" is closely associated with the concept of "constraint." Others tend to think of organization as something added to a system; Ashby's point of view is that organization represents a loss, restriction, or constraint on what might have happened. This novel and complementary point of view is highly compatible with the information-theoretic measurement of organization; it also implies that "organization" represents not an objective property of the system under study but a relation between system and observer. The relativity of organization is further brought out in the paper by showing that the same objective system may be viewed as either organized or disorganized according to the observer's perspective. Moreover, organization cannot be adjudged to be good or bad absolutely, but only relative to a given environment. "Self-organization," then, only makes sense in the context of an "organism" interacting its "environment", and here it is nearly trivial, since the laws of mechanism decree that the evolution of a complex system is bound to produce organisms which will "organize" themselves so as to attain adaptation to their environment. This observation is sometimes called the "Steady State Theory of Life and Intelligence". True, the adaptation of an organism may not be one that we like, but to itself, the resulting organization will be good.

"The Self-Reproducing System" is another gem of wit and clarity. How does an organism reproduce itself? Ashby's answer is: it doesn't. "No organism reproduces itself. The only thing that ever has had such a claim made for it was the phoenix, of which we are told that there was only one, that it laid just one egg in its life, and that out of this egg came itself." Disposing of the original illogical concept of self-organization, Ashby formulates an alternative which makes it clear that reproduction requires an environment or matrix and a "form" within the matrix; if more forms like the original eventually appear, reproduction is occurring. With this clear but abstract formulation it is apparent that reproduction is not rare and unusual but is commonplace and not restricted to biological entities; even the regular ripples on a gravel "washboard" road qualify. Indeed, through reading the paper the reader will undergo a conceptual shift: self-reproduction will not seem at all surprising, though its absense would be! Ashby points out that all sufficiently large dynamic systems will eventually become filled with self-reproducing forms. Indeed, in his view, self-reproduction represents an inevitable adaptation to an environment having the property that most disasters strike locally so that there is survival value in dispersing replicate forms, and therefore self-reproduction is simply a corollary of the Steady State Theory of Life and Intelligence.

Being trained in psychiatry, Ashby always displayed a deep interest in how the brain worked. The last three papers in this section represent his attempts to work out properties of the brain and other large systems of simple deterministic elements connected together in complex ways. His dream that the laws of mechanism would lead to deep insights into global behavior of the brain has not yet been fully realized, but these three works are interesting beginnings along that line. In "Instability of Pulse Activity in a Net with Threshold" he showed, with Heinz Von Foerster and

Crayton Walker, that in a large collection of "neurons" each having N inputs, the result of having neurons fire on threshold (i.e. when the density of incoming pulses from other neurons exceeds a preset minimum) is that the network is in stable equilibrium only when no neuron is firing or when all neurons are firing - nothing in between. Since the brain is known to be stable at an intermediate condition the paper is a proof that there must be some stabilizing mechanism at work there, in addition to the inherently unstable threshold mechanism.

The question in "Connectance of Large Systems...," with Mark Gardner, was this: If one takes a large collection of individually stable parts and connects them, by a set of linear equations $dx/dt = Ax$, in such a way that on the average each part is affected by a percentage C of other parts, how is the stability of the whole affected by the value of the connectance C? This Monte Carlo study indicated that there is a threshold phenomenon: for large systems there is a critical value of C. Below the critical value the system is almost certainly stable, and above it, almost certainly unstable. The work has implications not only for brain behavior but for society at large. As the "connectance" in our culture has risen dramatically over the past centuries, might we be approaching a catastrophic threshold of which this paper is a warning?

The last and most ambitious study in this category is reported in "On Temporal Characteristics of Behavior in Certain Complex Systems," which is largely a report on doctoral thesis work done by Crayton Walker under Ashby's direction. It was a Monte Carlo study of behavior in networks of 100 identical automata, connected randomly. There are 256 distinct 2-input, 2-state automata and each type was explored in the study, with particular interest being given to the length of the terminal cycle and the transient time needed to reach it from a random initial network state. Walker has since continued this work along similar lines.

Ashby performed his own experiment with random networks by connecting 200 triodes to one another, using a table of random numbers to determine the connections. He reported that he never could figure out what the network was doing, being utterly defeated by the sheer quantity of information the network displayed in its behavior. This was one of the influences which led to his deep interest in information theory as a tool for investigating behavior in complex systems, the topic of the following chapter.

EFFECT OF CONTROLS ON STABILITY

During the War the introduction of governmental controls has led to many matters being dealt with by an order *fixing* some quantity, price or other variable where a *laissez-faire* system would have allowed them to find their own levels. As examples we have rates of foreign exchange, wages, and prices. Not only has this fixing occurred in many instances during the War, but a further extension of control or planning in peace will probably lead to even more variables being fixed in this way.

It is the purpose of this communication to point out the danger that in any dynamic system the fixing of one variable may render the rest unstable; and it will be shown that there is one type of variable particularly likely to lead to this result. (In a social or economic system the change to an unstable state would be shown by the subsequent growth of various peculiar and undesirable "vicious circles".)

The theory may be shown in the following way: a dynamic system in general, of n variables, has equations of form

$$\frac{dx_i}{dt} = f_i(x_1, \ldots, x_n) \quad (i = 1, \ldots, n).$$

Near a point of equilibrium (at which the fluxions are zero) the equations may without serious loss of generality, be considered linear:

$$\frac{dx_i}{dt} = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \quad (i = 1, \ldots, n).$$

For a system to be stable at the equilibrial point, it is necessary and sufficient that the real parts of the roots of the equation

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0$$

are all negative. (Since we are discussing an actual system, the quantities a_{ij} will all be real.) Further, since we are discussing an equilibrial point which has existed for some time under free conditions, we may suppose it stable.

Now suppose we fix x_n. The stability of the remainder will depend on the real parts of the roots of the equation

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2,n-1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-1,1} & a_{n-2,2} & \cdots & a_{n-1,n-1}-\lambda \end{vmatrix} = 0$$

The stability of the first system by no means implies the stability of the second system. It is clear, then, that fixing a variable may render the rest of a system unstable.

As a numerical illustration, the system

$$\left.\begin{aligned} x_1 &= 6x_1 + 5x_2 - 10x_3 \\ x_2 &= -4x_1 - 3x_2 - x_3 \\ x_3 &= 4x_1 + 2x_2 - 6x_3 \end{aligned}\right\}$$

leads to the equation

$$\lambda^3 + 3\lambda^2 + 26\lambda + 60 = 0;$$

and this has roots $-2{\cdot}44$, $-0{\cdot}28 \pm 4{\cdot}95\,i$, where $i = \sqrt{-1}$. The real parts being all negative, the system is stable. But if we fix x_3, we have a system with determinant

$$\begin{vmatrix} 6 & 5 \\ -4 & -3 \end{vmatrix},$$

and as the roots are now + 1 and + 2, the system is unstable.

We can, however, go further than this. Since the sum of the roots is equal to the sum of the elements in the main diagonal, Σa_{ii}, any change making this less negative will tend to make the system less stable – other things being equal (the argument here is admittedly imprecise). So the fixing of x_n would be particularly likely to lead to instability if a_{nn} was large and negative. We can identify such variables without difficulty; for, as they behave in accordance with the equation

$$\frac{dx}{dt} = \xi + ax,$$

where ξ is independent of x, but changes with time, while a is large and negative, such a variable (x) will always have the properties that (1) it always moves towards $-\xi/a$, (2) it moves towards $-\xi/a$ *quickly*, (a) as $-\xi/a$ has a as denominator it will be small, and therefore the fluctuations of x will be small.

It is concluded, therefore, that: (1) To fix a sociological or economic variable by order carries some danger of rendering the system, or parts of it, unstable

(the latter being shown by the subsequent development of various "vicious circles").
(2) The type of variable more particularly dangerous from this point of view is one which, under free conditions, changes value at high speed, and, by these quick changes affecting the other variables, fluctuates only through a narrow range.

Not being an economist, I cannot give detailed instances, but I have little doubt that some could be provided.

CURRENTS IN MODERN BIOLOGY 1 (1967) 95-104. NORTH-HOLLAND PUBLISHING COMPANY AMSTERDAM

THE PLACE OF THE BRAIN IN THE NATURAL WORLD

W. Ross ASHBY
Electrical Engineering Research Laboratories
University of Illinois, Urbana, Illinois, USA

Received 15 January 1967

A great deal is known already about the brain, but most of our knowledge of it is still in the form of experimental and observational facts. With the growing interest in the brain's more general properties however, such as in "artificial intelligence" in its various forms, the time has come for an abstract formulation of the nature of "brain", a formulation suitable for a *direct* translation to the computer or hardware. The paper gives such a formulation on the basis of set theory and the concept of the state-determined system.

1. THE NEW MECHANIST

We can, of course, at once characterize the brain by saying that it is a collection of physico-chemical parts, each part acting on many other parts, and each part determined by the physico-chemical laws (to the extent demonstrated by the neurophysiologists). But the brain has so many parts that our usual methods for dealing with parts acting together become practically inapplicable, and we must stop to reconsider the situation.

After the properties of the nerve cell have been ascertained, there remains the task of relating these properties to those shown by organizations of such nerve cells in large numbers. The properties that emerge are those of the behaving organism, and it is important to appreciate at once that the properties of the behaving organism are by no means to be deduced directly from the properties of the single nerve cell, for most of the organism's behavioral properties are due to *interactions* between nerve cells - they are what the physicist calls "cooperative" phenomena. For this reason, any study of the relation between nerve cell and behavior must pay great attention to questions of interaction. The subject tends to be complex, and used to be thought forbiddingly so; but during the past twenty years such great advances have been made in our understanding of complex mechanisms that the subject can now be treated with some clarity.

Attempts to treat the relation of nerve cell to behavior were much hampered, before about 1940, by the fact that the would-be scientific mechanist possessed, as examples, only such simple machines as the clock, the lathe and the typewriter. So he saw that, obviously, no "machine" could correct its own errors, could predict, could have initiative, and so on. These generalizations were correct enough over the machines of pre-1940 type, but as machines have since then developed altogether new powers, many of the old generalizations are today quite wrong.

The nature of the new machines (and of the ideas of the New Mechanist, as we might call him) may be most simply made clear by a brief mention of the events in history. Prior to about 1940, mechanisms (and the "classical" physics that thought about them) were of a "cause-effect" type in which typically one cause led to one effect, then the process was complete: the wound watch ran for 24 hr, then stopped: the lathe, switched on, ran round a cycle endlessly: the typewriter, when a key was pressed, printed the letter and stopped. With these machines before him, the psychologist theorized similarly: stimulus elicits response, stop; a dog is subjected to a cycle of flashes and reinforcements and it develops a conditioned reflex, stop. More complex theories of behavior could not be formed because no one knew how to think about complex behavior.

Then Howard H. Aiken built Mark I. Charles Babbage had understood the principles of sequential machinery a century earlier but had been unable to solve the purely mechanical problems. But Mark I worked. Here was a proof by construction that "mechanism" could include the type that strung causes and effects into chains of unlimited length, each cause evoking its effect, and each effect being itself the cause of the next step. Go-stop became go-go-go-..., and an entirely new wealth of computational behavior had

been shown to be possible.

A second essential contribution occurred (a little earlier, in fact) when the radio engineers tamed "feedback". It was very early discovered that adding "reaction" to a radio receiver was a most powerful method of improving its performance, but at the price of making the receiver almost crazily uncontrollable. For 30 years the battle raged, but by 1940 the engineers had won - they understood feedback and could use it as a useful tool. This understanding was a second essential contribution to the new science of behavior, for most biological systems are rich in circular actions. Before 1940 the lack of understanding made any realistic treatment impossible; even to mention the existence of the circular actions was apt to bring the discussion to a shocked silence, as each person reflected on his inability to think clearly about such things. Today, however, we are no longer afraid of the topic for we know that it can be understood and that it has its own logic, theorems and methods. We can see, too, that the advance has consisted in understanding the properties of long chains of cause and effect, in this case acting round and round the same physical circuit (instead of linearly onwards as in Mark I's processes).

With the radio receiver and Mark I before them, it has been possible for the mathematicians and engineers to extend their methods over the same realm. Here they were much helped by meeting the current of mathematical thought developing from Whitehead and Russell's work (1925) that had been attempting to free mathematical thought from its excessive pre-occupation with the continuous, the linear, and the analytic. Their work was fully developed by the French school that writes collectively under the pseudonym of N. Bourbaki, who showed in detail how all mathematical processes could be seen as specializations from certain basic operations on "sets of elements". All that is required of the "elements" is that they are unambiguously identifiable. They may be the numbers $1, 2, 3, \ldots$ or the points on a line, but they may just as well be the five types of cloud distinguished by the meteorologist, or the three types of cry emitted by a species of bird, or the four modes of progression of a horse. Thus, if the biologist or psychologist has ideas definite enough to serve as a basis, he has the necessary material to which modern mathematical logic can be applied. Thus arises the possibility of a fully rigorous science of behavior. It starts with the data of the *observer*, uses these as elements in modern set theory, and so enters the rigorous world of mathematics and logic (e.g., Ashby, 1952, 1966).

The new mathematics, or logic, of behavior should not be confused with the "mathematical biology" of the beginning of this century. At that time the sole mathematical methods available were the analytic, continuous and linear forms developed originally for the solution of Newtonian problems; after the biological data or concepts had been modified to fit the unyielding mathematical frame they were often only a caricature of the biological reality. The new mathematics, however, is quite free from any need to distort. Its first steps have been guided rather by the needs and outward forms of mechanical computation, but the biologist should not allow the presentation to mislead him. The theories, now well developed, of the "finite-state machine" (Gill, 1962), of the "noiseless transducer" (Shannon and Weaver, 1949), of the "state-determined system" (Ashby, 1952), and of the "sequential circuit", are essentially homologous. All treat, basically, the case of the system whose next state is determined by the immediately preceding state, a case so common in natural systems as to be regarded by many as absolutely universal. All the theories show that this (apparently) simple restriction carries, in fact, deep and wide implications.

The statement of this fundamental property of the state-determined system may take several forms, and the student of the theory of behavior must be prepared to recognize them in whatever form they occur. The simplest form states explicitly what the next state (x') is as a function of the earlier (x):

$$x' = f(x). \tag{1}$$

Thus if $x' = x + 0.7$, the state $x = 0.2$ would be followed by x becoming 0.9, and then 1.6, 2.3, etc. If x is thought of as having a sequence of values ($x_1, x_2, x_3, \ldots$, say) at steps $1, 2, 3, \ldots$, then the same equations would appear as

$$x_{n+1} = f(x_n). \tag{2}$$

Sometimes x is thought of as a function of the time t and written $x(t)$; if time advances by steps of duration Δt the equation would then appear as

$$x(t + \Delta t) = f[x(t)]. \tag{3}$$

An equivalent method is to specify not x's new value, but by how much, Δx, it has changed from its previous value. Then $\Delta x = x' - x = x_{n+1} - x_n = x(t + \Delta t) - x(t)$, and the equations would take the form

$$\Delta x = g(x_n) \quad \text{or} \quad g[x(t)], \tag{4}$$

where g is the function: $f(x) - x$.

Should the steps become infinitesimal, over time interval dt, Δx becomes dx, and the equation becomes that of an ordinary first-order differential equation:

$$dx/dt = g(x). \tag{5}$$

The symbol x, assumed above to represent one value of a set of values, may be a vector, with n components say. If these component variables are $x_1, x_2, \ldots, x_n$ (where the subscripts distinguish variables, and not steps as in eq. (2)) the equations may take the extended form:

$$\begin{aligned} x_1(t + \Delta t) &= f_1[x_1(t), \ldots, x_n(t)], \\ x_2(t + \Delta t) &= f_2[x_1(t), \ldots, x_n(t)], \\ &\cdots\cdots\cdots \\ x_n(t + \Delta t) &= f_n[x_1(t), \ldots, x_n(t)]. \end{aligned} \tag{6}$$

If the steps are infinitesimal, the equations become a set of simultaneous ordinary first-order differential equations:

$$\begin{aligned} dx_1/dt &= g_1(x_1, \ldots, x_n), \\ &\cdots\cdots\cdots \\ dx_n/dt &= g_n(x_1, \ldots, x_n). \end{aligned} \tag{7}$$

Sometimes the subscript is itself continuous and the g's may show some special relation. Such is the case with diffusion (of heat or solute) when the temperature or concentration (x) changes with time in a way that depends on the neighboring temperatures, along a linear distance measure y: then the equations become (e.g.)

$$\frac{\partial x}{\partial t} = k\frac{\partial^2 x}{\partial y^2}. \tag{8}$$

What is important here is that *all systems that behave in ways specifiable by any of the types above are subject to the new logic of mechanism.* In this way many branches of knowledge that started independently can be brought together and given a unified theory of behavior.

It is important that the reader appreciates that though this theory includes many of the results of mathematical physics (such as eq. 8 above) it is not restricted by them. Its basic concept is the "mapping". It is always from one set (its "domain") to a set (its "range") and is a rule (or process or transition or change or any other correspondence) that gives, for each element in the domain, one and only one element in the range. It is the "one and only one" that characterizes the "mapping", not some physical action; thus if the two sets are mothers and daughters, the mapping is *from* daughters *to* mothers; for while each daughter has one and only one mother, each mother may have more than one daughter. Below, if mapping μ turns element x (in the domain) to y (in the range), we shall write

$$\mu(x) = y. \tag{9}$$

The domain and range may sometimes be the same set; thus "square it" maps the set of integers into the set of integers. When this is so, the mapping can be repeated, generating from the elements

$$\mu(x), \ \mu^2(x), \ \mu^3(x), \text{ etc.}$$

In general

$$\mu^{n+1}(x) = \mu[\mu^n(x)], \tag{10}$$

and by comparing this equation with eq. (2) before we see that the mathematical concept of a mapping gives us what is needed to represent a state-determined system.

The subject can hardly be taken further without technicalities, out of place here. But I hope enough has been said to make clear that the following propositions are defensible: 1) the modern theory of mechanism, being founded on the concept of a mapping, includes the scientific knowledge gained in the past; 2) the modern theory of mechanism, by considering cause-effect relations in great numbers (both in long sequences and with feedback), provides a technique and logic adequate for the complex facts of biology and psychology; 3) the modern logic of mechanism is fundamentally a logic of behavior (not of matter or of energy).

The New Mechanist, then, feels equipped to attempt the bridging of the gap between neuron and behavior, but he is also aware, perhaps exceptionally so, of the vastness of the gap to be bridged! Fifty years ago it seemed so simple: stimulus goes in, response comes out - what more is required? It was assumed that all that was necessary was plenty of reflexes, with a little integration to weld things together. Unfortunately, with the growing understanding of mechanism has come a growing appreciation of the conceptual distance that separates the activity of the neuron from the behavior of the whole organism. The situation today is not unlike that in physics, when the designer of a steel bridge reflects that his art rests on quantum physics. No bridge-designer today appeals directly to the laws of quantum physics: the connection has to be in several stages, through atomic motions, crystal structure, the strength of metals, prac-

tical metallurgy, the strengths of girders, to the whole structure. It seems likely that the gap from neuron to behavior will similarly have to be bridged in stages. Such an attempt will be sketched in the sections that follow.

2. EVOLUTION AND EQUILIBRIUM

"The behaving organism" is discussable from three very different points of view. One can discuss its inner consciousness, its awareness. I shall say nothing from this point of view for I have nothing to say; the problem is one of extraordinary difficulty, involving the subtlest questions of philosophy and scientific method.

The second point of view considers its creative aspect, as when a man invents a new system of musical harmony or produces Joyce-like prose. I shall say nothing of this matter, for I have no objective criterion by which I can distinguish such productions from the squeakings of a gate, or from a sequence of words generated by a dictionary and a table of random numbers. When "anything goes", science has little to say.

The third point of view regards the behaving organism as one fashioned for survival: as a system highly adapted to its environment, molded by evolution and natural selection, and able, especially in the species *Homo*, to produce extremely complex patterns of behavior that show (to the Mechanist) astonishingly complex adaptations to the environment. Here the brain is seen simply as an organ that furthers survival. This third point of view has today achieved some completeness, in that there remain no large gaps that are wholly mysterious. What we see today will be sketched below.

Our starting point is the well established fact that this earth solidified about five billion years ago and that ever since the conditions affecting its surface have either been quite constant - the laws of energy, the law of gravity, the properties of carbon, of water, for instance, - or have changed only slowly - its temperature, the quality of sunlight reaching it, the composition of the ocean.

The logic of mechanism now becomes applicable. The earth's surface will have, at each moment of time, a well defined state - the position of each sand grain, the temperature at each point, the distribution of each species, and so on. The laws of nature, acting at each point, determine how that state will change. Since every state goes to *some* state, and never goes to two, the laws of nature specify a *mapping* of the set of possible states into the same set. (Whether this formulation is wholly true is not yet known; it is certainly true to a major degree, and it is also the universally held hypothesis that guides the scientist in his daily work. We shall, in this article, write on the assumption that it is wholly true; the complications caused by atomic indeterminacy would cause us repeatedly to say "statistically determinate" or "on the average", but the modification would cause no major alteration from the outline given below, so we shall ignore the complication.)

Let us call the mapping (induced by the laws of nature and its basic forces) L, so that, if it acts on state s, it changes s to $L(s)$. Now, saying that the laws and conditions on the earth's surface have been largely unchanging means that the mapping (or operator) L has been unchanging in time, so that the sequence of operators at work has been the repetitive sequence

$$L, L, L, L, L, L, L, \ldots,$$

and not, say

$$L, P, K, M, L, J, M, \ldots,$$

as would be the case if the laws or conditions had varied appreciably. This (trivial-looking) observation will in fact give us a secure origin for a rigorous treatment of the origin of life and intelligence.

We start from the fact that the first sequence shows high redundancy (in the sense defined by Shannon and Weaver, 1949) and thereby shows constraint; for the larger set (all sequences composed of $J, K, L, M, P, \ldots$ in some order), is restricted to a subset (those composed only of element L). Bourbaki has shown, especially in the section "Echelles d'ensembles et structures", that restriction to a subset is always the essential operation that generates properties, relations, patterns, structures (as the words are used in ordinary language or with special precision in mathematics). Thus, from the theory of mappings we would expect the sequence of states generated by repetition of one operator to show special features. One way in which such special features appear is at the states of equilibrium - those that satisfy the relation

$$L(s) = s. \tag{11}$$

Such states are of the highest importance in the study of the brain as an organ for survival.

It should be noticed that most of the "classic" examples of equilibrium - the pendulum hanging motionless, the run-down watch, the mixture of chemicals when all reaction is exhausted - are

far from typical today, for they occur in systems of extreme simplicity and complete dependence on a finite quantity of energy. The "open" system, on the other hand, can still be a "machine" (as a computer is by being state-determined) but may show equilibria (sometimes distinguished as "steady states") of vastly greater richness and interest. The richness comes from the fact that what is regarded as a "state" from one level of study (and therefore unanalyzed) may, on closer inspection, be found to have a rich internal structure. Thus the Roman Empire remained recognizably the same entity over hundreds of years, in spite of many disturbances, while a closer examination shows that in fact a vast number of personal activities and changes were contributing to the stability of the Empire as a whole. In such cases a rigorous treatment is still possible: the set S of states s is "stable" under L if (by definition)

$$L(S) \subset S, \tag{12}$$

i.e., if L, acting on S, produces no new states. Thus, though L may change state s_i to s_j, its action is only to cause changes within the set S: and if S has some characteristic property, this property is not lost when L acts. ("Stability after displacement" is a special case in which L is the composite mapping $\lambda^n\delta$, where δ is the operator that effects some displacement, so that $\delta(s) \neq s$, and λ is the mapping whose repeated operation eventually brings the state from $\delta(s)$ back to s.)

The statement "all systems tend to equilibrium" embodies much experience but is too vague for a rigorous theory of behavior. The cases where it is not true, however, all seem to be highly specialized and to demand exact construction. Even if a system has no states of equilibrium (or cycles) it will tend towards certain "preferred" regions or sets of states, there being no preference (no convergence or divergence of the trajectories in the phase space), only when, if the system is specified by

$$dx_i/dt = g_i(x_1, \ldots, x_n) \quad (i = 1, \ldots, n), \tag{13}$$

the g's have the special property that everywhere

$$\frac{\partial g_1}{\partial x_1} + \frac{\partial g_2}{\partial x_2} + \ldots + \frac{\partial g_n}{\partial x_n} = 0. \tag{14}$$

Similarly, if the system is stochastic and Markovian it will also tend to some "preferred" states unless the matrix of transition probabilities has not merely its rows but also its columns adding to 1. Thus, there is certainly some justification for the statement "systems tend to equilibrium". In any particular case, the mathematically acceptable form would have to be developed to suit the details of the case.

We can thus say that most systems with unchanging laws (i.e., systems in unchanging conditions) change towards states (or sets of states) in which they linger; by showing a convergence towards such states they generate relations between the law and the system's state (and between its component parts). The relation generated is that of "adapted for survival". At the primary level the relation is truistic; in its consequences in complex systems it develops unlimited complexity.

At equilibrium, the relation between the parts is necessarily holistic and one of *coordination*. One example will be given to show what is meant. Let the matrix

$$\begin{bmatrix} 1 & 0 & 4 \\ 3 & 1 & 2 \\ 7 & 2 & 5 \end{bmatrix}$$

be the operator, or law, or drive, that changes (by multiplication modulo 12) such a state as the vector

$$\begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix} \quad \text{to the state or vector} \quad \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix}.$$

In this case the state has been changed. The vector

$$\begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}$$

however, is unchanged by the operator:

$$1 \times 2 + 0 \times 5 + 4 \times 3 = 14$$
$$3 \times 2 + 1 \times 5 + 2 \times 3 = 17$$
$$7 \times 2 + 2 \times 5 + 5 \times 3 = 39$$

and these exceed 12's multiples by 2, 5, and 3, respectively: thus the state is regenerated, and is equilibrial. What is important here is that the three component values - 2, 5, and 3 - act *cooperatively* to preserve each other. The 5, for instance, depends on the 2, since the 5 was obtained from

$$3 \times 2 + 1 \times 5 + 2 \times 3.$$

Had the 2 been a 1, the outcome would have been

$$3 \times 1 \;+\; 1 \times 5 \;+\; 2 \times 3,$$

i.e., 14 or 2 (as in the first example).

The cooperative action is quite general. At any state of equilibrium (whether simple or having complex internal structure), the parts always interact so that the action of all is to regenerate the state of each. Thus, if we were simply to watch one act of regeneration we might well say: this set of parts is acting (within these laws of nature) so as to preserve its condition unchanged.

Given, then, that a dynamic system is isolated, there exists today a completely rigorous theory showing why it should pass to states or forms characterized by being self-preservative in their behavior. But the "state or form" must be equated to the system "organism *plus* environment". At an oasis, for instance, the well keeps alive the villagers, and the villagers repair the well: the permanence of both is due to their appropriate interaction. The "adaptation" shown by the living organism is always to some property of its environment: let the environment change, and what was an adaptive way of behaving may become grossly inappropriate.

The relation (that ensures survival) becomes conceptually simpler in those cases in which the stable set of states is "cylindrical", that is, in which some of the variables stay within clearly marked limits. These variables are then recognized as the "essential" variables of the adapted system - those that must be kept within "physiological" limits if the whole is to survive: the supply of food, the volume of circulating blood, the pressures that threaten the continuity of bone and skin, the body temperature in the warm-blooded animals.

The rigorous theory of behavior thus joins the physiologist by regarding *homeostasis* as the core of all adaptive behavior. In a study of behavior the essential variables are easily overlooked. They stay almost constant, while the non-essential variables range widely, change rapidly and generally catch the observer's eye. Nevertheless, the dramatic activities of the non-essential variables are secondary: they have meaning and relevance only because their vigorous changes act to keep the essential variables within limits. Should they have other effects, these effects are, to the student of fundamentals, mere by-products.

The word "homeostasis" was originally coined by Walter B. Cannon (1932) to describe these processes in so far as they occurred in the autonomic, vegetative and internal processes of the living organism. The rigorous theory of behavior, however, stresses that any boundary here, between internal and external, is ultimately arbitrary, that many of the processes use both internal and external factors, and that common principles govern both. For these reasons it is of the highest importance, if the basic unity of behavior theory is not to be lost, that the observer should be able to relate the free-ranging activities of the non-essential variables (the organismic behaviors) to the homeostases that they ultimately achieve, and that gives them their ultimate significance. Only in this way can the purposeful free-ranging activities be distinguished from those activities that are merely the expression of force in action. Intrinsically, of course, there is no difference: they differ only in their relevance to some understood homeostasis.

From this point of view, *Homo* is simply a species that has specialized in the development of extremely complex free-ranging movements and that has managed to obtain some of their advantages while avoiding most of their dangers. He can melt steel without burning himself, he can build and run motors without being torn to pieces, and he can send electricity over a continent without being electrocuted. By viewing Man's activities in this way, we can trace an unbroken line of deduction from the primary fact that the earth has long been isolated, to the emergence of extremely complicated systems that always tend to behave homeostatically.

Let us now examine these free-ranging activities, these behaviors, in more detail, so as to see how the well known behaviors of Man can be related to their underlying mechanisms. We will examine first the very simple pieces of behavior called "reflexes", then the more complex types called "instinctive", and finally those that depend largely on "learning". All will be regarded as manifestations of the primary mappings that underly (and drive) all physico-chemical events, shaped to homeostatic form by natural selection.

3. THE REFLEX ORGANISM

The reflex, in its many forms, presents little difficulty in general theory since it is now known, and has been proved in many ways, that *if the parts have certain minimal properties, a sufficiently large and complex set of them can produce any well defined behavior*. Thus the old question: can a machine do it? is always to be answered yes; provided that the required activity is capable of unambiguous description in operationally meaningful terms. All the reflexes of

physiology are so capable, and the "problems of the reflex" are essentially those of discovering the particular details of each particular reflex.

That many of them are regulatory, i.e., homeostatic, offers no unusual difficulty today, for it is now well known that such regulation demands only the provision (by the gene pattern) of a feedback (any chain of cause and effect from the main effect back to the original cause) so arranged as to be "negative", i.e., so that the returning effect subtracts from, and thus tends to annul, the original disturbance. In complex cases the "subtractive" operation may have to be understood in a technically sophisticated way, but the basic idea remains. The theory of such feedbacks is now extensively developed, partly in the theory of servo-mechanisms. The physiologist who wishes to use the essentially homeostatic aspect of such feedbacks should not deprive himself of the greatly increased knowledge and technical assistance that he can get from the theory of feedback mechanisms.

The worker in biological subjects, however, will notice that such theories tend to be unduly specialized to those cases in which the feedback is linear and continuous. Biological regulators often tend to be grossly non-linear (the response not proportional to the stimulus) and non-continuous. The study of such regulations at the reflex level calls for no new principles, only new techniques. Such non-linear and non-continuous regulations appear in their most complex forms in the learned reaction (referred to later).

4. THE INSTINCTIVE ORGANISM

While the reflex organism offers little difficulty in a mechanistic representation, the difficulty is small only so long as the reflex remains simple. The modern theory of mechanism, however, envisages mechanism unbounded in complexity, both in respect of their numbers of working parts and in respect of the complex conditionalities governing their internal activities.

The "instincts", as they are recognized and listed today, have each of them a goal, but this fact does not remove them from the class of mechanisms, for any describable piece of behavior that has some permanence and some ending must have a bound and therefore stability in some sense - the purely transient and easily diverted is neither noticed nor named. Thus any system with many states of equilibrium offers the possibility of being described as having as many goals and "instincts". The machines of everyday life seem to lack them only because these machines are too poor in ways of behaving and in complex steady states. Let a machine be made with modern activity and richness of possibilities and the observer can soon name many trends in its behavior that, in a living organism, would claim recognition as "instincts".

There was a time when the instincts were thought to be fundamentally different from the reflexes because the instinct was often evoked by some situation or event that could not be identified with any specific physical or chemical event. Thus dogs tend to bark whenever "something strange" occurs, and "something strange" cannot be identified with any particular sound or any particular stimulation of the retina. It is now known, however, that this property of reacting to combinations and *relations* between stimuli, is readily obtained from the mechanism, if the mechanism works in stages or levels so that the first level "computes" various functions of the primary stimuli, then the later levels compute functions of these functions, and the final stage acts only if these "computational" processes have resulted in some actual physical event at the penultimate stage. In this way any defined function over the primary stimuli, however complex or subtle it may be, can be transformed, in a purely mechanistic way, to a physical event suitable to act as physical cause for the instinctive action. The apparent distinction between reflex and instinct arose partly because the older theories were based, mostly unconsciously, on a one-level model: stimulus-to-response, without intermediate processing.

The organism, developing through evolution, thus develops ever more complex mechanisms, improving its ability to react homeostatically in ever more complex ways to the disturbances and threats of the environment. How orderly is this progression? Here the logic of mechanism is adamant: in general, every new addition, every extension, leads to an essentially new total system whose properties are also new. Only when there are special simplicities will the new extension give new behaviors that merely add to the set of behaviors already available. In general, no matter how large the machine and how small the alteration made to it, if the machine is not restricted we can put no limit to the size of the change that may occur in its behavior. The mathematician knows the corresponding fact that if the f_i's are unrestricted in

$$dx_i/dt = f_i(x_1 \ldots x_n; \alpha) \quad (i = 1, \ldots, n), \tag{15}$$

then the behavior of this system, when the parameter α has a particular value α_1, does not restrict in any way the behavior of the system in which α has been changed to $\alpha_1 + \Delta\alpha$.

This fact, unpleasant though it be to those looking for simplicities, is fundamental in the theory of mechanism and the analysis of behavior. Just as a few chemical elements (C, H, N, O) can be put together in many combinations to form the many compounds of organic chemistry, so any other units (transistors, neurons, molecules) can be put into combinations whose variety of behavior is not limited by the uniformity of the units but *can be as various as the number of their combinations*. This richness is precisely the richness that comes from the presence of active interactions between the units, simplicity occurring (with many units) only when the interactions are small.

It should be noticed that the behavior of the whole system cannot be predicted by our regarding the whole system as made up of various feedback loops, finding the behavior inherent in each separate loop, and then adding (or combining in any way whatever) the separate behaviors. Systems have been constructed, for instance, in which every possible loop has the feedback negative (so that each, by itself, would be stable), yet the whole shows the ever increasing divergence of instability (e.g., Ashby, 1952). In a similar way the logic of mechanism shows that we have no right to expect that the instincts (as complex trends observed in the behavior) will be simple, tidy, or neatly classifiable. Here the field worker or clinician has the last word, for only he can say what instincts are worth distinguishing, defining and naming.

5. THE LEARNING ORGANISM

The nature of learning and memory, once so mysterious, has been completely clarified in the last twenty years, and it is now possible to see them as processes entirely homologous with the other processes occurring in matter. On the basic hypothesis (strongly supported by two centuries of scientific work) that all processes in matter (above the atomic level) are state-determined, the concept of "memory" becomes appropriate when an observer, unable to observe every variable of the system (and thus finding it unpredictable), restores predictability by taking into account earlier events in what he can observe. Thus, if the system of three variables - x, y, and z - can only be observed at x, the observer may well find that the value of x at time $t + 1$ is predictable provided he knows the values of x at times t, $t - 1$ and $t - 2$. From the point of view of information theory the change is quite simple: variables $y(t)$, and $z(t)$ are replaced by variables $x(t-1)$ and $x(t-2)$, exactly as they might be replaced by any two functions of x, y and z.

The parallel is exact. If the test (in Shannon's notation)

$$H(x) + H(y) - H(x, y) \neq 0$$

shows that "transmission" (in his sense) is occurring between x and y we have the ordinary case of transmission between two spatially separated variables, between two points in the nervous system, say. If the test

$$H[x(t)] + H[x(t-k)] - H[x(t),\ x(t-k)] \neq 0$$

holds, then "transmission" (defined in just the same way) is occurring between the events at x at time t and those that occurred at x, k units of time earlier. When this is so we have the essential property that allows us to think of "memory of duration k" at x. (The fact that the value at $t - k$ must necessarily become the value of $x(t)$ at k units of time later does not necessarily enter into the computations and is, from this point of view, irrelevant.) Thus the modern logic of mechanism is able to treat the basic epistemological properties of "memory" as a transmission over time exactly homologous with the well understood transmission over space.

That transmission should be physically possible between two times demands some special physical mechanism exactly as does transmission between two places. The mechanism used, though varying widely in some ways, will always use some form of equilibrium, for the attempt to carry some state from one time to another, without corruption and without loss of information, demands that something be invariant over the interval, and "invariance" is the core of "equilibrium".

What physical or chemical "state" is used to be invariant, to carry the "memory", is of little importance in the larger questions of behavior: all that is necessary is that the state should have certain properties; how these properties are achieved may be decided by matters of purely local significance. Clearly, as no one expects one method to be used for all the many transmissions from place to place in brain and body, there is as little reason to expect that only one method will be used as "memory basis" for transmission from earlier to later. Much more likely is that the organism will use a variety of

methods, each adapted to the needs of its particular purpose.

The details of the memory trace are thus of little significance in the larger questions of behavior. What is of more significance is the method used for making the record in the first place and for using it advantageously later. At the moment, our scientific thinking tends to be grossly misled by the example of the big digital computer. It has a big memory store, kept far from the working parts, which send recordable facts to special places, and then later go back to exactly the same places to regain the information. Such a method, demanding vast numbers of exactly connected lines, can hardly be achieved in biological machinery, especially as such machinery must use parts subject to injury, starvation, infection and similar disturbances. More likely is it that most of the brain's memory traces occur, and are retained, at the site of their action. It seems likely, therefore, that the traces that contribute to a particular reaction (e.g., to answering "What is your name?") will be widely scattered, each having only a very small effect, yet amounting in their total effect to a decisive determination of the behavior. The concept of "memory" will have to become that of "the memories", rather like the "animal heat" of the Middle Ages, as a unity, became all that is known today of metabolism and oxidation. At one end of the types of memory are those simple events that leave a permanent mark on behavior, those often called "painful" or "terrifying". These learnings often use innate mechanisms developed by natural selection, ready to learn and record what is painful but requiring the details to be provided by the child's particular environment. The trained mechanism (giving the behavior of the "burned child") is clearly homeostatic. It is a homeostatic mechanism whose final details of design have been postponed until the information necessary has been supplied by the environment. Such a method of developing a homeostatic mechanism shows in essence all the necessary features of the learning process; the "higher" forms are essentially similar, carried to a far higher degree of complexity.

6. COMPLEXITY

The logic of mechanism, used quantitatively, shows that any mechanism as complex as the human brain could never have been brought to an adequately self-preserving form, either by evolution or by personal learning, if it and its environment had been richly connected (both internally and between the two) - the possibilities would be so vast that all geological time would not be adequate for the working-up of the unorganized nerve net to an adapted form (Ashby, 1952). The arrival of most children at a reasonably well adapted adult state is possible only because the adaptation can be developed piece-meal. Thus our terrestrial environment allows the child to learn how to pour water into a cup independently of what he has just learned about the English language, and these again are independent of his learning what a dog does if you pinch it. Sometimes what has to be learned is not wholly separable, but allows the learning to occur in stages, each of which can be established with reference only to what was established earlier. Thus arithmetics can be learned in the order: addition, subtraction, multiplication, division, but not in the reverse order. And pole jumping must be preceded by learning to stand, to walk, to run and to manipulate long objects. Because our terrestrial environment allows the full adaptation of the adult to be developed largely in small stages, the process is much simpler than would be the case in the full generalization.

Nevertheless, as soon as a quantitative estimate is made of how much information must come to the observer if he is to understand fully what is before him in a human organism, so soon do we find that the quantity of information is likely to exceed all bounds of what is possible (e.g., Bremermann, 1965), even with the most generous allowances. It seems clear that when we leave the old methods of thinking about the brain, with their gross oversimplifications, and change to the modern methods, we shall have to take the question of the quantity of information seriously, lest we waste time attempting the impossible.

The question of "complexity" must play a dominating role in our attempts to understand the brain (whether natural or artificial), for once we leave the mechanisms that we knew before 1940, we arrive at forms whose complexity increases with overwhelming rapidity. Most properties in them increase, not as the volume or mass but with combinatorial speed, so that the order of their increase is either with e^n or $n!$, or much faster still (e.g. Ashby, 1966). Thus, the possibility of adaptation occurring in any reasonable time is fundamentally dependent on the presence of simplicities. The case in which interaction is incomplete or weak does, in fact, occur very commonly in our terrestrial environment, and what is known suggests that the brain

has been profoundly shaped by evolution to take advantage of this fact. Thus, the extremely common method of working to a major goal by the achievement of a sequence of sub-goals is one expression of this adaptation.

More can hardly be said at the moment, for the general study of complex dynamic systems reacting with a complex environment has only just begun. Today, however, enough is known of the logic of mechanism to show that the general principles by which the properties of neuronic units can be related to the larger behaviors of the whole organism can be traced with a rigor limited only by our resources of time and patience.

ACKNOWLEDGEMENT

The work on which this article is based was supported by the U. S. Air Force Office of Scientific Research under Grant AF-OSR 7-63 and by the U. S. Department of Public Health under Grant GPH 10718-01.

REFERENCES

Ashby, W. Ross, 1952. Design for a Brain (John Wiley and Sons, New York).
Ashby, W. Ross, 1956. An Introduction to Cybernetics (John Wiley and Sons, New York).
Ashby, W. Ross, 1966. Mathematical models and computer analysis of the function of the central nervous system. Ann. Rev. Physiol. 28, 89.
Bourbaki, N., Eléments de mathématique, esp. Théorie des ensembles; fascicule de résultats. ASEI 1141 (Hermann and Cie., Paris).
Bremermann, H. J., Quantum noise and information, in: Fifth Berkeley Symposium on Mathematical Statistics and Probability, 27 December 1965 (in press).
Cannon, W. B., 1932. The Wisdom of the Body (Norton, London).
Gill, Arthur, 1962. An Introduction to the Theory of Finite State Machines (McGraw-Hill Book Cy., New York).
Shannon, C. E. and W. Weaver, 1949. The Mathematical Theory of Communication (University of Illinois Press, Urbana).
Whitehead, A. N. and B. Russell, 1925. Principia mathematica (Cambridge University Press, Cambridge).

The Set Theory of Mechanism and Homeostasis†

W. ROSS ASHBY‡
University of Illinois

1. INTRODUCTION

The last twenty years has seen science invading regions hitherto avoided—the world of dynamic systems that are intrinsically complicated. For a hundred years such dynamic systems as the cerebral, the social, the economic, the protoplasmic, the colloidal, were treated mostly by the methods of classical science; by the attempt to reduce the whole system to one of many simple units, with only infinitesimal interactions. The advent of statistical and matrix methods, however, began to enable the scientist to deal more successfully with the moderately complex. Then came the large general-purpose computer; while it confronted him with yet another extremely complex system, its clear logic of behavior so educated him that today the whole logic and strategy for dealing with the highly complex system has become immeasurably clearer; helped too by the discipline of information theory, he has been able to achieve a new clarity and a new rigor.

To the biologist, the need for a new rigor may not be at once apparent. Yet

† Part of this work was supported by the National Science Foundation, Grant 25148.

if biology is to study and understand the really complex system the methods it uses must be appropriate. Foremost among these is that of "simplification": not by the intuitive rules-of-thumb commonly used so far but by the more developed methods that use homomorphisms. The method promises much, but its use demands rigor and technique; where are they to be found?

The method described in this paper is offered because the author during the last twenty years [1, 2] has found it invaluable as a guide. As its concepts are initially quite free from any implication of either continuity, or of order, or of metric, or of linearity (though in no way excluding them), the method can be applied to the facts of biology without the facts having to be distorted for merely mathematical reasons.

The method described here is based on the work of the French school that writes under the pseudonym of N. Bourbaki. As their great work [3, 4, 5, 6] has shown that *all* mathematics, and therefore all products of accurate thinking, can be based on set theory, so there is considerable advantage in keeping the method in this paper wholly aligned with theirs; we can thus ensure ready and safe interchangeability between this method and all mathematics. Their "Fascicule de résultats" [3] has therefore been taken as basis for this method. (Their full, three-volume "Théorie des ensembles" [4] seems to me to add little of value to the biologically oriented worker.) I have also drawn substantially on the work of J. Riguet [7, 8, 9, 10], who has extended Bourbaki's work in the direction of making it algebraic and of providing it with a calculus. Theorems due particularly to him are acknowledged in the text.

One final advantage of the method is that it is ready at every stage to admit the various measures of the "quantity of information", such as those of Shannon [11] and of McGill and Garner [12, 13]. The study of the really large and complex system is dominated everywhere by the extent of the quantity of information and whether it exceeds the information-processing resources of the investigator. That the method makes almost intuitively obvious how the quantity of information would be measured is not the least of its advantages.

2. THE ALGEBRAIC SET THEORY OF MAPPINGS AND RELATIONS

2.1 SET THEORY

A. *Set and element*

We start with the ideas of "set" and "element" taken as understood. What is essential is that we must be able to say with certainty of any element x and of any set A whether element x is or is not contained in set A. The fact will be written as $x \in A$ or as $x \notin A$. If the set is described by the naming of its individual constituent elements, it will be written within braces, for example as $\{a, b, c\}$. Repetitions of an element within a set (should they occur for any reason) will be ignored; we shall assume that the elements are all distinct. The empty set, that with no elements, will be represented by $\{\}$, as a single symbol. (In general, capitals will be used for sets and lower case letters for elements.)

If two sets are such that every element of A is also an element of B, we write $A \subset B$. If A and B are composed of the same set of elements we write $A = B$. The relation $A \subset B$ does not exclude that of $A = B$.

B. *Complement*

Given two sets A and B, the set $A - B$ is defined as consisting of those elements that are in A but not in B. If A is some basic set that can be taken for granted, $\bar{B}$ will signify the set of elements not in B ("but in A" understood). $\bar{B}$ is the *complement* of B; it has no meaning in the Bourbaki set theory unless some total set is defined or understood.

C. *Implication*

If statement P implies statement Q, i.e. if P's being true implies that Q must be true, or if the condition P holding implies that condition Q must hold, we shall write $P \Rightarrow Q$. When both $P \Rightarrow Q$ and $Q \Rightarrow P$, we shall write $P \Leftrightarrow Q$.

D. *Quantifiers*

"$\exists x: P \ldots$" is to be read as meaning: "There exists (within an already defined or understood set) at least one element, let's call it x, that has the property P, or that makes the statement P true".

"$\forall x: P \ldots$" is to be read as meaning: "Every element (within an already defined or understood set) has the property P, or makes the statement P true".

"$\exists x: x \in A$ and $\ldots$" may be abbreviated to "$\exists x \in A: \ldots$"
"$\forall x: x \in A$ and $\ldots$" may be abbreviated to "$\forall x \in A: \ldots$"

The following *formulae* are easily verified. More formal proofs are discussed in Section 2B. A and B are any sets. (Parentheses and brackets are used freely in the formulae below to help make the meaning clearer.)

1D.1 $(A = B) \Leftrightarrow [(A \subset B) \text{ and } (B \subset A)]$.
1D.2 $(A \subset B) \Leftrightarrow \forall x: [(x \in A) \Rightarrow (x \in B)]$.

(The expression on the right reads: For every element, *if* it is in A then it must also be in B.)

1D.3 $(A = B) \Leftrightarrow \forall x: [(x \in A) \Leftrightarrow (x \in B)]$ (read correspondingly).
1D.4 $a \in \bar{A} \Leftrightarrow a \notin A$ (some total set being understood).
1D.5 $A \subset B \Leftrightarrow \bar{B} \subset \bar{A}$ (some total set being understood).
1D.6 $A = B \Leftrightarrow \bar{A} = \bar{B}$ (some total set being understood).

E. *Union and intersection*

Given two sets A and B, their *union*, written $A \cup B$, is the set of elements that belong either to A or to B or to both.

1E.1 $a \in (A \cup B) \Leftrightarrow (a \in A)$ or $(a \in B)$.
1E.2 $(A \subset B) \Leftrightarrow (A \cup B = B)$.
1E.3 $(A \subset B) \Rightarrow [(A \cup C) \subset (B \cup C)]$.

The *intersection* of A and B, written $A \cap B$, is the set of elements that belong to both A and B.

1E.4 $a \in (A \cap B) \Leftrightarrow (a \in A)$ and $(a \in B)$.
1E.5 $(A \subset B) \Leftrightarrow (A \cap B = A)$.
1E.6 $(A \subset B) \Rightarrow [(A \cap C) \subset (B \cap C)]$.
1E.7 $\overline{A \cup B} = \bar{A} \cap \bar{B}$.
1E.8 $\overline{A \cap B} = \bar{A} \cup \bar{B}$.
1E.9 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
1E.10 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

F. *Mappings*

Given two sets E and F, a *mapping* (from E to F) is any correspondence, rule, method, diagram, indication, construction, process, algorithm, computation, machine, device, force, drive, reflex, instinct, command, or any other cause whose effect is that, given any element in E, *one and only one* element in F results. (In Bourbaki: une application.)

E is the mapping's *domain*; F is its *range*, in which the mapping takes its *values*. F is not necessarily different from E. At this point it should be noticed that whether the sets E and F are finite or infinite, ordered or not, discrete or continuous, with a metric or not, are all irrelevant.

If the mapping μ, operating on e of E, gives f in F, we write $\mu(e) = f$. (Greek lower case letters will be reserved in this paper for mappings.)

If A is a subset of E, and μ acts on each element of A, the set generated is some subset of F. Thus, given each subset of E, the action of μ on the elements generates *one and only one* subset of F. There is thus defined a mapping of the set of all subsets of E into the set of all subsets of F. Though essentially distinct from μ, experience has shown that the use of the same symbol μ to represent it is convenient and rarely a source of confusion. Thus, if $A = \{a_1, a_2, a_3, \ldots\}$, we have

1F.1 $\mu(A) = \{\mu(a_1), \mu(a_2), \mu(a_3), \ldots\}$

with the original μ on the right hand and the new μ on the left.

(The one-one mapping, so popular in much of mathematics, need not be defined here, since it is nowhere used in this paper.)

1F.2 $x \in \mu(A) \Leftrightarrow \exists y: y \in A$ and $x = \mu(y)$
1F.3 $A \subset B \Rightarrow \mu(A) \subset \mu(B)$.
1F.4 $\mu(A \cup B) = [\mu(A)] \cup [\mu(B)]$.
1F.5 $\mu(A \cap B) \subset [\mu(A)] \cap [\mu(B)]$.

The particular mapping that maps a set into itself by the rule: $\forall x: \lambda(x) = x$, is the *identity* mapping. It can conveniently be represented by symbol 1, as there is little in set theory to be confused with it. (It is, of course, quite different from the Boolean 1.) Its domain is an essential characteristic (there are many 1's differing in their domains), and the domain must always be borne in mind or indicated. The identity mapping on the domain A will be written as 1_A.

1F.6 $\forall x \in A: 1_A(x) = x$
1F.7 $1\ (B) = A \cap B$
1F.8 If $B \subset A$, $1_A(B) = B$.

G. *Representations of a mapping*

Several are possible. Skill in their selection may convert a difficult and obscure argument to one that is almost immediately obvious.

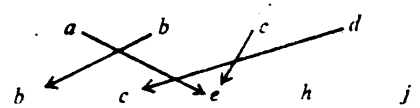
The *explicit* representation simply names, for each element in the domain, its transform in the range; for example if the domain is $\{a, b, c, d\}$ and the range $\{b, c, e, h, j\}$, μ might be represented by

$$\mu(a) = e, \quad \mu(b) = b, \quad \mu(c) = e, \quad \mu(d) = c$$

More compactly, it may be written

$$\mu: \begin{array}{|cccc} a & b & c & d \\ \downarrow & & & \\ e & b & e & c \end{array}$$

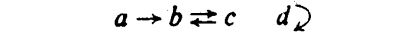
The *sagittal* representation shows, by a set of arrows, how each element of the domain goes to one in the range; for example.



If the mapping is of a set into itself, for example of E into E by

$$\begin{array}{|cccc} a & b & c & d \\ \downarrow & & & \\ b & c & b & d \end{array}$$

the sagittal representation would be



The *tabular* representation is given by a rectangle with columns corresponding to the elements of the domain, with rows corresponding to those of the range, and a mark at those intersections (one to each column) that correspond to the mapping.

The *matrix* representation is a tabular representation with a 1 at each intersection and a 0 elsewhere. The operations of set theory then correspond to those of matrix algebra, provided that the orientation of the representation is that described above and the multiplication is of rows into columns.

While it is essential, especially when the systems are biological, that we should be able to use mappings and sets that are wholly arbitrary or structureless, many of the applications in chemistry and physics use mappings of restricted or specialized types. As these are, in a sense, classical, their relations to the wholly arbitrary will be indicated. They all represent some way of taking advantage of some *redundancy* in the details of the mapping. Thus the mapping

$$\mu: \begin{array}{|cccccc} 0 & 1 & 2 & 3 & 4 & 5 \\ \downarrow 1 & 2 & 3 & 4 & 5 & 0 \end{array}$$

can obviously be condensed to

$$\mu(x) = x + 1 \quad \text{(modulo 6)}$$

In general, the mapping can be specified as $\mu(x) = f(x)$, where $f(x)$ is some well known function that can be written briefly. If $\mu(x)$ is written as x', the equation becomes of the form

$$x' = f(x);$$

and if x is a function of n or t, the mapping may be written as

$$x_{n+1} = f(x_n) \quad \text{or as} \quad x(t+1) = f(x(t)).$$

If the emphasis is on the *change* of x, and if x is numerical (so the subtraction is possible), $x_{n+1} - x_n$ can be written as Δx, and the mapping can be represented as the difference equation $\Delta x = g(x)$. If the steps become infinitesimal, dependent on an infinitesimal change in time, the mapping is naturally represented by an ordinary differential equation of the first order:

$$\frac{dx}{dt} = h(x).$$

(The functional symbols have been changed from f to g to h to make clear that the form f does not remain unchanged.)

All that has been said is in no way restricted to a single variable, for x may be an n-tuple of components (which may also be finite in number or infinite, discrete or continuous in values, with or without a metric.) When x is an n-tuple, further constraints between the variables composing x may be expressed by partial differential operators; such is the elementary equation of heat conduction

$$\frac{\partial x}{\partial t} = k \frac{\partial^2 x}{\partial y^2}.$$

Thus many of the well known equations of physics and chemistry that describe a system's behavior are in fact specifications of mappings.

H. *Inverse of a mapping*

Given a mapping μ of E into F, the inverse operation, written μ^{-1}, turns any element in F into all those elements in E that correspond to it under μ. Thus, from the example of Section 2.1G above,

$$\mu^{-1}(b) = b, \quad \mu^{-1}(c) = d, \quad \mu^{-1}(e) = \{a, c\},$$

while $\mu^{-1}(h)$ and $\mu^{-1}(j)$ have no elements. Thus the inverse of a mapping is commonly not a mapping.

1H.1 $x \in \mu^{-1}(y) \Leftrightarrow y = \mu(x)$.
1H.2 $x \in \mu^{-1}(A) \Leftrightarrow \exists y: y \in A$ and $y = \mu(x)$.
1H.3 The inverse of mapping 1 is the same mapping 1 (avoiding writing 1^{-1} as the two 1's have different meanings).

I. *Composition of two mappings*

If μ is a mapping of E into F, and λ is a mapping of F into G, there necessarily exists a mapping of E into G defined thus: each element in E gives, by μ, one and only one element in F; this element gives, by λ, one and only one element in G. The element in F is $\mu(e)$, and that in G is $\lambda(\mu(e))$. The rule thus gives, for each element in E, one and only one element in G; it therefore defines a mapping. It will be written $\lambda \circ \mu$, which may usually be conveniently abbreviated to $\lambda\mu$. Notice that μ operates first. $\mu \circ \mu$ will be written μ^2, etc. Only when the two mappings have a common dimension for elimination can the composition be performed.

(The sets E, F, and G being defined or understood):

1I.1 $(\lambda \circ \mu)(e) = \lambda(\mu(e)) = \lambda\mu(e)$
1I.2 $g \in \lambda\mu(e) \Leftrightarrow \exists f: f \in F$ and $f = \mu(e)$ and $g = \lambda(f)$.
1I.3 $g \in \lambda\mu(A) \Leftrightarrow \exists a: a \in A$ and $g = \lambda\mu(a)$.
1I.4 $g \in \lambda\mu(A) \Leftrightarrow \exists a: a \in A$ and $\exists f$: (etc. as in 2).
1I.5 $1 \circ \mu = \mu \circ 1 = \mu$ (provided that $E = F = G$, and 1's domain contains μ's.).

J. *Kinematic graph*

When a set is mapped into itself, the sagittal representation drawn on the one set used both as domain and range shows, by the chains of arrows, the sequence of values that will occur if any element e is operated on repeatedly by μ, giving the values $\mu(e)$, $\mu^2(e)$, $\mu^3(e)$, ... The successive values can be thought of as represented by one point that moves along the chain of arrows, generating a *trajectory*.

If the number of elements in the set is finite, the trajectory always ends in some final set, a *basin*, which it occupies infinitely often. The set of elements that are led by the kinematic graph to one basin is a *confluent*. The confluents are a partition of the elements in the domain.

2.2 LOGIC FOR SET THEORY

It is convenient here to tabulate and to make precise the logical methods that will be used later for proving theorems. The formulae below are readily established by examining their applications to elementary cases in which the correct answer is already known with certainty. They are listed here chiefly for convenience of reference.

A. *Formulae*

Let P etc. be a property that an element may or may not have, or a statement that may or may not be true.

2A.1 "*P and Q*" means "properties P and Q both hold", or "statements P Q are both true".
2A.2 "*P or Q*" means "at least one of P or Q holds, perhaps both", etc.
2A.3 P and (Q and R) $\Leftrightarrow$ (P and Q) and R.
2A.4 P or (Q or R) $\Leftrightarrow$ (P or Q) or R.
2A.5 P and (Q or R) $\Leftrightarrow$ (P and Q) or (P and R).
2A.6 P or (Q and R) $\Leftrightarrow$ (P or Q) and (P or R).
2A.7 $\exists y$: (P or Q) $\Leftrightarrow$ ($\exists y$: P) or ($\exists y$: Q).
2A.8 $\exists y$: (P and Q) $\Rightarrow$ ($\exists y$: P) and ($\exists y$: Q).
2A.9 (If P is a property of x and y)
$\exists\langle x, y\rangle$: $P \Leftrightarrow \exists x$: ($\exists y$: P) $\Leftrightarrow \exists y$: ($\exists x$: P).
2A.10 $\forall y$: (P and Q) $\Leftrightarrow$ ($\forall y$: P) and ($\forall y$: Q).
2A.11 ($\forall y$: P) or ($\forall y$: Q) $\Rightarrow \forall y$: (P or Q).
2A.12 $\forall\langle x, y\rangle$: $\bar{P} \Leftrightarrow \forall x$: ($\forall y$: P) $\Leftrightarrow \forall y$: ($\forall x$: P).
2A.13 $\exists x$: ($\forall y$: $\bar{P}$) $\Rightarrow \forall y$: ($\exists x$: P).
($\bar{P}$ is the negation of property or statement P.)
2A.14 $\overline{P \text{ and } Q} \Leftrightarrow \bar{P}$ or $\bar{Q}$.
2A.15 $\overline{P \text{ or } Q} \Leftrightarrow \bar{P}$ and $\bar{Q}$.
2A.16 $(P \Rightarrow Q) \Leftrightarrow (\bar{Q} \Rightarrow \bar{P})$.
2A.17 $(P \Leftrightarrow Q) \Leftrightarrow (\bar{P} \Leftrightarrow \bar{Q})$.
2A.18 $\overline{\exists y: P} \Leftrightarrow \forall y: \bar{P}$.
2A.19 $\overline{\forall y: P} \Leftrightarrow \exists y: \bar{P}$.
2A.20 $(P \Rightarrow Q) \Leftrightarrow (\forall y: (\bar{P} \text{ or } Q))$.
2A.21 $\forall y$: (P or Q) $\Leftrightarrow (\bar{P} \Rightarrow Q) \Leftrightarrow (\bar{Q} \Rightarrow P)$.

B. *Illustrative proofs*

1. To show (1E.9) that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

$x \in A \cup (B \cap C)$

(by 1E.1) $\Leftrightarrow (x \in A)$ or $(x \in (B \cap C))$.
(by 1E.4) $\Leftrightarrow (x \in A)$ or $(x \in B$ and $x \in C)$.

(by 2A.6) $\Leftrightarrow (x \in A$ or $x \in B)$ and $(x \in A$ or $x \in C)$.
(by 1E.1) $\Leftrightarrow (x \in A \cup B)$ and $(x \in A \cup C)$.
(by 1E.4) $\Leftrightarrow x \in (A \cup B) \cap (A \cup C)$.
$\therefore$ (by 1D.3)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

2. To show (1F.5) that $\mu(A \cap B) \subset [\mu(A)] \cap [\mu(B)]$.

$x \in \mu(A \cap B)$

(by 1F.2) $\Leftrightarrow \exists y$: $y \in (A \cap B)$ and $x = \mu(y)$.
(by 1E.4) $\Leftrightarrow \exists y$: $y \in A$ and $y \in B$ and $x = \mu(y)$.
$\Leftrightarrow \exists y$: $y \in A$ and $x = \mu(y)$ and $y \in B$ and $x = \mu(y)$.
(by 2A.3) $\Leftrightarrow \exists y$: $[\{y \in A$ and $x = \mu(y)\}$ and $\{y \in B$ and $x = \mu(y)\}]$.
(by 2A.8) $\Rightarrow \exists y$: $[y \in A$ and $x = \mu(y)]$ and $\exists y$: $[y \in B$ and $x = \mu(y)]$.
(by 1F.2) $\Leftrightarrow x \in \mu(A)$ and $x \in \mu(B)$.
(by 1E.4) $\Leftrightarrow x \in (\mu(A) \cap \mu(B))$.
(by 1D.2) $\mu(A \cap B) \subset [\mu(A)] \cap [\mu(B)]$.

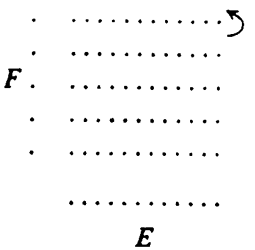
2.3 COMPONENTS

All that has been said so far has treated the elements as atoms, unanalyzed, and it is most important that we should recognize that the properties and relations so far established hold rigorously without further analysis of the elements into parts or components. Nevertheless, such analysis is often advantageous, so it will now be treated.

A. *Product set*

Given two sets, E and F, the *product* set $E \times F$ is the set formed by taking all possible pairs of elements, the first from E and the second from F. Each pair is an *element* of the set $E \times F$; it will be written $\langle e, f\rangle$. The elements have no particular order in $E \times F$, but within the couple the order is essential: $\langle e, f\rangle$ is not equal to $\langle f, e\rangle$ unless $e = f$.

A product set may be given the tabular representation for example



in which the direction should be clearly shown, i.e. which is the first factor and which the second.

3A.1 $\langle e, f\rangle \in E \times F \Leftrightarrow e \in F$ and $f \in F$.
3A.2 $\langle e, f\rangle \in E \times F \Leftrightarrow \langle f, e\rangle\ F \times E$.
3A.3 $(A \cup B) \times C = (A \times C) \cup (B \times C)$.
3A.4 $(A \cap B) \times C = (A \times C) \cap (B \times C)$.
3A.5 If $A \subset C$ and $B \subset F$:

$$\overline{(A \times B)} = (\bar{A} \times F) \cup (E \times \bar{B}).$$

3A.6 If $A \times B \neq \{\}$, then

$$A \times B \subset C \times D \Leftrightarrow A \subset C \text{ and } B \subset D.$$

Products of more than two sets are formed similarly, thus

3A.7 $\langle e, f, g\rangle \in E \times F \times G \Leftrightarrow e \in E$ and $f \in F$ and $g \in G$.
3A.8 $\langle e, \langle f, g\rangle\rangle \in E \times (F \times G) \Leftrightarrow e \in E$ and $\langle f, g\rangle \in F \times G$.

When there are many sets they may more conveniently be indexed (as E_1, E_2, E_3, ... if the indices are numerical), or, more generally, when the indices may be any arbitrary set, as $(E_i)_{i \in I}$, where I is the set of indices with i a typical element. The product of such a set may be written $\prod_{i \in I} E_i$.

B. *Partial mappings*

When a mapping has a product set as its domain, it defines a set of *partial* mappings based on the elements of the factor sets, in the following way. Suppose μ maps $E \times F \times G$ into H. If the domain is restricted to $\{e_1\} \times F \times G$, each couple $\langle f, g\rangle$ is mapped, like $\langle e_1, f, g\rangle$, into a unique element of H; so is defined a unique mapping of $F \times G$ into H. This mapping depends on e_1; had some other element, e_2, been used another mapping of $F \times G$ into H would have been obtained. Thus, from μ there can be obtained a set of partial mappings, which can be represented individually as μ_{e_1}, μ_{e_2}, μ_f, μ_g, etc., of various factor-sets of $E \times F \times G$ into H. Sagittally, the new mapping uses arrows only from certain "planes" or "lines" in the original space.

Conversely, any *set* M of mappings (of P into Q say) can be regarded as (is in one-one correspondence with) a *single* mapping of $M \times P$ into Q.

C. *Projection*

Projection means "picking out one (or more) components". If $\langle x, y\rangle$, for instance, is the point (2,4) in the x,y-plane, then the projection of it on to the y-axis is the value 4. In set theory, given a product set $E \times F$, pr_1 is the operator that converts each couple $\langle e, f\rangle$ to the element e of E. pr_1 is thus a mapping of $E \times F$ into E. Similarly, pr_2 is the mapping of $E \times F$ into F that converts $\langle e, f\rangle$ to f. If the basic set is $E \times F \times G$, $pr_1(\langle e, f, g\rangle) = e$; and so on. Similarly, one can write $pr_{23}(\langle e, f, g\rangle) = \langle f, g\rangle$; and so on.

E and F being given, $pr_1^{-1}(e)$ consists of all those couples that have e as first component; clearly, it is the set $\{e\} \times F$.

An operator required later (*I12g*) is that which, operating on any n-tuple, generates the set of n-tuples that differs in all possible ways from the initial n-tuple in certain components while leaving the other components unchanged. If the factors of the product set are $(E_i)_{i \in I}$, so that I is the set of all factors, and if $J \subset I$, and if it is the components of the set J that are to be varied, then the operator we require is easily found to be $pr_J^{-1} \circ pr_{I-J}$. It is not, of course, a mapping but a binary relation (2.5 below). As it induces variation in the components of set J it will be represented as V_J.

(If $S \subset E \times F$):

3C.1 $e \in pr_1(S) \Leftrightarrow \exists f: f \in F$ and $\langle e, f\rangle \in S$.
3C.2 $f \in pr_2(S) \Leftrightarrow \exists e: e \in E$ and $\langle e, f\rangle \in S$.
3C.3 $pr_1(1_E) = E$; $pr_1(1_A) = A$.

As pr_i is a mapping, the formulae of *1f* apply; in particular

3C.4 $pr_i(S \cup T) = [pr_i(S)] \cup [pr_i(T)]$.
3C.5 $pr_i(S \cap T) \subset [pr_i(S)] \cap [pr_i(T)]$.
3C.6 $S \subset T \Rightarrow [pr_i(S)] \subset [pr_i(T)]$.

2.4 PROPERTIES AND RELATIONS

A. *Properties*

It is fundamental in Bourbaki's method that a property is identified with the subset of elements that possess the property (some total set or "universe" always being defined, or at least clearly understood). Thus, if the "universe" is the set of positive integers, the property of "being even" would be identified with the subset $\{2,4,6,\ldots\}$, and the property "$x \leqslant 5$" with the subset $\{1,2,3,4,5\}$. (Again, whether the sets are finite or infinite, discrete or continuous, etc., is irrelevant.) In this way properties can undergo the same operations as sets, without the least ambiguity. Thus the *union* of the two *properties* just mentioned is the property corresponding to the set

$$\{1, 2, 3, 4, 5, 6, 8, 10, 12, \ldots\};$$

although there happens to be no ready-made English adjective for it, it is perfectly well defined. Similarly, within the same total set, the negation or complement of the property "$x \leqslant 5$" is the property $\{6,7,8,9,\ldots\}$, which can be expressed as "$x > 5$" (with the defined universe understood). In general, $P[\![x]\!]$ will be used to represent some particular property that x may (or may not) possess.

B. *Relations*

In the same manner, a relation is identified with a subset of a product set, a suggestion originally due to Wiener [14]. Thus, the relation "x further north than y" is satisfied by the couple ⟨Edinburgh, London⟩, but not by ⟨London,

Edinburgh⟩ nor by ⟨Rome, London⟩. The set of couples (or more generally n-tuples) that satisfy the relation is now, as a set, subject to all the ordinary set operations. Thus, in the universe of men, the *intersection* of the two relations "x has the same father as y" and "x has the same mother as y" is the relation "x is full brother of y". In general $R[\![x,y,\ldots]\!]$ will be used to represent some particular relation that may (or may not) hold between x and y and ..., each from its own set.

C. *Reduction of order*

In $R[\![x,y]\!]$, the fixing of x at a single element (for whatever reason) makes $R[\![x,y]\!]$ a property of y. Thus if $R[\![x,y]\!]$ is "x is twice as big as y" and x is then fixed, at 10 say, the phrase "10 is twice as big as y" specifies a property of a single number, not of a couple $\langle x,y\rangle$, such that 5 has it but 6 has not. So R has decreased in order from binary to unary (equivalent to a property).

The expression $\forall x\colon R[\![x,y]\!]$ classifies the possible values of y according to whether each particular value makes the expression true or false. Thus it defines a property of y, not a relation between x and y. The quantifier $\forall$, operating on one of the variables in $R[\![x,y,\ldots]\!]$ thus lowers the order by one.

Similarly, so does $\exists$.

With these facts in mind, the following formulae are readily established; some obvious abbreviations are used to save space.

4C.1 If $A \subset E$: $\exists x \in A\colon R[\![x, \ldots]\!] \Rightarrow \exists x \in E\colon R[\![x,\ldots]\!]$

4C.2 If $A \subset E$: $\forall x \in E\colon R[\![x,\ldots]\!] \Rightarrow \forall x \in A\colon R[\![x,\ldots]\!]$

4C.3 $\overline{\exists x\colon R[\![x,\ldots]\!]} \Leftrightarrow \forall x\colon \overline{R}[\![x,\ldots]\!]$

4C.4 $\overline{\forall x\colon R[\![x,\ldots]\!]} \Leftrightarrow \exists x\colon \overline{R}[\![x,\ldots]\!]$

4C.5 $\exists x\colon (R \text{ and } S[\![x,\ldots]\!]) \Leftrightarrow R$ and $\exists x\colon S[\![x,\ldots]\!]$ provided x does not occur in R.

4C.6 (Similarly for $\forall$ and "or").

4C.7 $\exists x\colon (R \text{ or } S) \Leftrightarrow (\exists x\colon R)$ or $(\exists x\colon S)$.

4C.8 $\forall x\colon (R \text{ and } S) \Leftrightarrow (\forall x\colon R)$ and $(\forall x\colon S)$.

4C.9 $\exists x\colon (R \text{ and } S) \Rightarrow (\exists x\colon R)$ and $(\exists x\colon S)$.

4C.10 $(\forall x\colon R)$ or $(\forall x\colon S) \Rightarrow \forall x\colon (R \text{ or } S)$.

4C.11 $\exists\langle x,y\rangle\colon R \Leftrightarrow \exists x\colon (\exists y\colon R) \Leftrightarrow \exists y\colon (\exists x\colon R)$.

4C.12 $\forall\langle x,y\rangle\colon R \Leftrightarrow \forall x\colon (\forall y\colon R) \Leftrightarrow \forall y\colon (\forall x\colon R)$.

4C.13 $\exists x\colon (\forall y\colon R) \Rightarrow \forall y\colon (\exists x\colon R)$.

4C.14 $\exists x\colon \forall y\colon \forall z\colon R \Rightarrow \forall y\colon \exists x\colon \forall z\colon R \Rightarrow \forall y\colon \forall z\colon \exists x\colon R$.

4C.15 $\exists x\colon \exists y\colon \forall z\colon R \Rightarrow \exists x\colon \forall z\colon \exists y\colon R \Rightarrow \forall z\colon \exists x\colon \exists y\colon R$.

2.5 BINARY RELATIONS

As the binary relations are of special importance to us they will be given a more extended description.

A binary relation is any subset of the product of two sets. Conversely, any subset of a product set defines a binary relation. It can be both an active operator and a passive operand; so arises the possibility of operations *on* operations, with rich possibilities of dynamic structure.

A. *Section*

(This is Bourbaki's *coupe*; his *section* is not this.)

Given a product set and subset of it, $S \subset E \times F$ say, and an element x in E, the *section* of S corresponding to x, written $S(x)$, is the set of those elements in F that, with x, make a couple in S.

5A.1 $y \in S(x) \Leftrightarrow \langle x,y\rangle \in S$.

(Notice that on the left S is an operator, converting x to a set of F-elements; on the right S is simply a subset of $E \times F$.) The section corresponding to a subset A of E is defined as with a mapping (1F.1): if $A = \{a_1, a_2, a_3, \ldots\}$ then

5A.2 $S(A) = S(a_1) \cup S(a_2) \cup S(a_3) \cup \ldots$

5A.3 $y \in S(A) \Leftrightarrow \exists x\colon x \in A$ and $y \in S(x)$.

(If $S \subset E \times F$, $T \subset E \times F$, $A \subset E$, $B \subset E$):

5A.4 $S(A \cup B) = S(A) \cup S(B)$. (Riguet)

5A.5 $S(A \cap B) \subset S(A) \cap S(B)$. (Riguet)

5A.6 $(S \cup T)(A) = S(A) \cup T(A)$. (Riguet)

5A.7 $(S \cap T)(A) \subset S(A) \cap T(A)$. (Riguet)

5A.8 $A \subset B \Rightarrow S(A) \subset S(B)$. (Riguet)

5A.9 $S \subset T \Rightarrow S(A) \subset T(A)$. (Riguet)

B. *Inverse of a relation*

If $S \subset E \times F$, the subset of $F \times E$ defined by

5B.1 $\langle y,x\rangle \in S^{-1} \Leftrightarrow \langle x,y\rangle \in S$.

defines the binary relation S^{-1} between F and E. It has the usual properties of a binary relation (5A). In addition

5B.2 $x \in S^{-1}(y) \Leftrightarrow y \in S(x)$.

5B.3 $(S^{-1})^{-1} = S$.

5B.4 $S \subset T \Leftrightarrow S^{-1} \subset T^{-1}$.

5B.5 $\overline{S^{-1}} = (\overline{S})^{-1}$.

5B.6 $(A \times B)^{-1} = B \times A$.

C. *Composition*

If $S \subset E \times F$ and $T \subset F \times G$, so that S and T share the set F, the *composition* of S and T, written $T \circ S$ (in that order) is a new binary relation, a subset of $E \times G$, defined by

5C.1 $\langle x,z\rangle \in T \circ S \Leftrightarrow \exists y\colon y \in S(x)$ and $z \in T(y)$.

Thus the new relation, or set, consists of those elements in E and G that can find, through S and T, a common element in F.

$T \circ S$ may often conveniently be contracted to TS; and, if $R \subset E \times E$, RR may be written R^2. (If $E \neq F$, S^2 does not exist.)

(If $S \subset E \times F$ and $T \subset F \times G$ and $R \subset E \times E$):

5C.2 $\langle x, z \rangle \in TS \Leftrightarrow z \in (TS)(x)$. (by 5a:1)
5C.3 $\langle x, z \rangle \in TS \Leftrightarrow z \in T(S(x))$.
5C.4 $S \subset SS^{-1}S$ (true for every S).
5C.5 $T \circ (S_1 \cup S_2) = TS_1 \cup TS_2$. (Riguet)
5C.6 $T \circ (S_1 \cap S_2) \subset TS_1 \cap TS_2$. (Riguet) (The two sides become equal if T^{-1} is single-valued; see below.)
5C.7 $(T_1 \cup T_2) \circ S = T_1 S \cup T_2 S$. (Riguet)
5C.8 $(T_1 \cap T_2) \circ S \subset T_1 S \cap T_2 S$. (Riguet) (The two sides become equal if S is single-valued.)
5C.9 $S_1 \subset S_2 \Rightarrow TS_1 \subset TS_2$
5C.10 $S_1 \subset S_2 \Rightarrow S_1 T \subset S_2 T$.
5C.11 $1_B \circ S \circ 1_A = S \cap (A \times B)$ $(A \subset E, B \subset F)$. (Riguet)
5C.12 $(TS)^{-1} = (S^{-1}) \circ (T^{-1})$.
5C.13 $(B \times C) \circ S = [S^{-1}(B)] \times C$ $\left.\begin{matrix} A \subset E \\ B \subset F \\ C \subset G \end{matrix}\right\}$ (Riguet)
5C.14 $T \circ (A \times B) = A \times [T(B)]$

D. *Transitive closure*

A binary relation important in generalized dynamics is that obtained from a mapping by applying it repeatedly. More generally, if $R \subset E \times E$, so that R^2, R^3, etc. exist, the *transitive closure* of R, written R^T, is the binary relation, also a subset of $E \times E$, defined by

5D.1 $x \in R^T(e) \Leftrightarrow x \in \{R(e) \cup R^2(e) \cup R^3(e) \cup \ldots\}$.

If N is the set of integers, zero excluded,

5D.2 $x \in R^T(e) \Leftrightarrow \exists n \in N: x \in R^n(e)$.

R may be a mapping (6C); its transitive closure, however, is not usually a mapping.

2.6 SPECIAL FORMS OF BINARY RELATIONS

A. *Single-valued*

S is single-valued if, for all e, $S(e)$ has never more than one element. Tabularly, no column may have more than one mark; sagittally, no element may emit more than one arrow. The algebraic condition may be found by this method, due to Riguet:

S is single-valued
$\Leftrightarrow$ If $S(e)$ has apparently two elements, x and y, they must be really the same element
$\Leftrightarrow$ If $[\exists e: x \in S(e)$ and $y \in S(e)]$ then $[x = y]$
$\Leftrightarrow$ If $[\exists e: x \in S(e)$ and $e \in S^{-1}(y)]$ then $[x = y]$
$\Leftrightarrow$ If $[x \in SS^{-1}(y)]$ then $[x = 1_F(y)]$
$\Leftrightarrow$ If $[\langle y, x \rangle \in SS^{-1}]$ then $[\langle y, x \rangle \in 1_F]$
thus:

6A.1 S is single-valued $\Leftrightarrow SS^{-1} \subset 1_F$.

B. *Everywhere defined*

S is everywhere defined if, for all e, $S(e)$ has at least one element. Tabularly, no column may have no mark; sagittally, every element must emit at least one arrow. The algebraic condition may be found thus (Riguet):

S is everywhere defined
$\Leftrightarrow \forall e$: $S(e)$ contains at least one element
$\Leftrightarrow \forall e: [\exists y: y \in S(e)]$
$\Leftrightarrow \forall e: [\exists y: y \in S(e)$ and $e \in S^{-1}(y)]$
$\Leftrightarrow \forall e: [\exists y: e \in S^{-1}(y)$ and $y \in S(e)]$
$\Leftrightarrow \forall e: e \in S^{-1}S(e)$
$\Leftrightarrow \forall e: \langle e, e \rangle \in S^{-1}S$
thus:

6B.1 S is everywhere defined $\Leftrightarrow 1_E \subset S^{-1}S$.

C. *Mappings*

A mapping can now be defined simply as a binary relation that is both single-valued and everywhere defined. It is essentially identical with the well known "function". It is an operator (of course), but now, as a mere set it can be operated on; we thus now have a calculus in which operators can be operated on, properties can *have* properties (by being joined into sets), relations can be related, and so on.

D. *Reflexive*

A binary relation is reflexive if every element e has this relation to itself: i.e. $\forall e: \langle e, e \rangle \in R$, or (Riguet)

6D.1 R is reflexive $\Leftrightarrow 1_E \subset R$.

The property is, of course, not possible if $S \subset E \times F$ and $E \neq F$.

E. *Transitive*

A binary relation is transitive if, whenever two couples share a common element, in the appropriate order: $\langle x, y \rangle \in R$ and $\langle y, z \rangle \in R$, then $\langle x, z \rangle \in R$.

In this case, for all x, z:

$$[\exists y: \langle x,y\rangle \in R \text{ and } \langle y,z\rangle \in R] \Rightarrow [\langle x,z\rangle \in R];$$
$$[\exists y: y \in R(x) \text{ and } z \in R(y)] \Rightarrow [\langle x,z\rangle \in R];$$
$$\langle x,z\rangle \in RR \Rightarrow \langle x,z\rangle \in R.$$

6E.1 R is transitive $\Leftrightarrow R^2 \subset R$. (Riguet)

F. *Symmetric*

A binary relation is symmetric if, whenever x has the relation R to y, y has it to x; i.e. $\forall\langle x,y\rangle: \langle x,y\rangle \in R \Leftrightarrow \langle y,x\rangle \in R$; or: $\forall\langle x,y\rangle: [\langle x,y\rangle \in R \Leftrightarrow \langle x,y\rangle \in R^{-1}]$.

6F.1 R is symmetric $\Leftrightarrow R = R^{-1}$. (Riguet)

G. *Equivalence*

A binary relation is an equivalence relation if and only if it is reflexive and transitive and symmetric.

The classes into which an equivalence relation divides the elements are the *elements* of the "quotient" set.

Every equivalence relation is of the form $\rho^{-1}\rho$, where ρ is the mapping of the elements of the basic set into the quotient set. (ρ says, of each element, which class it belongs to).

H. *Anti-symmetric*

A binary relation is anti-symmetric if, for all $\langle x,y\rangle$,

$$[\langle x,y\rangle \in R \text{ and } \langle y,x\rangle \in R] \Rightarrow [x = y];$$

i.e.
$$[\langle x,y\rangle \in R \cap R^{-1}] \Rightarrow [x = y]:$$

6H.1 R is anti-symmetric $\Leftrightarrow R \cap R^{-1} \subset 1_E$. (Riguet)

I. *Order*

A binary relation is one of order if and only if it is reflexive and transitive and anti-symmetric.

J. *Rectangular*

A binary relation is rectangular if it can be expressed as a product set. For this to be so, $\langle e_1,f_1\rangle$ and $\langle e_1,f_2\rangle$ and $\langle e_2,f_1\rangle$ in S must imply $\langle e_2,f_2\rangle$ in S. Forming the composition as before, this gives:

6J.1 S is rectangular $\Leftrightarrow SS^{-1}S = \{\}$. (Riguet)

K. *Difunctional*

A binary relation S is difunctional (Riguet) if and only if, for every pair e_1 and e_2 in E, $S(e_1)$ and $S(e_2)$ are either identical or have no intersection, i.e. for all e_1 and e_2 in E, and f_1 and f_2 in F:

$$[f_1 \in S(e_1) \text{ and } f_1 \in S(e_2) \text{ and } f_2 \in S(e_1)] \Rightarrow [f_2 \in S(e_2)]$$

i.e.
$$f_2 \in SS^{-1}S(e_2) \Rightarrow f_2 \in S(e_2);$$

6K.1 S is difunctional $\Leftrightarrow SS^{-1}S \subset S$.

6K.2 S is difunctional $\Leftrightarrow SS^{-1}S = S$ (using 5C.4).

L. *Cyclic*

A binary relation R is cyclic if, for every couple $\langle x,y\rangle$ such that $y \in R(x)$, it is also true that $x \in R^T(y)$; i.e. $\langle x,y\rangle \in R \Rightarrow \langle x,y\rangle \in (R^T)^{-1}$.

6L.1 R is cyclic $\Leftrightarrow R \subset (R^T)^{-1}$.

The *cyclic content*, $\overset{\circ}{R}$, of a binary relation R is defined by

6L.2 $$\overset{\circ}{R} = \bigcup_{\substack{P \subset R \\ P \text{ is cyclic}}} P.$$

The cyclic content of a mapping is the set of states in its basins; it is therefore the set of states to which every state in the domain will be converted under incessant repetition of the mapping.

M. *Re-arrangement*

Cancellation, and the re-arrangement of equations, cannot be performed by a mere copying of the rules of ordinary algebra, but must be based on first principles. Some examples of possible methods are given below. Notice that the calculus is now sufficiently developed to allow operations directly on the sets and relations themselves, the elements being out of sight.

Example 1. $AB \subset C$, and A is everywhere defined: what can be said of B?

By 6B.1, $1 \subset A^{-1}A$, so by 5C.10 and 5C.11, $B \subset A^{-1}AB$; and from $AB \subset C$ and 5C.9, $A^{-1}AB \subset A^{-1}C$; $\therefore B \subset A^{-1}C$.

Example 2. $SR \supset Q$, and S^{-1} is single-valued; what can be said of R?

By 6A.1 and 5B.3, $S^{-1}S \subset 1$, and so $S^{-1}SR \subset R$; and as $Q \subset SR$, $S^{-1}Q \subset S^{-1}SR$; $\therefore R \supset S^{-1}Q$.

Example 3. $T \subset SRR^{-1}$ and R is single-valued; can R be eliminated? $RR^{-1} \subset 1$, so $SRR^{-1} \subset S$; but $T \subset SRR^{-1}$; $\therefore T \subset S$.

2.7 TERNARY AND HIGHER RELATIONS

Examples are given by such statements as:

1. Ship x is at longitude y and latitude z.
2. Mr. x purchased object y for z dollars.

3. x is between y and z.
4. $2x + y - z = 1$
5. p is to q as r is to s.

The matter is somewhat simplified by the fact that many ternary relations are more naturally treated as a binary relation between a variable and a couple; thus the first example is naturally equivalent to the binary relation—Ship x is at position p—with the subsidiary fact that p, on this Earth, must be the couple $\langle y, z \rangle$.

Extension of the earlier method to ternary and higher relations is at most points obvious. An n-ary relation is a subset of a product of n sets. The quantifier $\exists$, or $\forall$, lowers the order by one.

Care, however, is needed in the dimensions. The "inverse" of a ternary or higher relation is no longer unique (for while the change from $\langle x, y \rangle$ to $\langle y, x \rangle$ is a unique permutation, the permutations of $\langle x, y, z \rangle$ are more than one). Composition, too, must be specially defined to show what components are eliminated; thus, while subsets of $E \times F \times G \times H$ and subsets of $G \times H \times J$ might eliminate the common $G \times H$, they might eliminate only G, or only H. When the component sets are all the same, for example $E \times E \times E \times E$ and $E \times E \times E$, the elimination might be done in many ways, so the way selected must be specially defined. The selection must, of course, be guided by the primary aim of the work.

3. APPLICATIONS TO HOMEOSTASIS

The account given previously is purely mathematical, for it makes no appeal to any source of structure or of justification other than to its own axioms. Only in this way can we be sure that we are not unconsciously appealing to parallel known facts in the real world; only so can we be sure that the structure has a strength of its own. We shall now consider how this structure is related to certain structures already well known in the worlds of biology, physics, and control mechanisms.

As a first step it will be convenient to consider Sommerhoff's concept of "directive correlation"[15]. It shows well the peculiar advantages of the method of set theory, and will form a natural transition to the cases in which the dynamic, or time, aspect is outstanding.

3.1. DIRECTIVE CORRELATION

In 1950, Sommerhoff gave a rigorous and operational definition that attempted to catch the essence of what is meant in biology and psychology by coordination, integration, purposeful action, adapting the means to the end. His attempt was in the spirit of the mathematicians who, a century ago, attempted to give rigorous forms to such common ideas as continuity, torsion,

convergence. Today no one doubts the value of their achievements, for vague common-sense and personal intuition were replaced by a disciplined rigor. Sommerhoff succeeded in his attempt, and I have no doubt that his precise definition will come to be recognized as fundamental in scientific biology and psychology.

In 1950, however, the methods of set theory in Bourbaki's form were little known, and Sommerhoff used the analytic language that had become classic in physics and chemistry. Here I hope to show that his basic idea can be stated very much more simply, and therefore perhaps more clearly, in the concepts of set theory. (It should again be borne in mind that the sets used below may be either finite or infinite, with elements differing finitely or infinitesimally, with

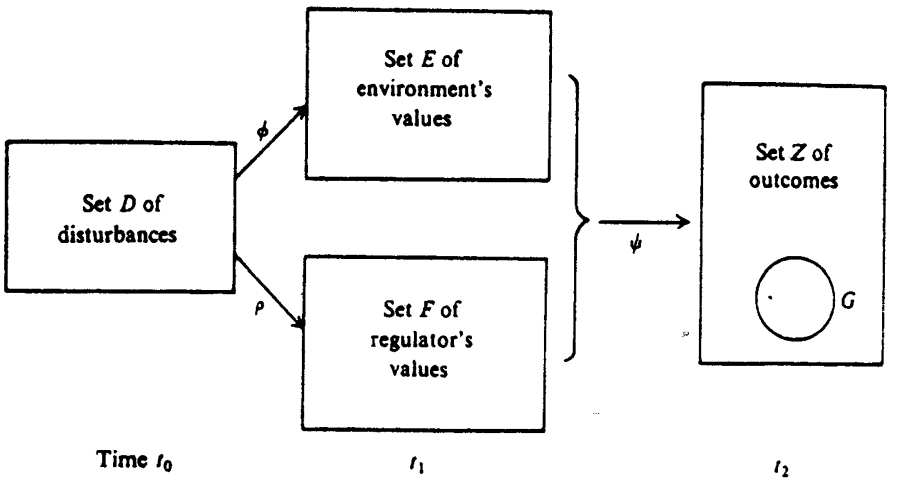


FIG. 1.

or without a metric, as the reader wishes; the formulation is the same for all cases.) The formulation is as follows.

There is a set D of disturbances d; this is the set of values of the coenetic variables. They cause, in the environment, values e of the environment's set E of possible values. As the environment always does something, even if only to make a change of zero degree, the effect of D on E is everywhere defined; and as the environment cannot do two things at once, the effect is single-valued; thus the D–E relation defines a mapping, ϕ say, of D into E. It is assumed that the disturbance d occurs at time t_0; ϕ produces e, where $e = \phi(d)$, at time t_1.

The organism, brain, regulator, or whatever it is that claims to show directive correlation, is similarly specified by a set F, of elements f; and similarly its behavior, its response to d, specifies a mapping ρ of D into F. For directive correlation to be shown, the mapping ρ ("how the brain reacts"), must bear some special relation to ϕ. Our question is: what relation?

When disturbance d has evoked e and f—responses $\phi(d)$ and $\rho(d)$ respectively—these two values interact to give some final outcome, at t_2. Again, because the outcome must be everywhere defined and single-valued, the interaction to an outcome must correspond to a mapping, ψ say, of $E \times F$ into Z, where Z is the set of possible outcomes when E and F range uncorrelatedly over all their values. Z in fact must be $\psi(E \times F)$. Within Z is the subset, call it G, of outcomes that are "good", that satisfy the focal condition. The relations may be clarified by Fig. 1.

"Directive correlation" is now defined as being shown by ρ in respect of D, ϕ, ψ, and G if and only if:

1.1 $\forall d \in D: \psi(\langle \phi(d), \rho(d) \rangle) \in G.$

(Some further restrictions could be added if one wished to exclude degenerate cases, such as when D has only one element (or is even empty!), or if $G = Z$, or $G = \{\}$; but it seems simpler to leave them in and merely to notice their degeneracy should they occur. Directive correlation would be said to be present, but to zero degree. This condition, it should be noticed, is fundamentally different from cases in which the primary conditions are not met, or are left undefined.)

The expression above can be simplified algebraically by noticing that the set specified by

$$\forall d \in D: \langle \phi(d), \rho(d) \rangle$$

is identical with the set $\rho \circ \phi^{-1}$, with D as the set eliminated by the composition. So the criterion becomes

1.2 $\psi(\rho \circ \phi^{-1}) \subset G.$

As ψ and ϕ are mappings, this expression may be rearranged (as in 2.6M) to give the final *formulation*:

1.3 $\left.\begin{array}{l}\rho \text{ shows directive correlation} \\ \text{in respect of } D, \phi, \psi, \text{ and } G\end{array}\right\} \Leftrightarrow \rho \subset [\psi^{-1}(G)] \circ \phi$

The expression, one should notice, is wholly operational, referring everywhere to what the parts *do*. It also shows exactly what components any discussion of directive correlation must be based on: omit any and the discussion becomes meaningless. (These features were also present, of course, in Sommerhoff's original formulation.) (At the present degree of generality, no distinction is made between major disturbances that threaten and minor disturbances that are used as signals: d may represent both.)

3.2 MACHINES

While Sommerhoff's concept is basically dynamic, for it treats of three different events that must occur in time in a certain order, it uses only three

We will now consider systems that progress through an indefinitely large number of successive times.

Various definitions and methods are possible; here I shall outline one that has been well tested and that has shown, over twenty years, its peculiar appropriateness for questions relating to homeostasis.

Of central importance in science is the system whose present state, if known completely, determines its next state. Laplace took for granted that the universe is of this type; Temple [16] refers to "the fundamental assumption of macrophysics that a complete knowledge of the present state of a system furnishes sufficient data to determinate definitely its state at any future time or its response to any external influence".

A. *States*

Formally, he who would define a particular machine (such as a typewriter, the solar system, the mosquito) must start by specifying a set of states. To define this set is nothing other than to make unambiguous what is being talked about. "The mosquito's susceptibility to DDT" obviously refers to the adult, but does the reference include the larva, and the egg? "Mosquito" includes many forms—old and young, male and female, hungry and fed, flying and resting, healthy and malarial—so before we can proceed we must indicate, with sufficient precision, the extent of the mosquito-states under consideration. Here "mosquito" is the set, its various states are the elements.

These states provide the basic set on which all Bourbaki's concepts rest. Not every real thing that is nameable qualifies; for set theory the elements must have both individuality and permanence of individuality. The rain-drops running down a window-pane, for instance, cannot be used; for, as they fuse and break, their individualities are lost, and such operations as union become undefinable. It is assumed here that any states used to describe a system are such as allow the set operations to be performed unambiguously. The states may be defined quantitatively, as astronomy gives the state of a planetary system by numerical positions and momenta, or by arbitrary names, as the meteorologist identifies the type of cloud in the sky.

B. *Mapping*

Given a sufficiently defined set M of states m, the set shows *state-determined* behavior if and only if its succeeding state m' is a single-valued and everywhere defined function of its present state: $m' = \mu(m)$. Such behavior corresponds to a mapping of M into M, with the form μ corresponding to, or being due to, whatever natural forces are operating in real time to cause the change; μ represents the laws of nature so far as they are showing in the set M.

Such a mapping represents an *isolated* system. The repetition of μ generates a sequence of states: $\mu(m)$, $\mu^2(m)$, $\mu^3(m)$, ..., a line of behavior or trajectory; such might be shown by an ant colony, when given a piece of meat and then

observed over 24 hours. "Isolation" has thus nothing to do with "closed to energy and matter"; a system isolated in our present sense may be wide open thermodynamically. (The question is taken up more rigorously in Section 2I.)

C. *Structure*

It is with repetition that the characteristic "structure of machine" appears, where "structure" is used in Bourbaki's sense to refer to a characteristic form or pattern. The "structure" of a group or algebra has long been well known in mathematics; Bourbaki has identified the more general concept, and has shown [3] that it depends always on *restriction*, commonly that given by the axioms (of the group, etc.). The "structure of machine" appears because, as the sequence of states $m, m', m'', m''', \ldots$ appears, related by $m' = \mu_1(m), m'' = \mu_2(m')$, etc., all the μ_i's are the same: $\mu_1 = \mu_2 = \mu_3 = \ldots$ etc., and it is this *redundancy* (by repetition) that gives the structure. In other words, the structure of "machine" appears when the laws that govern the system are *invariant* in time.

D. *Succession*

A deeper insight into the meaning of "machine" can be obtained by these methods. Any empirical study of a system gives, as basis, a record of what happened at what times. It thus specifies a *mapping* from a domain T of time-values t into the set M of possible states. Call this observed mapping λ. If m is an observed state, $\lambda^{-1}(m)$ is the set of times at which this state was observed to occur. Now let σ be the mapping of T into T (with some qualification about the ends of the domain) that converts t to $t + \Delta t$, i.e. that moves t along by one unit of time. $\sigma\lambda^{-1}(m)$ is then the set of times one unit later than the times just mentioned; and $\lambda\sigma\lambda^{-1}(m)$ is the set of states that followed m. Thus, for the system to be state-determined it is necessary and sufficient that

2D.1 $\forall m \in M$: $\lambda\sigma\lambda^{-1}(m)$ is single-valued.

As any real system must be everywhere defined, we have:

2D.2 The record is that of a machine $\Leftrightarrow$ $\lambda\sigma\lambda^{-1}$ is a mapping.

Now the form ABA^{-1} is well known in many branches of mathematics. It can always be interpreted as what B looks like when seen through, or coded by, some operation A. Thus, the criterion just given (2D.2) shows that the core of the concept of "machine" is that the system should show *a coded version of simple succession.*

E. *Machine with input*

Sometimes the mapping μ is not constant, but the variations are at least constrained by being always from a well defined set $\{\mu_1, \mu_2, \mu_3, \ldots\}$. By Section 2.3B, this set of mappings (with subscripts taken from some set I, not necessarily numerical) corresponds to one mapping of $I \times M$ into M. This is the

basic form of the "machine with input". If the value of I stays constant at i, the system becomes an isolated one with mapping μ_i. If I is made to vary in time, by some external source of variation, the system becomes identical with Shannon's "noiseless transducer". Since the mapping of $I \times M$ into M specifies, for every input trajectory, the resulting output trajectory, it corresponds to, and generalizes, the well known "transfer function". The mapping is also identical with that of an "algebra with external operators" (Bourbaki [5]). There can therefore be no doubt of the very great range and applicability of this basic concept.

F. *Joining and analysis*

Joining two (or more) machines corresponds to adding to the descriptions of two (or more) machines a new function (mapping) which specifies one's input values (no longer allowed to vary arbitrarily) as a function of the other's state. "Feedback" occurs if two machines are joined reciprocally. It is easily shown that (as is obviously necessary) the new system still accords with the definition of a machine.

The converse process, of analyzing a machine with input into parts, whose joining gave the whole machine, can *always* be done algebraically (for the kinematic graph has only to be arranged in a suitable product space), but not all such algebraic possibilities correspond to cases of real interest. Given, however, that the analysis of the whole into parts is wanted, the whole states will be specified as n-tuples and the transformation-mapping becomes a simultaneous one on n sets, perhaps on n numerical variables. Each partial mapping (Section 2.3B) then gives the canonical representation of each part.

G. *Diagram of immediate effects*

When the whole is to be considered as made of parts, a very common and important question is: to what extent are the parts independent (what acts on what)? Answering this question implies construction of the diagram of immediate effects. The specification in algebraic set theory is due to Riguet [9].

Consider a machine with parts $X_1, X_2, X_3, \ldots, X_k, \ldots$ (where each part k is identified by a set X_k of elements x_k, so that the state of the whole machine is specified by the n-tuple $\langle x_1, x_2, \ldots, x_k, \ldots\rangle$. As machine, let there be defined a mapping μ of $\Pi_k X_k$ into itself. Let the set of parts be I, so that (in this notation) $I = \{1, 2, \ldots, k, \ldots\}$ (where the numbers are mere labels for the parts).

Suppose that variable (or part) i has no immediate effect on variable j. This means that if we take a generic state x and observe the various transitions that occur from it and from all the states that differ from it only in component i, then we shall find, in the transforms, that the values at component j are all the same, in spite of the variations at i. More formally, use the operator V of Section 2.3C to express these sets and relations. The transition from x is to $\mu(x)$. The states that differ from x only in the i-component are the set $V_i(x)$.

The transitions from these states will be to $\mu(V_i(x))$. The set restricted only by being equal to $\mu(x)$ in its j-component is $V_{I-j}(\mu(x))$. The previous set, to have the same j-components, must occur *within* this set. So the property "i has no immediate effect on j" is equivalent to

2G.1 $\forall x: \mu V_i(x) \subset V_{I-j}\mu(x)$.

Or, after re-arrangement as in Section 1.6M,

2G.2 Variable i has no immediate effect on variable $j \Leftrightarrow \mu V_i \mu^{-1} \subset V_{I-j}$.

Again the form is suggestive, for it says that when there is no immediate effect from i to j, variation at i, as seen through or transmitted by the machine μ, remains within the class that has the property or corresponds to "no variation at j".

H. *Diagram of ultimate effects*

The expression 2G.2 above (when μ and I are given) is true of certain $\langle i,j\rangle$-couples. It thus defines (by being $\bar{Q}$) a binary relation Q in $I \times I$. Clearly, the diagram of ultimate effects is simply the transitive closure, Q^T, of Q.

I. *Isolation*

The set formulation of an "isolated" system can now be stated rigorously. As before, let I be a set of variables or parts; and let $Q(\subset I \times I)$ be its diagram of immediate effects. Suppose, within this total system, that the subsystem composed of parts $J(J \subset I)$ is "isolated" from the other parts (from I–J).

Events in I–J have no effect on those in J; this means that in the diagram of immediate effects no arrow goes from any point in I–J to any point in J. Thus Q, acting on the set I–J, gives only points in I–J; or, equivalently, Q^{-1}, acting on J gives only points in J. Thus,

2I.1 The set J of variables is isolated $\Leftrightarrow Q^{-1}(J) \subset J$.

The proposition may be made intuitively more evident by noticing that as Q, acting on some variables, gives the set that is disturbed by their activities, so Q^{-1}, acting on a set, gives the set that does disturb the given set. $Q^{-1}(J) \subset J$ says that the disturbers of J are to be found only in set J; i.e. J is not subject to outside disturbance; J is isolated.

J. *Simplification*

As cybernetics progresses to the treatment of more and more complex systems, so will the methods of simplification have to become more powerful and sophisticated. The foundations of method have already been made clear by Bourbaki; here the foundations will be given as they apply to the theory of machines. I assume here that every simplification is achieved by the application

of an equivalence relation; I am not aware that the matter has been discussed exhaustively but I know of no reason for rejecting this *axiom*.

If a machine $\mu(\subset M \times M)$ is advantageously to be simplified by the application of an equivalence relation, the quotient set must *still be a machine* (if the work is to develop further in the same region of discourse). What are the conditions that this will be so?

Take an equivalence relation $R(\subset M \times M)$ and a generic state m of the machine. The states grouped with it by R is the set $R(m)$. The transforms, by μ, of all these states must lie in the same equivalence class (or the quotient machine will not be single-valued). m goes to $\mu(m)$, and the equivalence class of the transform is $R(\mu(m))$; as just said, all of $\mu(R(m))$ must lie in it. The condition for compatibility (that the merging does not destroy the structure of machine) is thus

2J.1 $\forall m \in M: \mu R(m) \subset R\mu(m)$.

After re-arrangement this becomes

2J.2 Machine μ is compatible with equivalence relation R $\Leftrightarrow \mu R \mu^{-1} \subset R$.

Again the form can be interpreted: the equivalence relation, as seen through or coded by, the machine, must not have its classes broken.

When R is expressed as $\rho^{-1}\rho$ (cf. Section 2.6G), the new (simplified) machine's mapping σ is given at once by

2J.3 $\sigma = \rho\mu\rho^{-1}$.

Again the interpretation is clear: the new mapping is simply the old one, seen through the simplifying mapping ρ.

The "difunctional" relation is important here. Any relation that satisfies $RR^{-1}R = R$ (Section 2, 5C.4 and 6K.1), if not a mapping, can be made one by the application of a suitable equivalence relation to its range. It thus provides an immediate indication that a simplification is possible.

K. *Markovian machines*

It may conveniently be noticed here that all these results can be generalized to the case in which the transitions are not determinate but have well defined probabilities. The machine, specified above by a mapping, then becomes specified by a matrix of transition probabilities, and the trajectory becomes a Markov chain. The basins cease, in general, to be absolute retainers but become places in which the machine spends an unusually large fraction of its time. Most properties that we have discussed above go over essentially unchanged, being merely distributed to some degree. From this point of view the Markovian machine could be the fully general form, the determinate machine being the special case in which all probabilities are 0 or 1.

3.3 EQUILIBRIA

The study of equilibria will always be important in the treatment of systems of high complexity, for the equilibria, in their various forms, are those states, or sets of states, in which the system's behavior no longer depends to a major degree on the time. By effectively losing a variable, the functional relation becomes simpler; and the change may reduce the impossibly complex to the manageable.

A. *Stable set, state*

The state that is in equilibrium under a mapping μ is abstractly identical with an element invariant under an operator, for both satisfy $\mu(x) = x$. A natural extension is the stable set of states, satisfying $\mu(A) \subset A$. These expressions mean, it should be noticed, that μ has, in a sense, lost its change-making power when reduced to A or x as a domain.

Obvious corollaries are that $\mu^T(x) = x$, and $\mu^T(A) \subset A$.

Often, when $\mu(x) = x$, it is of interest to know what will happen if the state operated on is not x but some state x^* near to it (obviously a topology over M must previously have been defined). If μ, applied repeatedly to x^*, brings it back to x, so that $\lim \mu^n(x^*) = x$, the machine μ is said to be "stable" to displacements from x; otherwise it is "unstable". When stable, the whole operation may be thought of as compound, with $\mu = \lambda^n \delta$, in which δ is an impulsive displacement-operator, such that $\mu(x) = x$. The displacements consequent on δ, which may be multiple-valued, are then ready to be equated to the set D of disturbances in Sommerhoff's formulation. In this way directive correlation may rigorously be demonstrated at any stable equilibrium.

B. *Trapping*

If the set A is stable under μ, so that $\mu(A) \subset A$, and the trajectory, under repeated action of μ, enters the set A then the trajectory can never leave it. Stable sets thus act as traps. Should A contain a subset B which is stable, then if the trajectory, confined to A, enters B it will remain trapped in the even smaller subset. So trajectories tend to get trapped in smaller and smaller sets.

Again, as it is easily shown that

3B.1 $\mu(C) \subset C$ and $\mu(D) \subset D \Rightarrow \mu(C \cap D) \subset C \cap D$

it follows that if any two sets C and D are trapping sets, so is their intersection. Thus if there are many stable, or trapping, sets occurring in complex overlapping patterns, all the intersections will be trapping, and the system will tend to be caught in some very small set that is the intersection of a number of primary sets.

C. *Lingering and convergence*

If the system is Markovian with fractional probabilities it may not be trapped absolutely in any subset, but there will usually be subsets within which it will spend a disproportionate amount of time. Some lingering in preferred subsets will, in fact, occur in all cases except when both rows and columns of the transition matrix add to 1 (the "doubly stochastic"). If the system is determinate and continuous (specified, say, by $\dot{x} = \phi(x)$, with x a vector), preferred regions will also always occur, the exceptions being the regions in which the ϕ_i's have the special relation

3C.1 $$\frac{\partial \phi_1}{\partial x_1} + \frac{\partial \phi_2}{\partial x_2} + \ldots + \frac{\partial \phi_n}{\partial x_n} = 0,$$

sometimes written as $\text{div}\,\phi = 0$. (The system will show a preference for those regions where div is negative, for there the phase-volume shrinks.) The fact that homogeneity of distribution occurs only in the special cases when all the rows and columns of the matrix add to 1, and when the divergence is exactly 0, shows that we may usually expect the distribution to be non-homogeneous, with preferred regions existing at which the system either sticks indefinitely or remains for an undue length of time.

D. *Selection*

As a mapping reduces its original domain to a subset, it performs the physical act of selection (though there may not exist any ready-made English noun to describe the subset). It follows, from Section 2.4, that any machine, in getting trapped in a stable set, *automatically generates properties and relations.*

Formula 5A.8, Section 2, now shows a new significance in an extremely broad range of application. "$A \subset M \Rightarrow S(A) \subset S(M)$" says that if a set M is cut down to a subset A, then every related set (related through S or any other) is also cut down. In other words, selection, at M, for some property will cause the emergence of some property *in every set related to M*. Thus the movement of a machine to an equilibrium may cause the emergence of all sorts of properties by no means having an obvious relation to the equilibrium.

If the phenomenon is to show strikingly the selection must be intense, and for it to show to a major degree the system must be large to the second order of largeness; for it must be so large that even after intense selection has shrunk the initially possible set of states to a small fraction, yet this small fraction must still be large enough to contain noticeable features, or to show more than trivial behavior. The general theory of equilibria has suffered much from the fact that most of the easily understood examples of equilibrium occur in systems so small that the equilibrium itself leaves no room for interesting events: after a watch has run down, for instance, or a hot body come to uniformity of temperature, little more can happen. A hive of bees, however,

with rich pastures around, can show an equilibrium in that year after year the state "living hive" transforms to "living hive"; within this equilibrium, however, there is room for a great deal of interesting activity.

Here these ideas join those of Sommerhoff, making possible a rigorous formulation of the general idea that at and around a state of equilibrium some coordination is essential. The correspondence can be traced in detail. The set Z of possible outcomes is the set M of all the system's states. The set G (of the goal or focal condition) is the set of stable states. R and E are the two parts obtained when the observer divides the whole system conceptually into "organism" and "environment". The set D of disturbances is here the set of possible initial states (essentially the same as Z). ϕ, ρ and ψ are determined by the various dynamic forces (of whatever nature) that make the system change with time. The system shows its coordination by showing that ρ is so matched to ϕ and ψ that goal G is arrived at, even after displacements from it.

3.4 HOMEOSTASIS

A. *Natural selection*

Within this framework of ideas homeostasis finds its exact representation. Natural selection, now the operator μ, has acted on a system that is actually large to the third order; for one order is used in the elimination of the vast number of subsystems that are dynamically unsuitable—for instance the interiors of earth and sun, the frozen regions, the regions that have only inert elements, for instance. The remainder, a small fraction of the totality, is still large enough to develop intensively selected equilibria; and then the equilibria themselves are sufficiently large to show a rich internal structure. When these localized equilibria are split, conceptually, by an observer into "organism" and "environment", he finds that the selection that has led to the equilibrium now shows as a special relation between the parts, that of 3.1.

B. *Ultrastability*

A host of special cases occur within this general formulation, as the sets D, E, F (of 3.1) are given different physical realizations. If they are product sets. for instance, the entities will be seen by the observer as built of *parts*, and he will often be interested in how the mapping ρ (of 3.1) is built out of partial mappings (2.3B), i.e. with how the parts are coordinated to make a whole. Much of the study of the brain and its functions is directed to just this question.

Among the special cases is that in which the disturbances D fall into two distinct classes—many small impulsive and a few large step-function form. In this case the relation between the components of ρ implied by the stability of the whole is that described, somewhat fully, as "ultrastability"[17].

A host of other interesting special cases could be described. Some of them are already developed in linear servo-theory and other branches of regulation-

theory. Much remains to be done, especially when the systems contain Markovian, or stochastic, components.

This account has been written in the hope that those who are working in the *general* theory of systems may find in algebraic set theory a tool that is general enough to be unrestrictive to the biologist while rigorous enough to satisfy the mathematician. It has also been written to serve as basic reference for further applications now being made.

REFERENCES

1. ASHBY, W. Ross (1940). Adaptiveness and equilibrium. *J. ment. Sci.* **86**, 478–483.
2. ASHBY, W. Ross (1962). "An Introduction to Cybernetics". Chapman and Hall, London. 4th impress.
3. BOURBAKI, N. (1958). "Eléments de Mathématique. Théorie des ensembles; *fascicule de résultats*" ASEI 1141 Hermann & Cie, Paris, 3me. Ed.
4. BOURBAKI, N. (1954; 1956; 1957). "Théorie des ensembles". ASEI 1212, 1243, 1258. Hermann & Cie, Paris.
5. BOURBAKI, N. (1951). "Algèbre". ASEI 1144. Hermann & Cie, Paris.
6. BOURBAKI, N. (1951). "Topologie générale". ASEI 1142. Hermann & Cie, Paris.
7. RIGUET, J. (1948). Relations binaires, fermetures, correspondences de Galois. *Bull. Soc. math. Fr.* **76**, 114–155.
8. RIGUET, J. (1951). "Fondements de la Théorie des Relations Binaires". Thèse, Paris.
9. RIGUET, J. (1953). Sur les rapports entre les concepts de machine de multipole et de structure algébrique. *C.r. hebd. Séanc. Acad. Sci., Paris*, **237**, 425–427.
10. RIGUET, J. (1953). Systèmes de coordonnées relationnels. *C.r. hebd. Séanc. Acad. Sci., Paris*, **236**, 2369–2371.
11. SHANNON, C. E. and WEAVER, W. (1949). "The Mathematical Theory of Communication". University of Illinois Press, Urbana.
12. MCGILL, W. J. (1954). Multivariate information transmission. *Psychometrika*, **19**, 97–116.
13. GARNER, W. R. and MCGILL, W. J. (1956). The relation between information and variance analysis. *Psychometrika* **21**, 219–228.
14. WIENER, N. (1914). A simplication of the logic of relations. *Proc. Camb. phil. Soc.* **17**, 387–390.
15. SOMMERHOFF, G. (1950). "Analytical Biology". Oxford University Press, London.
16. TEMPLE, G. (1942). "General Principles of Quantum Theory". Methuen and Co., London. 2nd Ed.
17. ASHBY, W. Ross (1960). "Design for a Brain". Chapman and Hall, London. 2nd Ed.

W. ROSS ASHBY
University of Illinois

PRINCIPLES OF THE SELF-ORGANIZING SYSTEM*

Questions of principle are sometimes regarded as too unpractical to be important, but I suggest that that is certainly not the case in *our* subject. The range of phenomena that we have to deal with is so broad that, were it to be dealt with wholly at the technological or practical level, we would be defeated by the sheer quantity and complexity of it. The total range can be handled only piecemeal; among the pieces are those homomorphisms of the complex whole that we call "abstract theory" or "general principles". They alone give the bird's-eye view that enables us to move about in this vast field without losing our bearings. I propose, then, to attempt such a bird's-eye survey.

WHAT IS "ORGANIZATION"?

At the heart of our work lies the fundamental concept of "organization". What do we mean by it? As it is used in biology it is a somewhat complex concept, built up from several more primitive concepts. Because of this richness it is not readily defined, and it is interesting to notice that while March and Simon (1958) use the word "Organizations" as title for their book, they do not give a formal definition. Here I think they are right, for the word covers a multiplicity of meanings. I think that in future we shall hear the *word* less frequently, though the *operations* to which it corresponds, in the world of computers and brain-like mechanisms, will become of increasing daily importance.

The hard core of the concept is, in my opinion, that of "conditionality". As soon as the relation between two entities *A* and *B*

* The work on which this paper is based was supported by ONR Contract N 049–149.

becomes conditional on C's value or state then a necessary component of "organization" is present. Thus *the theory of organization is partly co-extensive with the theory of functions of more than one variable.*

We can get another angle on the question by asking "what is its converse?" The converse of "conditional on" is "not conditional on", so the converse of "organization" must therefore be, as the mathematical theory shows as clearly, the concept of "reducibility". (It is also called "separability".) This occurs, in mathematical forms, when what looks like a function of several variables (perhaps very many) proves on closer examination to have parts whose actions are *not* conditional on the values of the other parts. It occurs in mechanical forms, in hardware, when what looks like one machine proves to be composed of two (or more) sub-machines, each of which is acting independently of the others.

Questions of "conditionality", and of its converse "reducibility", can, of course, be treated by a number of mathematical and logical methods. I shall say something of such methods later. Here, however, I would like to express the opinion that the method of Uncertainty Analysis, introduced by Garner and McGill (1956), gives us a method for the treatment of conditionality that is not only completely rigorous but is also of extreme generality. Its great generality and suitability for application to complex behavior, lies in the fact that it is applicable to any arbitrarily defined set of states. Its application requires neither linearity, nor continuity, nor a metric, nor even an ordering relation. By this calculus, the *degree* of conditionality can be measured, and analyzed, and apportioned to factors and interactions in a manner exactly parallel to Fisher's method of the analysis of variance; yet it requires no metric in the variables, only the frequencies with which the various combinations of states occur. It seems to me that, just as Fisher's conception of the analysis of variance threw a flood of light on to the complex relations that may exist between variations on a metric, so McGill and Garner's conception of uncertainty analysis may give us an altogether better understanding of how to treat complexities of relation when the variables are non-metric. In psychology and biology such variables occur with great commonness; doubtless they will also occur commonly in the brain-like

processes developing in computers. I look forward to the time when the methods of McGill and Garner will become the accepted language in which such matters are to be thought about and treated quantitatively.

The treatment of "conditionality" (whether by functions of many variables, by correlation analysis, by uncertainty analysis, or by other ways) makes us realize that the essential idea is that there is first a product space—that of the *possibilities*—within which some sub-set of points indicates the actualities. This way of looking at "conditionality" makes us realize that it is related to that of "communication"; and it is, of course, quite plausible that we should define parts as being "organized" when "communication" (in some generalized sense) occurs between them. (Again the natural converse is that of independence, which represents non-communication.)

Now "communication" from A to B necessarily implies some constraint, some correlation between what happens at A and what at B. If, for given event at A, all possible events may occur at B, then there is no communication from A to B and no constraint over the possible (A, B)-couples that can occur. Thus the presence of "organization" between variables is equivalent to the existence of a *constraint* in the product-space of the possibilities. I stress this point because while, in the past, biologists have tended to think of organization as something extra, something *added* to the elementary variables, the modern theory, based on the logic of communication, regards organization as a restriction or constraint. The two points of view are thus diametrically opposed; there is no question of either being exclusively right, for each can be appropriate in its context. But with this opposition in existence we must clearly go carefully, especially when we discuss with others, lest we should fall into complete confusion.

This excursion may seem somewhat complex but it is, I am sure, advisable, for we have to recognize that the discussion of organization theory has a peculiarity not found in the more objective sciences of physics and chemistry. The peculiarity comes in with the product space that I have just referred to. Whence comes this product space? Its chief peculiarity is that *it contains more than actually exists in the real physical world*, for it is the latter that gives us the actual, constrained *subset*.

The real world gives the subset of what *is*; the product space represents the uncertainty of the *observer*. The product space may therefore change if the observer changes; and two observers may legitimately use different product spaces within which to record the same subset of actual events in some actual thing. The "constraint" is thus a *relation* between observer and thing; the properties of any particular constraint will depend on both the real thing and on *the observer*. It follows that a substantial part of the theory of organization will be concerned with *properties that are not intrinsic to the thing but are relational between observer and thing*. We shall see some striking examples of this fact later.

WHOLE AND PARTS

"If conditionality" is an essential component in the concept of organization, so also is the assumption that we are speaking of a whole composed of parts. This assumption is worth a moment's scrutiny, for research is developing a theory of dynamics that does *not* observe parts and their interactions, but treats the system as an unanalysed whole (Ashby, 1958, a). In physics, of course, we usually start the description of a system by saying "Let the variables be $x_1, x_2, \ldots, x_n$" and thus start by treating the whole as made of n functional parts. The other method, however, deals with unanalysed states, $S_1, S_2, \ldots$ of the whole, without explicit mention of any parts that may be contributing to these states. The dynamics of such a system can then be defined and handled mathematically; I have shown elsewhere (Ashby, 1960, a) how such an approach can be useful. What I wish to point out here is that we can have a sophisticated *dynamics*, of a whole as complex and cross-connected as you please, that makes no reference to any parts and that therefore does *not* use the concept of organization. Thus the concepts of dynamics and of organization are essentially independent, in that all four combinations, of their presence and absence, are possible.

This fact exemplifies what I said, that "organization" is partly in the eye of the beholder. Two observers studying the same real material system, a hive of bees say, may find that one of them, thinking of the hive as an interaction of fifty thousand bee-parts, finds the bees "organized", while the other, observing whole states

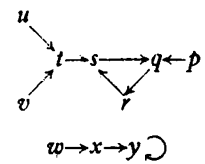
such as activity, dormancy, swarming, etc., may see *no* organization, only trajectories of these (unanalysed) states.

Another example of the independence of "organization" and "dynamics" is given by the fact that whether or not a real system is organized or reducible depends partly on the point of view taken by the observer. It is well known, for instance, that an organized (i.e. interacting) linear system of n parts, such as a network of pendulums and springs, can be seen from another point of view (that of the so-called "normal" coordinates) in which all the (newly identified) parts are completely separate, so that the whole is reducible. There is therefore nothing perverse about my insistence on the relativity of organization, for advantage of the fact is routinely taken in the study of quite ordinary dynamic systems.

Finally, in order to emphasize how dependent is the organization seen in a system on the observer who sees it, I will state the proposition that: given a whole with arbitrarily given behavior, a great variety of arbitrary "parts" can be seen in it; for all that is necessary, when the arbitrary part is proposed, is that we assume the given part to be coupled to another suitably related part, so that the two together form a whole isomorphic with the whole that was given. For instance, suppose the given whole, W of 10 states, behaves in accordance with the transformation:

W	↓	p	q	r	s	t	u	v	w	x	y
		q	r	s	q	s	t	t	x	y	y

Its kinematic graph is



and suppose we wish to "see" it as containing the part P. with internal states E and input states A:

		E	
	↓	1	2
A	1	2	1
	2	1	1

$\Big\}P$

With a little ingenuity we find that if part P is coupled to part Q (with states (F, G) and input B) with transformation Q:

		(F, G)						
	↓	1, 1	1, 2	1, 3	2, 1	2, 2	2, 3	
B	1	2, 1	1, 2	1, 2	2, 1	1, 2	1, 2	Q
	2	·	2, 3	·	2, 1	2, 2	2, 2	

by putting $A = F$ and $B = E$, then the new whole W' has transformation

W':	↓	1, 1, 1	1, 1, 2	1, 1, 3	1, 2, 1, etc.
		2, 2, 1	2, 1, 2	2, 1, 2	1, 2, 1, etc.

which is *isomorphic with* W under the one–one correspondence

	↓	1, 1, 1	1, 1, 2	1, 1, 3	1, 2, 1, etc.
		w	s	p	y , etc.

Thus, subject only to certain requirements (e.g. that equilibria map into equilibria) *any dynamic system can be made to display a variety of arbitrarily assigned "parts"*, simply by a change in the *observer's* view point.

MACHINES IN GENERAL

I have just used a way of representing two "parts", "coupled" to form a "whole", that anticipates the question: what do we mean by a "machine" in general?

Here we are obviously encroaching on what has been called "general system theory", but this last discipline always seemed to me to be uncertain whether it was dealing with *physical* systems, and therefore tied to whatever the real world provides, or with mathematical systems, in which the sole demand is that the work shall be free from internal contradictions. It is, I think, one of the substantial advances of the last decade that we have at last identified the *essentials* of the "machine in general".

Before the essentials could be seen, we had to realize that two factors must be *excluded as irrelevant*. The first is "materiality"—the idea that a machine must be made of actual matter, of the hundred or so existent elements. This is wrong, for examples can

readily be given (e.g. Ashby, 1958, a) showing that what is essential is whether the system, of angels and ectoplasm if you please, *behaves* in a law-abiding and machine-like way. Also to be excluded as irrelevant is any reference to energy, for any calculating machine shows that what matters is the *regularity* of the behavior—whether energy is gained or lost, or even created, is simply irrelevant.

The fundamental concept of "machine" proves to have a form that was formulated at least a century ago, but this concept has not, so far as I am aware, ever been used and exploited vigorously. A "machine" is that which behaves in a machine-like way, namely, that its internal state, and the state of its surroundings, defines uniquely the next state it will go to.

This definition, formally proposed fifteen years ago (Ashby, 1945) has withstood the passage of time and is now becoming generally accepted (e.g. Jeffrey, 1959). It appears in many forms. When the variables are continuous it corresponds to the description of a dynamic system by giving a set of ordinary differential equations with time as the independent variable. The *fundamental* nature of such a representation (as contrasted with a merely convenient one) has been recognized by many earlier workers such as Poincaré, Lotka (1925), and von Bertalanffy (1950 and earlier).

Such a representation by differential equations is, however, too restricted for the needs of a science that includes biological systems and calculating machines, in which discontinuity is ubiquitous. So arises the modern definition, able to include both the continuous and the discontinuous and even the discrete, without the slightest loss of rigor. The "machine with input" (Ashby, 1958, a) or the "finite automaton" (Jeffrey, 1959) is today defined by a set S of internal states, a set I of input or surrounding states, and a mapping, f say, of the product set $I \times S$ into S. Here, in my opinion, we have the very essence of the "machine"; all known types of machine are to be found here; and all interesting deviations from the concept are to be found by the corresponding deviation from the definition.

We are now in a position to say without ambiguity or evasion what we mean by a machine's "organization". First we specify which system we are talking about by specifying its states S and its

conditions I. If S is a product set, so that $S = \Pi_i T_i$ say, then the parts i are each specified by its set of states T_i. *The "organization" between these parts is then specified by the mapping f.* Change f and the organization changes. In other words, the possible organizations between the parts can be set into one–one correspondence with the set of possible mappings of $I \times S$ into S. Thus "organization" and "mapping" are two ways of looking at the same thing—the organization being noticed by the observer of the actual system, and the mapping being recorded by the person who represents the behavior in mathematical or other symbolism.

"GOOD" ORGANIZATION

At this point some of you, especially the biologists, may be feeling uneasy; for this definition of organization makes no reference to any *usefulness* of the organization. It demands only that there be conditionality between the parts and regularity in behavior. In this I believe the definition to be right, for the question whether a given organization is "good" or "bad" is quite independent of the prior test of whether it is or is not an organization.

I feel inclined to stress this point, for here the engineers and the biologists are likely to think along widely differing lines. The engineer, having put together some electronic hardware and having found the assembled network to be roaring with parasitic oscillations, is quite accustomed to the idea of a "bad" organization; and he knows that the "good" organization has to be searched for. The biologist, however, studies mostly animal species that have survived the long process of natural selection; so almost all the organizations he sees have already been selected to be good ones, and he is apt to think of "organizations" as *necessarily* good. This point of view may often be true in the biological world but it is most emphatically not true in the world in which we people here are working. We *must* accept that

(1) most organizations are bad ones;

(2) the good ones have to be sought for; and

(3) what is meant by "good" must be clearly defined, explicitly if necessary, *in every case*.

What then is meant by "good", in our context of brain-like mechanisms and computers? We must proceed cautiously, for the

word suggests some evaluation whose origin has not yet been considered.

In some cases the distinction between the "good" organization and the "bad" is obvious, in the sense that as everyone in these cases would tend to use the same criterion, it would not need explicit mention. The brain of a living organism, for instance, is usually judged as having a "good" organization if the organization (whether inborn or learned) acts so as to further the organism's survival. This consideration readily generalizes to all those cases in which the organization (whether of a cat or an automatic pilot or an oil refinery) is judged "good" if and only if it acts so as to keep an assigned set of variables, the "essential" variables, within assigned limits. Here are all the mechanisms for homeostasis, both in the original sense of Cannon and in the generalized sense. From this criterion comes the related one that an organization is "good" if it makes the system stable around an assigned equilibrium. Sommerhoff (1950) in particular has given a wealth of examples, drawn from a great range of biological and mechanical phenomena, showing how in all cases the idea of a "good organization" has as its essence the idea of a number of parts so interacting as to achieve some given "focal condition". I would like to say here that I do not consider that Sommerhoff's contribution to our subject has yet been adequately recognized. His identification of *exactly* what is meant by coordination and integration is, in my opinion, on a par with Cauchy's identification of exactly what was meant by convergence. Cauchy's discovery was a real discovery, and was an enormous help to later workers by providing them with a concept, rigorously defined, that could be used again and again, in a vast range of contexts, and always with exactly the same meaning. Sommerhoff's discovery of how to represent *exactly* what is meant by coordination and integration and good organization will, I am sure, eventually play a similarly fundamental part in our work.

His work illustrates, and emphasizes, what I want to say here—*there is no such thing as "good organization" in any absolute sense.* Always it is relative; and an organization that is good in one context or under one criterion may be bad under another.

Sometimes this statement is so obvious as to arouse no opposition. If we have half a dozen lenses, for instance, that can be

assembled this way to make a telescope or that way to make a microscope, the goodness of an assembly obviously depends on whether one wants to look at the moon or a cheese mite.

But the subject is more contentious than that! The thesis implies that there is no such thing as a brain (natural or artificial) that is good in any absolute sense—it all depends on the circumstances and on what is wanted. Every faculty that a brain can show is "good" only conditionally, for there exists at least one environment against which the brain is handicapped by the possession of this faculty. Sommerhoff's formulation enables us to show this at once: whatever the faculty or organization achieves, let that be *not* in the "focal conditions".

We know, of course, lots of examples where the thesis is true in a somewhat trivial way. Curiosity tends to be good, but many an antelope has lost its life by stopping to see what the hunter's hat is. Whether the organization of the antelope's brain should be of the type that does, or does not, lead to temporary immobility clearly depends on whether hunters with rifles are or are not plentiful in its world.

From a different angle we can notice Pribram's results (1957), who found that brain-operated monkeys scored higher in a certain test than the normals. (The operated were plodding and patient while the normals were restless and distractible.) Be that as it may, one cannot say which brain (normal or operated) had the "good" organization until one has decided which sort of temperament is wanted.

Do you still find this non-contentious? Then I am prepared to assert that there is not a single mental faculty ascribed to Man that is good in the absolute sense. If any particular faculty is *usually* good, this is solely because our terrestrial environment is so lacking in variety that its usual form makes that faculty usually good. But change the environment, go to really different conditions, and possession of that faculty may be harmful. And "bad", by implication, is the brain organization that produces it.

I believe that there is not a single faculty or property of the brain, usually regarded as desirable, that does not become *un*desirable in some type of environment. Here are some examples in illustration.

The first is Memory. Is it not good that a brain should have

memory? Not at all, I reply—only when the environment is of a type in which the future often *copies* the past; should the future often be the *inverse* of the past, memory is actually disadvantageous. A well known example is given when the sewer rat faces the environmental system known as "pre-baiting". The naïve rat is very suspicious, and takes strange food only in small quantities. If, however, wholesome food appears at some place for three days in succession, the sewer rat will learn, and on the fourth day will eat to repletion, and die. The rat without memory, however, is as suspicious on the fourth day as on the first, and lives. Thus, in *this* environment, memory is positively disadvantageous. Prolonged contact with this environment will lead, other things being equal, to evolution in the direction of diminished memory-capacity.

As a second example, consider organization itself in the sense of connectedness. Is it not good that a brain should have its parts in rich functional connection? I say, No—not *in general*; only when the environment is itself richly connected. When the environment's parts are *not* richly connected (when it is highly reducible, in other words), adaptation will go on faster if the brain is also highly reducible, i.e. if its connectivity is small (Ashby, 1960, d). Thus the *degree* of organization can be too high as well as too low; the degree we humans possess is probably adjusted to be somewhere near the optimum for the usual terrestrial environment. It does not in any way follow that this degree will be optimal or good if the brain is a mechanical one, working against some grossly non-terrestrial environment—one existing only inside a big computer, say.

As another example, what of the "organization" that the biologist always points to with pride—the development in evolution of specialized organs such as brain, intestines, heart and blood vessels. Is not this good? Good or not, it is certainly a specialization made possible only because the earth has an atmosphere; without it, we would be incessantly bombarded by tiny meteorites, any one of which, passing through our chest, might strike a large blood vessel and kill us. Under such conditions a better form for survival would be the slime mould, which specializes in being able to flow through a tangle of twigs without loss of function. Thus the development of organs is not good unconditionally, but is a specialization to a world free from flying particles.

After these actual instances, we can return to theory. It is here that Sommerhoff's formulation gives such helpful clarification. He shows that in all cases there must be given, and specified, first a *set of disturbances* (values of his "coenetic variable") and secondly a goal (his "focal condition"); the disturbances threaten to drive the outcome outside the focal condition. The "good" organization is then of the nature of a *relation* between the set of disturbances and the goal. Change the set of disturbances, and the organization, without itself changing, is evaluated "bad" instead of "good". As I said, there is no property of an organization that is good in any absolute sense; all are relative to some given environment, or to some given set of threats and disturbances, or to some given set of problems.

SELF-ORGANIZING SYSTEMS

I hope I have not wearied you by belaboring this relativity too much, but it is fundamental, and is only too readily forgotten when one comes to deal with organizations that are either biological in origin or are in imitation of such systems. With this in mind, we can now start to consider the so-called "self-organizing" system. We must proceed with some caution here if we are not to land in confusion, for the adjective is, if used loosely, ambiguous, and, if used precisely, self-contradictory.

To say a system is "self-organizing" leaves open two quite different meanings.

There is a first meaning that is simple and unobjectionable. This refers to the system that starts with its parts separate (so that the behavior of each is independent of the others' states) and whose parts then act so that they change towards forming connections of some type. Such a system is "self-organizing" in the sense that it changes from "parts separated" to "parts joined". An example is the embryo nervous system, which starts with cells having little or no effect on one another, and changes, by the growth of dendrites and formation of synapses, to one in which each part's behavior is very much affected by the other parts. Another example is Pask's system of electrolytic centers, in which the growth of a filament from one electrode is at first little affected by growths at the other electrodes; then the growths become

more and more affected by one another as filaments approach the other electrodes. In general such systems can be more simply characterized as "self-*connecting*", for the change from independence between the parts to conditionality can always be seen as some form of "connection", even if it is as purely functional as that from a radio transmitter to a receiver.

Here, then, is a perfectly straightforward form of self-organizing system; but I must emphasize that there can be no assumption at this point that the organization developed will be a good one. If we wish it to be a "good" one, we must first provide a criterion for distinguishing between the bad and the good, and then we must ensure that the appropriate selection is made.

We are here approaching the second meaning of "self-organizing" (Ashby, 1947). "Organizing" may have the first meaning, just discussed, of "changing from unorganized to organized". But it may also mean "changing from a bad organization to a good one", and this is the case I wish to discuss now, and more fully. This is the case of peculiar interest to *us*, for this is the case of the system that changes itself from a bad way of behaving to a good. A well known example is the child that starts with a brain organization that makes it fire-seeking; then a change occurs, and a new brain organization appears that makes the child fire-avoiding. Another example would occur if an automatic pilot and a plane were so coupled, by mistake, that positive feedback made the whole error-aggravating rather than error-correcting. Here the organization is bad. The system would be "self-organizing" if a change were *automatically* made to the feedback, changing it from positive to negative; then the whole would have changed from a bad organization to a good. Clearly, *this* type of "self-organization" is of peculiar interest to us. What is implied by it?

Before the question is answered we must notice, if we are not to be in perpetual danger of confusion, that *no machine can be self-organizing in this sense*. The reasoning is simple. Define the set S of states so as to specify which machine we are talking about. The "organization" must then, as I said above, be identified with f, the mapping of S into S that the basic drive of the machine (whatever force it may be) imposes. Now the logical relation here is that f determines the changes of S:—f is *defined* as the set of

couples (s_i, s_j) such that the internal drive of the system will force state s_i to change to s_j. To allow f to be a function of the state is to make nonsense of the whole concept.

Since the argument is fundamental in the theory of self-organizing systems, I may help explanation by a parallel example. Newton's law of gravitation says that $F = M_1M_2/d^2$, in particular, that the force varies inversely as the distance to power 2. To power 3 would be a different law. But suppose it were suggested that, not the force F but the *law* changed with the distance, so that the power was not 2 but some function of the distance, $\phi(d)$. This suggestion is illogical; for we now have that $F = M_1M_2/d^{\phi(d)}$, and this represents not a law that varies with the distance but *one* law covering all distances; that is, were this the case we would *re-define* the law. Analogously, were f in the machine to be some function of the state S, we would have to re-define our machine. Let me be quite explicit with an example. Suppose S had three states: a, b, c. If f depended on S there would be three f's: f_a, f_b, f_c say. Then if they are

↓	a	b	c
f_a	**b**	a	b
f_b	c	**a**	a
f_c	b	b	**a**

then the transform of a must be under f_a, and is therefore b, so the whole set of f's would amount to the *single* transformation:

	a	b	c
↓	b	a	a

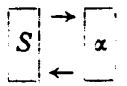
It is clearly illogical to talk of f as being a function of S, for such talk would refer to operations, such as $f_a(b)$, which cannot in fact occur.

If, then, no machine can properly be said to be self-organizing, how do we regard, say, the Homeostat, that rearranges its own wiring; or the computer that writes out its own program?

The new logic of mechanism enables us to treat the question rigorously. We start with the set S of states, and assume that f changes, to g say. So we really have a *variable*, $\alpha(t)$ say, a function of the time that had at first the value f and later the value g. This

change, as we have just seen, cannot be ascribed to any cause in the set S; so it must have come from some outside agent, acting on the system S as input. If the system is to be in some sense "*self*-organizing", the "self" must be enlarged to include this variable α, and, to keep the whole bounded, the cause of α's change must be in S (or α).

Thus the appearance of being "self-organizing" can be given only by the machine S being coupled to another machine (of one part):



Then the part S can be "self-organizing" within the whole $S+\alpha$.

Only in this partial and strictly qualified sense can we understand that a system is "*self*-organizing" without being self-contradictory.

Since no system can correctly be said to be self-organizing, and since use of the phrase "self-organizing" tends to perpetuate a fundamentally confused and inconsistent way of looking at the subject, the phrase is probably better allowed to die out.

THE SPONTANEOUS GENERATION OF ORGANIZATION

When I say that no system can properly be said to be self-organizing, the listener may not be satisfied. What, he may ask, of those changes that occurred a billion years ago, that led lots of carbon atoms, scattered in little molecules of carbon dioxide, methane, carbonate, etc., to get together until they formed proteins, and then went on to form those large active lumps that today we call "animals"? Was not this process, on an isolated planet, one of "self-organization"? And if it occurred on a planetary surface can it not be made to occur in a computer? I am, of course, now discussing the origin of life. Has modern system theory anything to say on this topic?

It has a great deal to say, and some of it flatly contradictory to what has been said ever since the idea of evolution was first considered. In the past, when a writer discussed the topic, he usually assumed that the generation of life was rare and peculiar,

and he then tried to display some way that would enable this rare and peculiar event to occur. So he tried to display that there is *some* route from, say, carbon dioxide to the amino acid, and thence to the protein, and so, through natural selection and evolution, to intelligent beings. I say that this looking for special conditions is quite wrong. The truth is the opposite—*every* dynamic system generates its own form of intelligent life, is self-organizing in this sense. (I will demonstrate the fact in a moment.) Why we have failed to recognize this fact is that until recently we have had no experience of systems of medium complexity; either they have been like the watch and the pendulum, and we have found their properties few and trivial, or they have been like the dog and the human being, and we have found their properties so rich and remarkable that we have thought them supernatural. Only in the last few years has the general-purpose computer given us a system rich enough to be interesting yet still simple enough to be understandable. With this machine as tutor we can now begin to think about systems that are simple enough to be comprehensible in detail yet also rich enough to be suggestive. With their aid we can see the truth of the statement that *every isolated determinate dynamic system obeying unchanging laws will develop "organisms" that are adapted to their "environments".*

The argument is simple enough in principle. We start with the fact that systems in general go to equilibrium. Now most of a system's states are non-equilibrial (if we exclude the extreme case of the system in neutral equilibrium). So in going from *any* state to one of the equilibria, the system is going from a larger number of states to a smaller. In this way it is performing a selection, in the purely objective sense that it rejects some states, by leaving them, and retains some other state, by sticking to it. Thus, as every determinate system goes to equilibrium, so does it select. We have heard *ad nauseam* the dictum that a machine cannot select; the truth is just the opposite: every machine, as it goes to equilibrium, performs the corresponding act of selection.

Now, equilibrium in simple systems is usually trivial and uninteresting; it is the pendulum hanging vertically; it is the watch with its main-spring run down; the cube resting flat on one face. Today, however, we know that when the system is more complex and dynamic, equilibrium, and the stability around it, can be

much more interesting. Here we have the automatic pilot successfully combating an eddy; the person redistributing his blood flow after a severe haemorrhage; the business firm restocking after a sudden increase in consumption; the economic system restoring a distribution of supplies after a sudden destruction of a food crop; and it is a man successfully getting at least one meal a day during a lifetime of hardship and unemployment.

What makes the change, from trivial to interesting, is simply the *scale* of the events. "Going to equilibrium" *is* trivial in the simple pendulum, for the equilibrium is no more than a single point. But when the system is more complex; when, say, a country's economy goes back from wartime to normal methods then the stable region is vast, and much interesting activity can occur within it. The computer is heaven-sent in this context, for it enables us to bridge the enormous conceptual gap from the simple and understandable to the complex and interesting. Thus we can gain a considerable insight into the so-called spontaneous generation of life by just seeing how a somewhat simpler version will appear in a computer.

COMPETITION

Here is an example of a simpler version. The competition between species is often treated as if it were essentially biological; it is in fact an expression of a process of far greater generality. Suppose we have a computer, for instance, whose stores are filled at random with the digits 0 to 9. Suppose its dynamic law is that the digits are continuously being multiplied in pairs, and the right-hand digit of the product going to replace the first digit taken. Start the machine, and let it "evolve"; what will happen? Now under the laws of this particular world, even times even gives even, and odd times odd gives odd. But even times odd gives even; so after a mixed encounter *the even has the better chance of survival.* So as this system evolves, we shall see the evens favored in the struggle, steadily replacing the odds in the stores and eventually exterminating them.

But the evens are not homogeneous, and among them the zeros are best suited to survive in this particular world; and, as we

the *run-in length.* Given a particular initial state, the length of the run-in is the time the system must operate before a cycle commences.

Disclosures. It is useful to refer to the trajectory segment from an initial state to the state just short of any repetition of states in the cycle. Such a trajectory segment is here called a *disclosure.* A *disclosure length* is the sum of a run-in length and a cycle length: the length of time before the system discloses that it is cycling.

Activity. Activity of a system is the fraction of elements that, from one instant in system-time to the next, have changed their element-state. The variable is so named because it is a measure of the extent to which the system is doing things, or communicating internally.

In this study activity is considered in systems which typically go to activity levels of zero. (A system in a cycle length one is maintaining itself in the same state, hence the activity of the system is zero.)

3. Method Used

The systems examined lend themselves readily to digital computer simulation: the method used is direct observation of behavior by computer modeling.

The programming of the computer used, an IBM 7094, is straightforward. Basically, the program calculates a state and compares that state with all previous states. If a previous state is the same as that calculated, a cycle has been disclosed. The necessary data are then printed out.

A parameter, the *aperture,* limits the number of states that are calculated and searched for cycles. The aperture setting used determines the maximum cycle, run-in, and disclosure length detectable in a particular simulation run. In the program written, the following relations hold on the aperture (a):

$$\text{maximum detectable cycle length} = a - 1, \tag{2}$$

$$\text{maximum detectable run-in length} = a - 2, \tag{3}$$

$$\text{Maximum detectable disclosure length} = a - 1. \tag{4}$$

4. Observational Procedure

4.1 Number of Elements in the Systems *(N)*

Since the program's time requirements increase at least linearly with N, and as one may reasonably expect cycle and run-in length to increase with N, considering the resources available, we concentrate on a convenient value of N, namely 100. All the systems observed have this number of elements.

4.2 How the Structures are Selected

To get an indication of the effect of structure on behavior, five randomly chosen structures are used with each transformation. The same five Ks are used throughout the investigation. Thus five systems using each T are examined.

Following the general procedure given in Section 2, Ks are constructed by the use of Kendall and Smiths' (1951) table of random decimal digits. Despite their convenience, pseudo-random number generators were avoided because they have been shown (Greenberger, 1962) to be quite capable of producing serially correlated numbers.

4.3 How the Initial States are Selected

Each system's behavior is sampled by initiating trajectory segments from states selected equiprobably and independently from among all states of the system. The actual selection of states is carried out by a computer program in which the high-order bit of N binary numbers generated by a pseudo-random number generator is assigned in turn to each of the system's elements. (The use here of a pseudo-random number generator is not objectionable, since even if the generated initial states come from a population that is correlated, the correlation can be reasonably assumed to be unrelated to characteristics of the field.) Ten states generated by this method are used as initial states throughout the investigation.

4.4 The Number of States in Each Observed Trajectory Segment

The primary search aperture is 500; however, at least one segment in each K is run with $a = 5$, 180 when no cycles are found with $a = 500$.

4.5 Summary

A set of Ts sufficient to represent the entire family of 256 is examined. For each T, fifty trajectory segments are generated and searched for cycles: ten trajectories, begun at randomly sampled states, in each of five structurally different systems. The systems examined all have 100 elements.

The data of the study can be visualized as filling a three-dimensional space such as that shown in Fig. 2. A separate space is occupied by cycle lengths, run-in lengths, and disclosure lengths. The same representation can be used for activity values, with the understanding that the points in the space in this case represent sets of numbers rather than single integers.

5. Results and Discussion

5.1 Introduction

It was mentioned previously that the data are discussed primarily from a biological point of view. The question of how a system behaves exactly is neglected in favor of how it behaves usually, an inquiry which is biologically more relevant: knowing factors which influence the style of behavior in these

watch, we shall see the zeros exterminating their fellow-evens, until eventually they inherit this particular earth.

What we have here is an example of a thesis of extreme generality. From one point of view we have simply a well defined operator (the multiplication and replacement law) which drives on towards equilibrium. In doing so it *automatically* selects those operands that are *specially resistant* to its change-making tendency (for the zeros are uniquely resistant to change by multiplication). This process, of progression towards the specially resistant form, is of extreme generality, demanding only that the operator (or the physical laws of any physical system) be determinate and unchanging. This is the general or abstract point of view. The biologist sees a special case of it when he observes the march of evolution, survival of the fittest, and the inevitable emergence of the highest biological functions and intelligence. Thus, when we ask: What was necessary that life and intelligence should appear? the answer is not carbon, or amino acids or any other special feature but only that the dynamic laws of the process should be *unchanging*, i.e. that the system should be *isolated. In any isolated system, life and intelligence inevitably develop* (they may, in degenerate cases, develop to only zero degree).

So the answer to the question: How can we generate intelligence synthetically? is as follows. Take a dynamic system whose laws are unchanging and single-valued, and whose size is so large that after it has gone to an equilibrium that involves only a small fraction of its total states, this small fraction is still large enough to allow room for a good deal of change and behavior. Let it go on for a long enough time to get to such an equilibrium. Then examine the equilibrium in detail. You will find that the states or forms now in being are peculiarly able to survive against the changes induced by the laws. Split the equilibrium in two, call one part "organism" and the other part "environment": you will find that this "organism" is peculiarly able to survive against the disturbances from this "environment". The *degree* of adaptation and complexity that this organism can develop is bounded only by the size of the whole dynamic system and by the time over which it is allowed to progress towards equilibrium. Thus, as I said, every isolated determinate dynamic system will develop organisms that are adapted to their environments. There is thus no difficulty

in principle, in developing synthetic organisms as complex or as intelligent as we please.

In *this* sense, then, *every* machine can be thought of as "self-organizing", for it will develop, to such degree as its size and complexity allow, some functional structure homologous with an "adapted organism". But does this give us what we at this Conference are looking for? Only partly; for nothing said so far has any implication about the organization being good or bad; the criterion that would make the distinction has not yet been introduced. It is true, of course, that the developed organism, being stable, will have its own essential variables, and it will show its stability by vigorous reactions that tend to preserve its own existence. To *itself*, its own organization will *always*, by definition, be good. The wasp finds the stinging reflex a good thing, and the leech finds the blood-sucking reflex a good thing. But these criteria come *after* the organization for survival; having seen *what* survives we then see what is "good" for that form. What emerges depends simply on what are the system's laws and from what state it started; there is no implication that the organization developed will be "good" in any absolute sense, or according to the criterion of any outside body such as ourselves.

To summarize briefly: there is no difficulty, in principle, in developing *synthetic organisms as complex, and as intelligent as we please*. But we must notice two fundamental qualifications; first, their intelligence will be an adaptation to, and a specialization towards, their particular environment, with no implication of validity for any other environment such as ours; and secondly, their intelligence will be directed towards keeping their own essential variables within limits. They will be fundamentally selfish. So we now have to ask: In view of these qualifications, can we yet turn these processes to our advantage?

REQUISITE VARIETY

In this matter I do not think enough attention has yet been paid to Shannon's Tenth Theorem (1949) or to the simpler "law of requisite variety" in which I have expressed the same basic idea (Ashby, 1958, a). Shannon's theorem says that if a correction-channel has capacity H, then equivocation of amount H can be

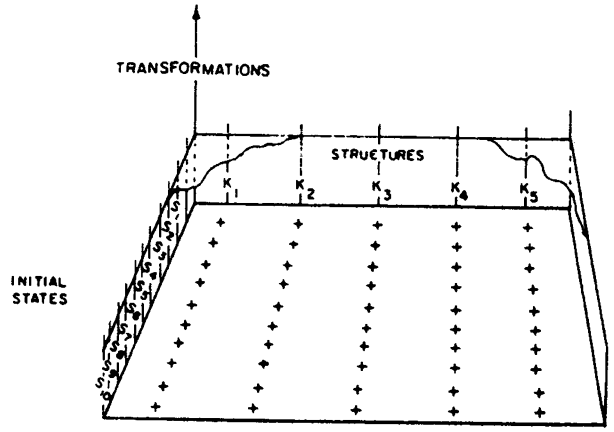


Fig. 2. The experimental plan

abstract systems may suggest actual control methods used by organisms, it being at least plausible that evolution may often seize and build on what happens typically. However, it should be noted that this biological orientation is not totally irrelevant to the concerns of machine designers. The problems faced by the designer or manager of truly complex systems may be problems already faced and largely solved by nature herself in existing organisms.

Section 5.2 sets out some of the observed regularities in behavior. In Section 5.3 an attempt is made to relate observed behavior to certain general characteristics of *Ts*.

5.2 Regularities in Behavior

Since behaviors and structures are sampled at random, diversity of behavior is to be expected, and indeed, it is found: over all trajectory segments examined, for example, cycle lengths range from one to 4,040, and run-in lengths from zero to 5,085 states. Under the conditions which prevail in this study, consistency of behavior would seem the more remarkable finding. Behavioral consistency is nevertheless seen frequently, not only within systems, but across systems using the same *T*.

Consistency of behavior is so frequent, in fact, that it is often useful to summarize the common behavior of the five systems that use the same *T* by speaking as if the *T* itself "shows" the common behavior, even though *Ts* themselves cannot properly be said to "behave." In this same connection: when some common behavioral trait is shown by all the systems using the same *T*, that trait is shown by systems whose structure are taken at random from a population of structures. In such cases the *T* may (within limits, due to sampling error, of course) be assumed to produce the observed common

trait without regard for the systems' structure. Such *Ts*, in this paper, are said to be *pragmatically structure-insensitive*, or, just *structure-insensitive* with respect to the common trait. One of the interesting findings is that such structural insensitivity is exhibited by many *Ts*, and with respect to behaviors whose relative constancy over structure appears to be unanticipated.

Disclosure Lengths. The significance of disclosures lies largely in the fact that disclosure lengths stand as a barrier to the understanding of a system by simple observation. If a naive observer is to understand the system, in the sense of being able to distinguish its temporary from its terminal behavior, his capacity to record and compare states (his aperture) must exceed the lengths of the system's longest disclosures.

In the systems of the present study, disclosures can conceivably be of any length between zero and, roughly, 10^{30} states. It is therefore reasonable to suspect that the disclosures of the systems examined might typically be too long for cycles to be detected. The facts are that the bulk of the trajectory segments examined have disclosures less than 500 states in length. Of the family of 256 *Ts*, there are 196 *Ts* that, using an aperture of 500 states, show disclosures in over half of their fifty trajectory segments. That is, for almost 75 per cent of the *Ts*, cycles are found using an aperture of 500, over half of the time in starts from random initial states, regardless of the systems' structure.

A condition that contributes importantly to the tendency of a *T* to have short disclosures is that *T* have a first column of the standard form (see Table 1), and reading from the top down, resembling 0000, or a second column resembling 1111. A condition that contributes to a *T's* having long disclosures is that *T* have a column resembling either 0110 or 1001. Reasons why these characteristics might be expected to influence the length of disclosures are suggested in Section 5.3.

Periodicity of Permanent Behavior (Cycle Lengths). Cycles of Length One. Of the 12,800 trajectory segments of the family of 256 *Ts*, 4,010 segments terminate in equilibrial states. The observed overall relative frequency of equilibrial states is therefore approximately 0.314. Taking into account 2,754 trajectory segments for which terminal cycles are not found, and since these segments *may* end in equilibrial states, an upper bound on the relative number of equilibrial states among all trajectory segments is 0.529. Therefore, for the family of systems examined, given the sampling conditions used in the present study, the frequency with which trajectory segments end in an equilibrial state is between roughly 1/3 and 1/2.

There are *Ts* observed for which equilibrial states are found very frequently. 55 *Ts*, about one out of five, produce equilibrial states, with an

removed, *but no more*. Shannon stated his theorem in the context of telephone or similar communication, but the formulation is just as true of a biological regulatory channel trying to exert some sort of corrective control. He thought of the case with a lot of message and a little error; the biologist faces the case where the "message" is small but the disturbing errors are many and large. The theorem can then be applied to the brain (or any other regulatory and selective device), when it says that the amount of regulatory or selective action that the brain can achieve is absolutely bounded by its capacity as a channel (Ashby, 1958, b). Another way of expressing the same idea is to say that any quantity K of appropriate selection demands the transmission or processing of quantity K of information (Ashby, 1960, b.) *There is no getting of selection for nothing.*

I think that here we have a principle that we shall hear much of in the future, for it dominates all work with complex systems. It enters the subject somewhat as the law of conservation of energy enters power engineering. When that law first came in, about a hundred years ago, many engineers thought of it as a disappointment, for it stopped all hopes of perpetual motion. Nevertheless, it did in fact lead to the great practical engineering triumphs of the nineteenth century, because it made power engineering more realistic.

I suggest that when the full implications of Shannon's Tenth Theorem are grasped we shall be, first sobered, and then helped, for we shall then be able to focus our activities on the problems that are properly realistic, and actually solvable.

THE FUTURE

Here I have completed this bird's-eye survey of the principles that govern the self-organizing system. I hope I have given justification for my belief that these principles, based on the logic of mechanism and on information theory, are now essentially *complete*, in the sense that there is now no area that is grossly mysterious.

Before I end, however, I would like to indicate, very briefly, the directions in which future research seems to me to be most likely to be profitable.

One direction in which I believe a great deal to be readily discoverable, is in the discovery of new types of dynamic process. Most of the machine-processes that we know today are very specialized, depending on exactly what parts are used and how they are joined together. But there are systems of more net-like construction in which what happens can only be treated statistically. There are processes here like, for instance, the spread of epidemics, the fluctuations of animal populations over a territory, the spread of wave-like phenomena over a nerve-net. These processes are, in themselves, neither good nor bad, but they exist, with all their curious properties, and doubtless the brain will use them should they be of advantage. What I want to emphasize here is that they often show very surprising and peculiar properties; such as the tendency, in epidemics, for the outbreaks to occur in waves. Such peculiar new properties may be just what some machine designer wants, and that he might otherwise not know how to achieve.

The study of such systems must be essentially statistical, but this does not mean that each system must be individually stochastic. On the contrary, it has recently been shown (Ashby, 1960, c) that no system can have greater efficiency than the determinate when acting as a regulator; so, as regulation is the one function that counts biologically, we can expect that natural selection will have made the brain as determinate as possible. It follows that we can confine our interest to the lesser range in which the sample space is over a set of mechanisms each of which is individually determinate.

As a particular case, a type of system that deserves much more thorough investigation is the large system that is built of parts that have many states of equilibrium. Such systems are extremely common in the terrestrial world; they exist all around us, and in fact, intelligence as we know it would be almost impossible otherwise (Ashby, 1960, d). This is another way of referring to the system whose variables behave largely as part-functions. I have shown elsewhere (Ashby, 1960, a) that such systems tend to show habituation (extinction) and to be able to adapt progressively (Ashby, 1960, d). There is reason to believe that some of the well-known but obscure biological phenomena such as conditioning, association, and Jennings' (1906) law of the resolution of physiological states may be more or less simple and direct expressions

aperture of 500, in at least 48 of their 50 trajectory segments. For obvious reasons, these *Ts* are here called ***equilibrial Ts***. Such *Ts* are, of course, structure-insensitive with respect to the production of equilibrial states. Characteristics in *T* that influence this extreme tendency to produce equilibrial states are considered in Section 5.3.

It bears mentioning that the equilibrial states produced by equilibrial *Ts* are not necessarily trivial. That is, the states are not "all 1's" or "all 0's". A particular example is *T:* 01 for which the equilibrial states observed show very
00
11
01

nearly half 1's and half 0's. (While these terminal densities of element-states are close to that predicted -- namely, exactly half 1's and half 0's -- by an elementary probability model of densities in these systems, in the case of other *Ts*, the discrepancies are large.)

Cycles of Length More than One. For the family of systems examined, it can be seen that the frequency with which trajectory segments end in a cycle of length more than one is between roughly 1/2 and 2/3. Thus, if a distinction is drawn between equilibrial states as being "steady" behavior, and cycles of length more than one as being "rhythmic" behavior, as compared with steady behavior, the systems studied most often show rhythmic terminal behavior.

Two regularities that hold across structure are found among the larger cycles: (1) where a *T*'s cycles are of length two, and (2) where a *T*'s cycle lengths are multiples of a (non-unity) common factor.

When all of a *T*'s cycles are of length two, the *T* is said here to be a ***doubled T***. Doubled *Ts* are not given much attention here, as their appearance is likely due in large measure to the fact that elements all have two element-states.

More surprising is that some *Ts* show cycle lengths which, while differing, are all multiples of a common factor greater than one. The factors (greatest common divisors) found are two, three, four, and eight. *Ts* showing such behaviors are here called ***multiple Ts***. As examples, producing cycles that have a greatest common divisor (g.e.d.) of three are *Ts:* 10 and 01;

01	11
10	10
00	10

producing cycles that have a g.e.d. of four are *Ts:* 01 where * indicates either 0 or 1.
10
0*
0*

While g.e.d.'s two, four, and eight suggest that the binary element-states may be of some importance in determining the magnitude of the g.e.d., it is difficult to see how the binarity of the element-states could produce a g.e.d. of three. For the present, the general conditions that produce multiple *Ts* will be left for future clarification.

It might be mentioned that some *Ts* show a tendency to have cycles with relatively long, prime lengths. For example, *T*: 00 shows cycles of length
11
01
10

653 (prime). This *T* also shows a cycle of length 2,391, which has a highest prime factor of 797. *Ts* with these behaviors are rather few, however, (four out of 256), and the appearance of the prime cycle lengths is sensitive to structural change.

Temporary Behavior (Run-In Lengths): Distribution of Run-In Lengths. Run-in lengths (of those *Ts* for which sufficient numbers of disclosures were observed) have distributions (within *Ts*) which can be largely described as unimodal, usually with a skew in the direction of longer run-ins. While there are exceptions, most *Ts'* run-in lengths are distributed in a manner not markedly different from a normal curve.

Progression of Activity in Equilibrial Ts. What is the general nature of the progression in equilibrial *Ts*? The answer is not unexpected: Activity, in systems which use equilibrial *Ts*, and which start from random initial states, falls off to zero in a fairly uniform, exponential decay.

5.3 Behavior as Related to Characteristics of Transformations.

In this section, disclosure, cycle, and run-in lengths are related to characteristics of transformations. The aim is to examine certain measures which can be extended without great difficulty to tables that give the behavior of components with many component-states, and many input wires.

The measures examined are:

(1) *Internal homogeneity* -- reflects the tendency for *T* to output the same state on successive occasions.

This measure is scored by separately counting the "0" and "1" entries in *T*; the greater number is the internal homogeneity of *T*. The measure scores *Ts* from 4 to 8, with increasing numbers indicating greater sameness of the entries of *T*.

(2) *Fluency* reflects the extent to which *T* resembles those *Ts* that provide maximum through-put of signalling from an element's input to output, when one input is being held constant.

Ts with maximum fluency have columns of the forms: 0110 and 1001. Fluency is scored (for the simple *Ts* of this study) by matching the entries in one column of a *T* against both column-forms (one at a time) given above,

of the multiplicity of equilibrial states. At the moment I am investigating the possibility that the transfer of "structure", such as that of three-dimensional space, into a dynamic system—the sort of learning that Piaget has specially considered—may be an *automatic* process when the input comes to a system with many equilibria. Be that as it may, there can be little doubt that the study of such systems is likely to reveal a variety of new dynamic processes, giving us dynamic resources not at present available.

A particular type of system with many equilibria is the system whose parts have a high "threshold"—those that tend to stay at some "basic" state unless some function of the input exceeds some value. The general properties of such systems is still largely unknown, although Beurle (1956) has made a most interesting start. They deserve extensive investigation; for, with their basic tendency to develop avalanche-like waves of activity, their dynamic properties are likely to prove exciting and even dramatic. The fact that the mammalian brain uses the property extensively suggests that it may have some peculiar, and useful, property not readily obtainable in any other way.

Reference to the system with many equilibria brings me to the second line of investigation that seems to me to be in the highest degree promising—I refer to the discovery of *the living organism's memory store*: the identification of its physical nature.

At the moment, our knowledge of the living brain is grossly out of balance. With regard to what happens from one millisecond to the next we know a great deal, and many laboratories are working to add yet more detail. But when we ask what happens in the brain from one hour to the next, or from one year to the next, practically nothing is known. Yet it is these longer-term changes that are the really significant ones in human behavior.

It seems to me, therefore, that if there is one thing that is crying out to be investigated it is the physical basis of the brain's memory-stores. There was a time when "memory" was a very vague and metaphysical subject; but those days are gone. "Memory", as a *constraint* holding over events of the past and the present, and a *relation* between them, is today firmly grasped by the logic of mechanism. We know exactly what we mean by it behavioristically and operationally. What we need now is the provision of adequate

resources for its investigation. Surely the time has come for the world to be able to find resources for *one* team to go into the matter?

SUMMARY

Today, the principles of the self-organizing system are known with some completeness, in the sense that no major part of the subject is wholly mysterious.

We have a secure base. Today we know *exactly* what we mean by "machine", by "organization", by "integration", and by "self-organization". We understand these concepts as thoroughly and as rigorously as the mathematician understands "continuity" or "convergence".

In these terms we can see today that the artificial generation of dynamic systems with "life" and "intelligence" is not merely simple—it is unavoidable if only the basic requirements are met. These are not carbon, water, or any other material entities but the persistence, over a long time, of the action of any operator that is both unchanging and single-valued. *Every* such operator forces the development of its own form of life and intelligence.

But will the forms developed be of use to *us*? Here the situation is dominated by the basic law of requisite variety (and Shannon's Tenth Theorem), which says that the achieving of appropriate selection (to a degree better than chance) is absolutely dependent on the processing of at least that quantity of information. Future work must respect this law, or be marked as futile even before it has started.

Finally, I commend as a program for research, the *identification of the physical basis of the brain's memory stores*. Our knowledge of the brain's functioning is today grossly out of balance. A vast amount is known about how the brain goes from state to state at about millisecond intervals; but when we consider our knowledge of the basis of the important long-term changes we find it to amount, practically, to nothing. I suggest it is time that we made some definite attempt to attack this problem. Surely it is time that the world had *one* team active in this direction?

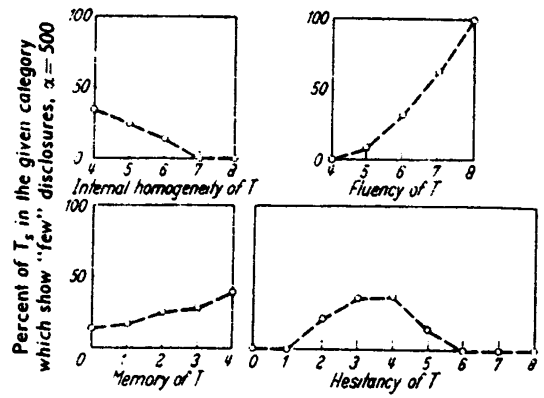


Fig. 3. Per cent of *Ts* which show "few" (in less than half of all trajectory segments examined) disclosures, $\alpha = 500$

ordinate, due to a high per cent of *Ts* that showed relatively few disclosures, indicates that, in that category, disclosures, with respect to an aperture of 500 states, tend to be long.

As can be seen in Fig. 3, the tendency for a system to show longer disclosures is encouraged by low internal homogeneity, high fluency, high memory, and intermediate hesitancy. Fluency, which can be thought of as a measure of the element's capacity to transmit information, is clearly the most potent factor; memory is the least effective, the apparent relationship shown being statistically questionable.

Considering the fact that increasing the hesitancy of the systems' elements can encourage an increase as well as a decrease the length of disclosures, it may be of interest to point out that where *T* has the lowest hesitancy it is trivially true that a system's field is then entirely filled with cycles of length two, and therefore, all disclosures are of length two. (In more complex systems using components with *n* component-states, the lowest hesitancy can give rise to *Ts* whose systems have fields entirely filled with cycles of length *n*, therefore, once again, all disclosures are relatively small, in this case of length *n*.) On the other hand, it should be obvious that the *T* with greatest hesitancy produces fields entirely filled with cycles (and disclosures) of length one. This makes clear a point easily overlooked: strong behavioral constraints (so far as disclosure lengths are concerned) are produced by *both* extremes of hesitancy.

Equilibrial Ts. The relations between the occurrence of equilibrial *Ts* and the four measures were tested in the same manner used for disclosure lengths. The number of *Ts* that are equilibrial and non-equilibrial, as judged by the standard chi-square test of independence, are reliably associated with internal homogeneity and hesitancy, while the effects of fluency and memory are reasonably accounted for by chance. (The chi-squares are: for internal homogeneity,

REFERENCES

1. W. Ross Ashby, The physical origin of adaptation by trial and error, *J. Gen. Psychol.* **32**, pp. 13–25 (1945).
2. W. Ross Ashby, Principles of the self-organizing dynamic system. *J. Gen. Psychol.* **37**, pp. 125–8 (1947).
3. W. Ross Ashby, *An Introduction to Cybernetics*, Wiley, New York, 3rd imp. (1958, a).
4. W. Ross Ashby, Requisite variety and its implications for the control of complex systems, *Cybernetica*, **1**, pp. 83–99 (1958, b).
5. W. Ross Ashby, The mechanism of habituation. In: *The Mechanization of thought Processes.* (Natl. Phys. Lab. Symposium No. 10) H.M.S.O., London (1960).
6. W. Ross Ashby, Computers and decision-making, *New Scientist*, 7, p. 746 1960, b).
7. W. Ross Ashby, The brain as regulator, *Nature, Lond.* **186**, p. 413 (1960, c).
8. W. Ross Ashby, *Design for a Brain; the Origin of Adaptive Behavior*, Wiley, New York, 2nd ed. (1960, d).
9. L. von Bertalanffy, An outline of general system theory, *Brit. J. Phil. Sci.* **1**, pp. 134–65 (1950).
10. R. L. Beurle, Properties of a mass of cells capable of regenerating pulses, *Proc. Roy. Soc.* **B240**, pp. 55–94 (1956).
11. W. R. Garner and W. J. McGill, The relation between information and variance analyses, *Psychometrika* **21**, pp. 219–28 (1956).
12. R. C. Jeffrey, Some recent simplifications of the theory of finite automata. Technical Report 219. Research Laboratory of Electronics, Massachusetts Institute of Technology (27 May 1959).
13. H. S. Jennings, *Behavior of the Lower Organisms*, New York (1906).
14. A. J. Lotka, *Elements of Physical Biology*, Williams & Wilkins, Baltimore (1925).
15. J. G. March and J. A. Simon, *Organizations*, Wiley, New York (1958).
16. K. H. Pribram, Fifteenth International Congress of Psychology, Brussels (1957).
17. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana (1949).
18. G. Sommerhoff, *Analytical Biology*, Oxford University Press, London (1950).

THE SELF-REPRODUCING SYSTEM*

W. ROSS ASHBY

High among the interesting phenomena of organization shown by life is that of reproduction. We are naturally led to ask. How can a system reproduce itself? And we go headlong into a semantic trap unless we proceed cautiously. In fact, the answer to the question, "How does the living organism reproduce itself?" is "It doesn't."

No organism reproduces *itself.* The only thing that ever has had such a claim made for it was the phoenix, of which we are told that there was only one, that it laid just one egg in its life, and that out of this egg came itself. What then *actually* happens when ordinary living organisms reproduce? We can describe the events with sufficient accuracy for our purpose here by saying:

(1) There is a matrix (a womb, a decaying piece of meat, a bacteriological culture tube perhaps).
(2) Into it is introduced a form (an ovum, a fly's egg, a bacterium perhaps).
(3) A complex dynamic interaction occurs between the two (in which the form may be quite lost).
(4) Eventually the process generates more forms, somewhat like the original one.

In this process we must notice the fundamental part played by the matrix. There is no question here of the ovum reproducing *itself.* What we see is the interaction between one small part of the whole and the remainder of the whole. Thus the outcome is a function of the *interaction* between two systems. The same is true of other forms. The bacterium needs a surrounding matrix which will supply

*The work on which this paper is based was supported by the Office of Naval Research, Contract N 62558-2404.

oxygen and food and accept the excretion of CO_2, *etc.* An *interaction* between the two then occurs such that forms somewhat resembling the initial bacterium eventually appear.

So, before we start to consider the question of the self-reproducing system we must recognize that *no organism is self-reproducing*. Further, we would do well to appreciate that Rosen [2] has recently shown that the idea of a self-reproducing automaton is logically self-contradictory. He uses an argument formally identical with that used by me [1] to show that a self-organizing system is, strictly, impossible. In each case the idea of a *self*-acting machine implies that a mapping must be able to alter itself—*i.e.*, that it is within its own domain. Mathematics and logic can do nothing with such a concept. It is in the same class as the fantasy that can see a man getting behind himself and pushing himself along.

I make these remarks, not in order to confuse or to obstruct, but simply to make sure, by clearing away sources of confusion, that we do really find the right approach to our topic. Though the adjective "self-reproducing" is highly objectionable semantically and logically, it does of course refer to a highly interesting process that we know well, even if we sometimes use inappropriate words to describe it.

I propose, then, to consider the question re-formulated thus:

A given system is such that, if there occurs within it a certain form (or property or pattern or recognizable quality generally), then a dynamic process occurs, involving the whole system, of such a nature that eventually we can recognize, in the system, further forms (or properties or patterns or qualities) closely similar to the original.

I ask what we can say about such systems.

CAN A MACHINE DO IT?

Having got the question into its proper form, we can now turn to the question whether a machine can possibly be self-reproducing. In a sense the question is pointless, because we know today that all questions of the type "Can a machine do it?" are to be answered "Yes." Nevertheless, as we are considering self-reproduction, a good deal more remains to be said in regard to the more practical details of the process. Our question then is: Does there exist a mechanism such that it acts like the matrix mentioned, in that, given a "form," the two together lead eventually to the production of other forms resembling the first?

I propose to answer the question largely by a display of actual examples, leaving the examples to speak for themselves.

The first example I would like to give is a formal demonstration in computer-like terms showing the possibility. Let us suppose a computer has only ten stores, numbered 0 to 9, each containing a two-digit decimal number, such as 72, 50, 07, or perhaps 00. The "laws" of this little world are as follows: Suppose it has just acted on store S−1. It moves to store S, takes the two digits in it, *a* and *b* say, multiplies them together, adds on 5 *and the store-number S*, takes the right-hand digit of the result, *c* say, and then writes the original two digits, *a* and *b*, into store c. It then moves on to the next store and repeats the process; and so on indefinitely.

At first sight, this "law" might seem to give just a muddle of numbers. At store No. 3 say, with 17 in the store, it multiplies together 1 and 7, adds 5 to the product, getting 12, adds the store number 3, getting 15, takes the right-hand digit, getting 5, and puts 17 into store 5. It then goes on to its next store, which is No. 4. There seems to be little remarkable in this process. On the other hand, a 28 in a store has a peculiar property. Suppose it is in store 7. $2 \times 8 = 16$, $16 + 5 = 21$, $21 + 7 = 28$, 28 gives 8, so 28 goes into store 8. When we work out the next step we find that 28 goes again into store 9, and so on into store after store. Thus, *once a 28 turns up in the store it spreads until it inhabits all the stores.* Thus the machine, with its program, is a dynamic matrix such that, if a "28" gets into it, the mutual interaction will lead to the production of more 28's. In this matrix, the 28 can be said to be self-reproducing.

The example just given is a formal demonstration of a process that meets the definition, but we can easily find examples that are more commonplace and more like what we find in the real world. Suppose, for instance, we have a number of nearly assembled screw drivers that lack only one screw for their completion. We also have many of the necessary screws. If now a single complete screw driver is provided, it can proceed to make more screw drivers. Thus we have again the basic situation of the matrix in which if one form is supplied a process is generated that results in the production of other examples of the same form.

On this example, the reader may object that a great deal of prefabrication has been postulated. This is true, of course, but it does not invalidate the argument, because the amount of prefabrication that occurs can vary over the widest limits without becoming atypical; and some prefabrication *has* to be allowed. After all, the liv-

32.1, *df* = 3; fluency, 4.2, *df* = 3; memory, 4.9, *df* = 4; hesitancy, 98.2, *df* = 4.) The results are given in Fig. 4 where ordinates are the per cent of *Ts* in a given category that are equilibrial.

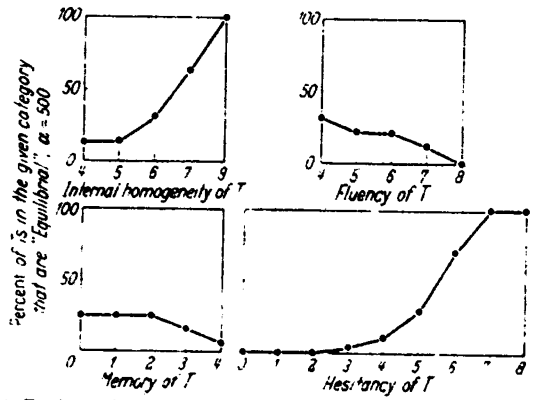


Fig. 4. Tendency for *Ts* to show primarily states of equilibrium. $\alpha = 500$

It was mentioned above that the *T* with the greatest hesitancy produces fields entirely filled with cycles of length one; it is for that reason equilibrial. On the other hand, the *T* with the lowest hesitancy is a doubled transformation. Therefore, the proportion of equilibrial *Ts* in each category of hesitancy, if the relationship is simple and well-behaved, should increase from zero to unity with increasing hesitancy. As can be seen in Fig. 4 such is the case, with the function in fact resembling the normal ogive. With respect to the other measures, it might be expected that greater internal homogeneity would lead to a greater production of equilibrial states due to a lack of variety in the element's signalling; and greater fluency, providing greater "informational transparency" in a system, might be expected to encourage a lesser tendency for the system to "stick" at a single state. (It is difficult to predict with confidence what effect increasing memory should have on the appearance of equilibrial states.)

Fluency, as expected, and memory curves exhibit trends toward an inverse relationship. The statistically significant measures internal homogeneity and hesitancy, as expected, both show increasing functions. The potency of hesitancy in the production of equilibrial states supports earlier theoretical work by Ashby (1960).

Modal Cycle Length and Run-In Length as Functions of Four Measures of Ts. An indication of the effect that the measures have on the typical cycle length can be gained from Fig. 5. Fig. 5 was obtained as follows: The modal cycle length was determined for each of the 198 *Ts* which showed cycles in sufficient numbers. The modal cycle lengths were then grouped into intervals according to the number of digits in the modal value. That is, each *T* was

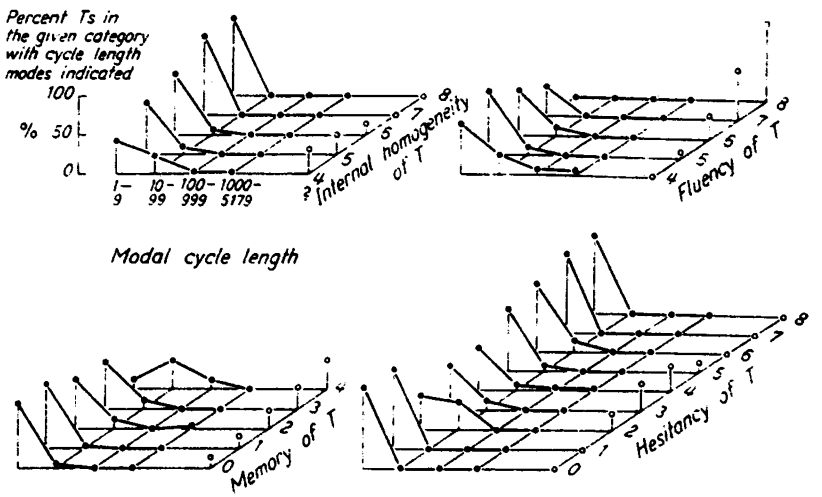


FIGURE 5. Modal (most frequent) cycle lengths as functions of four measures of *Ts*. (The question mark indicates the percent of *Ts* in the categories for which too few disclosures are shown to allow a reliable determination of modal values).

assigned a logarithmic interval describing its modal cycle length. For each of the four measures, *Ts* were grouped by their modal cycle length intervals and the category of the measure. The numbers of *Ts* with cycle lengths in the given intervals were compared to the total number of *Ts* in that category of the measure, expressed as percentages, and plotted as the ordinates of Fig. 5 with solid circles. The open circles in the figure give the percentages of *Ts* in the category with too few observed cycles to be classified by modal cycle lengths. Therefore, the open circles indicate the total extent to which the solid circles in a category are in doubt. (The open circles correspond closely to those of Fig. 3. The correspondence would be exact if two additional *Ts* had not been classed as to their cycle length modes.)

The fact that the distributions of modal cycle lengths are in doubt reduces the usefulness of Fig. 5, but the following trends are suggested: (1) increasing internal homogeneity tends to decrease cycle length, (2) extremely high memory appears to increase cycle length, (3) moderate hesitancy appears associated with larger cycles. (The high uncertainty associated with increasing fluency makes it difficult to pick out any trends with confidence.)

The same procedure used for cycle lengths applied to run-in lengths yields Fig. 6. It can be seen that except for memory, the trends found for cycle lengths are repeated in the run-in lengths, run-ins being, overall, longer than cycle lengths. Memory, however, affects cycle and run-in lengths differently

ing things that reproduce do not start as a gaseous mixture of raw elements.

(The same scale of "degrees of prefabrication" sometimes confuses the issue when a model maker claims that he has "made it all himself." This phrase cannot be taken in any absolute sense. If it were to be taken literally, the model maker would first have to make all the screws that he used, but before that he must have made the metal rods from which the screws were produced, then he must have found the ores out of which the metal was made, and so on. As there is practically no limit to this going backward, the rule that a model maker "must make it all himself" must be accompanied by some essentially arbitrary line stating how much prefabrication is allowed.)

The two examples given so far showed only reproduction at one step. Living organisms repeat reproduction: fathers breed sons, who breed grandsons, who breed great-grandsons, and so on. This possibility of extended reproduction simply depends on the scale of the *matrix*. It can be present or absent without appreciably affecting the fundamentals of the process.

FURTHER EXAMPLES

The subject of self-reproduction is usually discussed on far too restricted a basis of facts. These tend to be on the one hand simply the living organisms, and on the other hand machines of the most rudimentary type, such as the watch and the motor car. In order to give our consideration more range, let us consider some further examples. Those I give below will be found to be sometimes unorthodox but every one of them, I claim, does accord with the basic definition—that the bringing together of the first form and matrix leads to the production of later forms similar to the first.

Example 3. A factory cannot start producing because the power is not switched on. The only thing that can switch the power on is a spanner (wrench) of a certain type. The factory's job is to produce spanners of that type.

Example 4. A machine that vibrates very heavily when it is switched on can be started by a switch that is very easily thrown on by vibration. Such a system, if at rest and then given a heavy vibration, is liable to go on producing further heavy vibrations. Thus the form "vibration," in this matrix, is self-reproducing.

Example 5. Two countries, A and B, were at war. B discovered that country A was a dictatorship so intense that every document

bearing the dictator's initials (X.Y.Z.) had to be obeyed. Country B took advantage of this and ruined A's administration by bombing A with pieces of paper bearing the message: "Make ten copies of this sheet, with the initials, and send to your associates. X.Y.Z." In such a matrix, such a form is self-reproducing.

Example 6. A number of chameleons are watching one another, each affected by the colors it sees around it. Should one chameleon go dark it will increase the probability of "darkness" appearing around it. In this matrix, the property "darkness" tends to be self-reproducing.

Example 7. In a computer, if the order 0101010 should mean "type 0101010 into five other stores taken at random," then in this matrix the form 0101010 is self-reproducing.

Example 8. A computer has single digit decimal numbers in its various stores. It is programed so that it picks out a pair of numbers at random, multiplies them together, and puts the right-hand digit into the first store. In this condition, as any zero forces another zero to be stored, the zero is self-reproducing.

Example 9. Around any unstable equilibrium, any unit of deviation is apt to be self-reproducing as the trajectory moves further and further away from the point of unstable equilibrium. Thus, if a river in a flat valley happens to be straight, the occurrence of one meander tends to lead to the production of yet other meanders. Thus in this matrix the form "meander" is self-reproducing.

Example 10. A similar example occurs when a ripple occurs in a soft roadway. Under the repeated impact of wheels, the appearance of one tends to lead to the appearance of others. In this matrix, "ripple" is self-reproducing.

Example 11. (Due to Dr. Beurle) A cow prefers to tread down into a hole rather than up onto a ridge. So, if cows go along a path repeatedly, a hollow at one point tends to be followed by excessive wear at one cow's pace further on, and thus by a second hollow. And this tends to be followed by yet another at one pace further on. Thus, in this matrix, "hollow" is self-reproducing.

Example 12. Well known in chemistry is the phenomenon of "autocatalysis." In this class is the dissociation of ethyl acetate (in water) into acetic acid and alcohol. Here, of course, the dissociation is occurring steadily in any case, but the first dissociation that produces the acid increases the *rate* of the later dissociations. So, in this matrix, the appearance of one molecule of acetic acid

when very high: cycles tend to be increased in length, while run-ins tend to be decreased. Very high memory evidently provides these systems with an increased capability for complex terminal behavior relatively soon after the systems start operating from a random state.

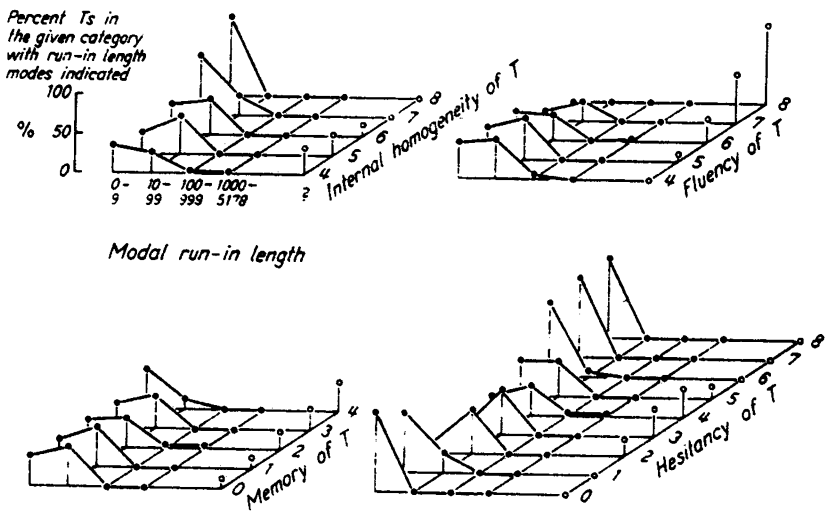


FIGURE 6. Modal (most frequent) run-in lengths as functions of four measures of *Ts*. (The question mark indicates the percent of *Ts* in the categories for which too few disclosures are shown to allow a reliable determination of modal values).

Summary. The relationships between measures and behaviors can be roughly summarized as follows (Table 2).

Table 2. *A Summary of the Effects of Increasing Certain Measures of Ts on Behavior*

Measure	Behavior			
	Disclosure Lengths	Per cent of Equilibrial Ts	Cycle Lengths	Run-In Lengths
Internal Homogeneity	↘	↗	↘	↘
Fluency	↗	?	?	?
Memory	?	?	→	↘
Hesitancy	↗↘	↗	↗↘	↗↘

5.4 Discussion

On Empirical Study of More Complex Systems. The regularity of the relationships between measures and behavior seen here suggests that empirical study of other interesting families of systems may be feasible, even though the numbers of transformations involved may preclude the exhaustive testing of whole families. Evidently, if behaviors and transformations are not specified too rigidly, abstract systems such as those examined here do not always show the discontinuities that might be expected of them.

On "Biological Clocks". While it is beyond the scope of the present paper to go into the question of how parallels can be usefully drawn between the abstract systems examined and naturally occurring systems, there are some interesting speculations that can be made. In the case of a multiple *T*, that is, a *T* for which the five systems which use that *T* all show cycle lengths with a common (non-unity) periodicity, the *T* is, of course, with respect to the common periodicity of the cycle length, pragmatically structure-insensitive. Therefore, for a multiple *T* it is reasonable that, if not all, at least a large number of systems using this *T* have cycle lengths that are multiples of that same factor. Were systems made of the appropriate parts to be found occurring naturally, and especially if the systems' structures were due to irregular, non-systematic circumstances, then it would not be uncommon for these natural systems, in constant surroundings, to show similar behavioral rhythms. Having previously noted the frequency with which rhythmic behavior is shown, a further implication is that "biological clocks" may be more the rule than the exception, and that when such rhythms are regularly found to occur, it is not to be concluded that there exist clock-mechanisms responsible for the rhythms which are spacially distinct and localizable. Like the periodicities of the systems studied, natural rhythms of real things may be due to spacially extended system-properties.

What Happens in Structurally Larger Systems? It is of particular interest to consider what should happen to the behavior of the systems of the present study if the number of their elements were increased. There is an interesting approach to this question suggested by the results. This approach would assume that because certain behavioral characteristics are observed to hold over a set of randomly sampled structures, the behaviors are therefore pragmatically independent of structural changes generally, and *independent of the number of elements in the systems in particular.* It would therefore be predicted that equilibrial *Ts* will remain equilibrial, doubled *Ts* will remain doubled, and multiple *Ts* will continue to be multiples of the same factors, no matter how large the number of elements in their systems. That such a prediction may not be entirely speculative is indicated by one piece of evidence. A trajectory segment run for a 1000-element

tends to encourage the appearance of further molecules of the same type.

Example 13. In the previous example the form has been a material entity, but the form may equally well be a pattern. All that is necessary is that the entity, whatever it is, shall be unambiguously recognizable. In a supersaturated solution, for instance, the molecular arrangement that one calls "crystalline" is self-reproducing, in the sense that in this matrix, the introduction of one crystalline form leads to the production of further similar forms.

Example 14. With a community of sufficiently credulous type as matrix, the introduction of one "chain letter" is likely to lead to the production of further such forms.

Example 15. In another community of suitable type as matrix, one person taking up a particular hobby (as form) is likely to be followed by the hobby being taken up by other people.

Example 16. Finally, I can mention the fact that the occurrence of one yawn is likely to be followed by further occurrences of similar forms. In this matrix, the form "yawn" is self-reproducing.

REPRODUCTION AS A SPECIALIZED ADAPTATION

After these examples we can now approach the subject more realistically. To see more clearly how special this process of reproduction is, we should appreciate that reproduction is not something that belongs to living organisms by some miraculous linkage, but is simply a specialized means of adaptation of a specialized class of disturbances. The point is that the terrestrial environments that organisms have faced since the dawn of creation have certain specialized properties that are not easily noticed until one contrasts them with the completely nonspecialized processes that can exist inside a computer. Chief among these terrestrial properties is the extremely common rule that if two things are far apart they tend to have practically no effect on one another. Doubtless there are exceptions, but this rule holds over the majority of events. What this means is that when disturbances or dangers come to an organism, they tend to strike locally. Perhaps the clearest example would be seen if the earth had no atmosphere so that the organisms on it were subject to a continuous rain of small shotlike particles traveling at very high speeds. Under such a rain the threat by each particle is local, so that a living form much increases its chance of survival if replicates of the form are made and dispersed. The rule of

course is of extremely wide applicability. Banks that may have a fire at one place make copies of their records and disperse them. If a computing machine were liable to sudden faults occurring at random places, there would be advantage in copying off important numbers at various stages in the calculation so as to have dispersed replicates. Thus, the process of reproduction should be seen in its proper relation to other complex dynamic processes as simply a specialized form of adaptation against a special class of disturbances. It is all that and nothing more. Should the disturbances not be localized there is no advantage in reproduction. Suppose, for instance, that the only threat to a species was the arrival of a new virus, that was either overwhelmingly lethal or merely slightly disturbing. Under such conditions the species would gain nothing by having many distinct individuals. The same phenomenon can be seen in industry. If an industry is affected by economic circumstances or by new laws, so that either all the companies in it survive, or all fail, then there is no advantage in the multiplicity of companies; a monopoly can be as well adapted as a multiplicity of small companies.

FUNDAMENTAL THEORY

After this survey we have at least reached a point where we can see "reproduction" in its proper nature in relation to the logic of mechanism. We see it simply as an adaptation to a particular class of disturbances. This means that it is at once subject to the theoretical formulations that Sommerhoff [3] has displayed so decisively. The fact that it is an adaptation means that we are dealing essentially with an invariant of some dynamic process. This means that we can get a new start, appropriate to the new logic of mechanism, that will on the one hand display its inner logic clearly, and on the other hand state the process in a form ready to be taken over by machine programing or in any related process. We start then with the fundamental concept that the dynamic process is properly defined by first naming the set S of states of the system and then the mapping f of that set into itself which corresponds to the dynamic drive of the system. *Reproduction is then one of the invariants that holds over the compound of this system and a set of disturbances that act locally.* If then f is such that some parts within the whole are affected individually, "reproduction" is simply a process by which these parts are invariant under the change-inducing actions of the dynamic drive f.

II.

INFORMATION FLOWS IN SYSTEMS

It must be emphasized that reproduction, though seeming a sharply defined process in living organisms, is really a concept of such generality that precise definition is necessary in all cases if it is to be clear what we are speaking of. Thus, in a sense every state of equilibrium reproduces itself; for if $f(x) = x$, then the processes f of the machine so act on x that at a moment later we have x again. This is exactly the case of the phoenix. It is also "self-reproduction" of a type so basic as to be uninteresting, but this is merely the beginning. It serves as a warning to remind us that processes of self-reproduction can occur, in generalized dynamic systems, in generalized forms that *far exceed in variety and conceptual content anything seen in the biological world.* Because they are nonbiological the biologist will hesitate to call them reproducing, but the logician, having given the definition and being forced to stick to it, can find no reason for denying the title to them. What we have in general is a set of parts, over some few of which a property P is indentifiable. This property P, if the concept is to be useful, must be meaningful at various *places* over the system. Then we show that "self-reproduction of P" holds in this system if along any trajectory the occurrence of P is followed, at the states later in the trajectory, by their having larger values for the variable "number of P's present."

It should be noted that because self-reproduction is an adaptation, which demands (as Sommerhoff has shown) a relation between organism and environment, and because the property P must be countable in its occurrences over the system, we *must* be dealing with a system that is seen as composed of parts. I mention this because an important new development in the study of dynamics consists of treating systems actually as a whole, the parts being nowhere considered. This new approach cannot be used in the study of reproduction because, as I have just said, the concept of reproduction *demands* that we consider the system as composed of parts.

The new point of view which sees reproduction simply as a property that may hold over a trajectory at once shows the proper position of an interesting extension of the concept. Reproduction, as I said, is a form of invariant. In general, invariants are either a state of equilibrium or a cycle. So far, we have considered only the equilibria, but an equally important consideration is the cycle. Here we reach the case that would have to be described by saying that A reproduces B, then B reproduces C, and then C reproduces A. Such a cycle is of course extremely common in the biological world. Not only are there the quite complicated cycles of forms through the egg, pupa, imago, and so on that the insects go through, there is of course also the simple fact that human reproduction itself goes regularly round the cycle: ovum, infant, child, adult, ovum, and so on.

A further clarification of the theory of the subject can be made. Let us define "reproduction" as occurring when the occurrence of a property increases the probability that that property will again occur elsewhere; this of course is positive reproduction. We can just as easily consider "negative" reproduction, when the occurrence of a property *decreases* the probability that the property will occur elsewhere. Examples of this do not appear to be common. We can of course at once invent such a system on a general-purpose computer; such "negative reproduction" would occur if, say, the instruction 00000 were to mean "replace all zeroes by ones." I have found so far only one example in real systems—namely, if, under electrodeposition, a whisker of metal grows toward the electrode, the chance of another whisker growing nearby is diminished. Thus "whiskers" have a negative net reproduction.

This observation gives us a clear lead on the question: Will self-reproducing forms be common or rare in large dynamic systems? The *negatively* self-reproducing forms clearly have little tendency to be obtrusive—they are automatically self-eliminating. Quite otherwise is it with the positively self-reproducing forms; for now, if the system contains a single form that is *positively* self-reproducing, that form will press forward toward full occupation of the system.

Suppose now we make the natural assumption that the larger the system, if assembled partly at random, the larger will be the number of forms possible within it. Add to this the fact that if any *one* is self-reproducing, then self-reproducing forms will fill the system, and we see that there is good reason to support the statement that *all sufficiently large systems will become filled with self-reproducing forms.*

This fact may well dominate the design of large self-organizing systems, forcing the designer to devote much attention to the question: "What self-reproducing forms are likely to develop in my system?" just as designers of dynamic systems today have to devote much attention to the prevention of simple instabilities.

INFORMATION FLOWS IN SYSTEMS

INTRODUCTION

The study of a large complex systems such as the brain, a city, a national economy, or the like is exceedingly difficult under any but the most favorable circumstances. Under special conditions - identity of all the system components, linearity of the system relationship, decomposability of the system into loosely coupled subsystems, the applicability of simplifying homomorphisms, etc. - some headway can be made, but in general the sheer quantity of information in the behavior of complex systems threatens to overwhelm and bewilder the investigator, even one armed with a computer. Ashby believed that information theory, generalized to N dimensions, would become an effective tool for the study of such systems, and indeed it has been used and developed by several researchers for that purpose since the time of his work. In "Principles of the Self-Organizing System" Ashby noted that the organization of a system is related to constraints it exhibits. These constraints can be measured by information theory (or uncertainty analysis, a nearly synonymous term) which therefore represents a tool for the measurement of organization. The basic idea is that if variables are related, a constraint exists between them which can be measured with the quantities of information theory; no "transmission", no relation. In this context information theory is invoked simply as a statistical tool for the measurement of multivariable correlations, though without the need for metric variables which the correlation coefficient, the analysis of variance, and related statistical devices require. The gain in using information theory in this way is that one may study organization in systems too complex for detailed comprehensive study; the loss is that all details, content, and meaning of the relationships are lost, leaving only "quantity of relationships" as the outcome of the investigation. In more recent work [102, 104, 111] the theory has in fact proved to be a very useful tool for the study of structure in multivariable systems, but it is not yet practical for computations unless the number of variables is quite modest (less than 25, say). In the absense of simplicities such as decomposability, Ashby's expectation that information theory would allow study of huge systems - he mentions 10 billion elements - seems forever doomed on statistical grounds alone, and on computational grounds besides.

In "Setting Goals in Cybernetic Systems" another sort of information "flow" is considered - the quantities of information involved in design, and in setting goals. Insofar as design can be understood as the appropriate selection of one design from a collection, the amount of selection can be measured and is subject to the laws of information theory. This is a unique perspective; it is much more common to think of design as creation, while Ashby, characteristically viewing it from the inverse perspective, sees it as selection.

In "Information Flows within Coordinated Systems" Ashby illustrates a basic methodology, in which the constraint within a system of four variables is partitioned in various ways using the entropy and transmission measures, the point of the paper being that for a certain degree of coordination a certain minimum of information "flow" is required. In "Information Processing in Everyday Human Activity" he attempts to estimate how much information is required, at a minimum, to perform a simple household task. Both papers may leave the reader with an uncomfortable feeling that the

numerical results are directly dependent upon rather arbitrary numerical assumptions made at the outset. This is quite so but is merely a reflection and verification of Ashby's observation, in the paper on self-organizing systems, that constraint and organization are not absolute properties of a system but have to do with the relation between the system and the observer. The papers assume this relationship and then illustrate a method, and it is the latter which represents the main thrust of these two publications.

"Measuring the Internal Information Exchange in a System" provides certain key identities in information theory, and "Two Tables..." adds more. The former paper gives explanations and interpretations for the quantities involved. At the time, Ashby was very keen on the Q terms, called "interactions," since he believed them to be intimately associated with a system's inherent complexity which, of course, he was always interested in quantifying, understanding and unravelling. As the paper shows there is considerable justification for that hope. Subsequent work mostly by Klaus Krippendorff [112] has shown, however, that Q has not borne out these initial hopes and is seriously misleading as an indicator of what we understand intuitively by the word interaction. The problem is mainly that Q arises from two opposing causes of opposite sign, one representing "genuine" interaction and one a statistical effect; the latter pollutes Q and makes it unusable as an indicator of the former. Krippendorff has contrived a much better indicator and the reader should be aware of that recent development so as to avoid uncritical acceptance of Ashby's paper - which has, nevertheless, been very influential.

Setting Goals in Cybernetic Systems

W. ROSS ASHBY

Borden Neurological Institute
England

Getting clear the matter of goals is of the first importance in cybernetics, for most applications of cybernetics start with someone saying "I want . . ". Here I am thinking of cybernetics not as a way of explaining things, but as a new science and technical method enabling us to tackle practical problems that would otherwise defeat us by their complexity. Coordinating the traffic around an airport, stabilizing the flows of money between international banks, normalizing the composition of the blood in a patient without kidneys—all these must start with the question "What do you want?" The process itself will end at the goal: the cybernetician's thoughts must start there.

What is a goal? We all know something about it, and a child of three frequently says "I want . . ." so we all start with a personal awareness of intention, purpose, need, desire. But when we ask how a machine can have a desire, we find ourselves in difficulties. The difficulty becomes even greater if the system that is to have the goal is not even one machine but a mixture of machines and men, with the goal involving

W. ROSS ASHBY

SUMMARY

Reproduction has, in the past, usually been thought of as exclusively biological, and as requiring very special conditions for its achievement. The truth is quite otherwise: it is a phenomenon of the widest range, tending to occur in all dynamic systems, if sufficiently complex.

The brain may well use this tendency (for self-reproducing forms to occur) as part of its normal higher processes. The designer of large self-organizing systems will encounter the property as a major factor, as soon as he designs systems that are really large and self-organizing.

REFERENCES

1. W. Ross Ashby, "Principles of the Self-organizing System," Symposium on Self-organizing Systems, University of Illinois, June 7-10, 1960. Pergamon Press, 1962.
2. R. Rosen, "On a Logical Paradox Implicit in the Notion of a Self-reproducing Automaton," Bull. Math. Biophysics, Vol. 21, pp. 387-394, 1959.
3. G. Sommerhoff, "Analytical Biology," Oxford University Press, London, 1950.

(Reprinted from Nature, Vol. 196, No. 4854, pp. 561-562, November 10, 1962)

INSTABILITY OF PULSE ACTIVITY IN A NET WITH THRESHOLD

By PROFS. W.R. ASHBY, H. Von FOERSTER and C.C. WALKER

Electrical Engineering Research Laboratory, University of Illinois, Urbana

For half a century, threshold has been known to be important in the activities of nerve cell and synapse, but little is known of the general properties that so ubiquitous a feature must impose on the system's behavior in the large. Beurle[1], in his investigations of the transmission of waves of activity over a conducting net, noticed that a net with threshold would have a marked tendency to be unstable, but his assumptions were complex, and the origin of the instability is not easily identified. Here we propose to show that an instability, similar to his, can readily be traced from a simpler origin.

The instability we refer to would soon show itself if one made an artificial nerve-net with unspecified, rich connectivity, with threshold at the junctions, and then attempted, as the net transmitted pulses, to keep the net's activity at a moderate value. The net would tend to run either down to a state of total inactivity, from which one could scarcely get it to stir, or up to a state of total activity, which would lesson only at the onset of exhaustion or some other extraneous factor.

The origin of this instability can be traced as follows. Suppose the whole net is composed of a large number of interconnected units which handle information represented everywhere by identical pulses of some physical activity. Each of these units has n identifiable inputs. A unit 'fires in time interval t (to $t + \Lambda t$) – emits a pulse of duration $\delta t \ll \Lambda t$ – if and only if at least θ of the inputs have received a pulse within the preceding time interval ($t - \Lambda t$ to t. The following argument is applicable both to nets with essentially one-directional information flow (from a network-input to a network-output) and also to nets with rich internal cross-connections (provided none of the feed-back loops is short).

We shall describe the activity on the inputs of a particular unit by first associating a probability p with the occurrence of a pulse in a particular time interval Λt on a specified input. We define:

$$p = \lim_{t \to \infty} \left(\frac{N}{L} \cdot \frac{\Delta t}{t} \right) \qquad (1)$$

where N is the number of pulses counted in time t which have passed through a sufficiently large bundle of L randomly selected inputs. From this we have the

only the whole, not the parts. How can such a system have a purpose or desire?

The solution of this first difficulty, I suggest, is to do what the psychologists have done for a century: drop the introspectional aspect and turn to the aspect of *behavior*. Stop asking "Does this system feel a want?" and ask instead "How does it behave?"

Those whose knowledge of these matters is mainly introspectional may well hesitate to abandon their main source of information. But a century's work in psychology has shown that the introspectional approach, though vivid and apparently unquestionable, is in fact grossly unreliable. Look, for instance, at a piece of uncolored (white) paper: if anything is obvious and trustworthy, it is that there is no red present. Yet the physicists have convinced us that what we see is not the paper but a message from the retina saying that the three primary colors "balance." The introspective viewer sees only his own retina, not beyond it.

A report based on introspection is in fact simply the output of the brain's final, verbalizing stage. Such report can give only a coded version of what is happening earlier in the processes; to take a coding literally is an evident mistake. Psychoanalytic studies have shown in innumerable cases how mistaken a person can be when he describes his own motives or goals. Briefly, the introspectional approach, in science, has so far proved to be either just useless or positively misleading.

But if a goal is not a want, what is it? Ever since McDougall,[1] psychologists have understood that it can be treated equivalently as *a way of behaving*. "Take a timid animal," he wrote, "such as a guinea-pig, from its hole or nest, and put it upon the grass plot. Instead of remaining at rest, it runs back to its hole; push it in any other direction, and, as soon as you withdraw your hand, it turns back towards its hole." Just the same behavior is characteristic of the missile that persists in going toward a source of infrared rays, that reasserts its direction if diverted, and that will change direction if the source moves.

The experiences of a century in psychology and of thirty years in automatic control systems have shown that, for *practical* purposes, we can achieve clarity by replacing the idea of a felt need with the idea of a *focus* in a stable dynamic system.

When the system is as simple, essentially, as a missile that "seeks" infrared rays, the thesis will probably not be disputed. But what of the more complex? What of natural evolution, for instance, with organisms apparently developing their own goals? What of man? Can he not choose his own goals? Cannot the cybernetician make a machine that can choose its own goal?

To get our ideas untangled, let us first take the case of natural evolution, since here the facts are, today, beyond dispute. This case is that of a planet, subject both to unchanging laws (such as those of gravity, optics, hydrodynamics) and, over all of 10^{10} years, to a constant energy input. Through all this time, photons of the visible wavelengths have poured in and have left at infrared lengths to the night sky, in a steady flux of about 10^{20} ergs per day. This unceasing flow of free energy has kept the molecules on the planet in a mild but unceasing turmoil, during which the less stable combinations have incessantly been superseded by the more stable. Today, after 10^{10} years, what remains is mostly of extremely high stability, ranging from such minerals as granite to such dynamically stable forms as the mammal, a form older than the Alps and the survivor of several ice ages.

Looked at in this way, natural evolution and the emergence of forms such as the guinea pig, with its well-developed goal, are in no way unusual: they merely exemplify the fact that almost all state-determined systems tend to preferred regions. It is the exceptional system that does not show such preferred regions. Thus, the continuous state-determined system with equations

$$x_i = \phi_i(x_1, \ldots, x_n) \qquad (i = 1, \ldots, n)$$

shows *non*convergence only where div ϕ has the special value

probability p_i that precisely i inputs on a particular unit receive a pulse in the time interval Δt:

$$p_i = \binom{n}{i} p^i (1-p)^{n-i} \tag{2}$$

Consequently, the probability p' that at least θ inputs are active in this time interval, that is, that the unit fires, is given by the cumulative binomial probability function[2]:

$$p' = \sum_{i=\theta}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(\theta-1)!\,(n-\theta)!} \int_0^p x^{\theta-1} (1-x)^{n-\theta}\, dx \tag{3}$$

The instability arises when we allow the resulting pulse frequency $f' = p'/\Delta t$ to be itself the generator of a further frequency $f'' = p''/\Delta t$, and so on. Under these conditions we ask what will happen in the net as time goes on. Equilibrium is reached if the pulse activity on each input equals the pulse activity on each output, that is, if:

$$p' = p = p^* \tag{4}$$

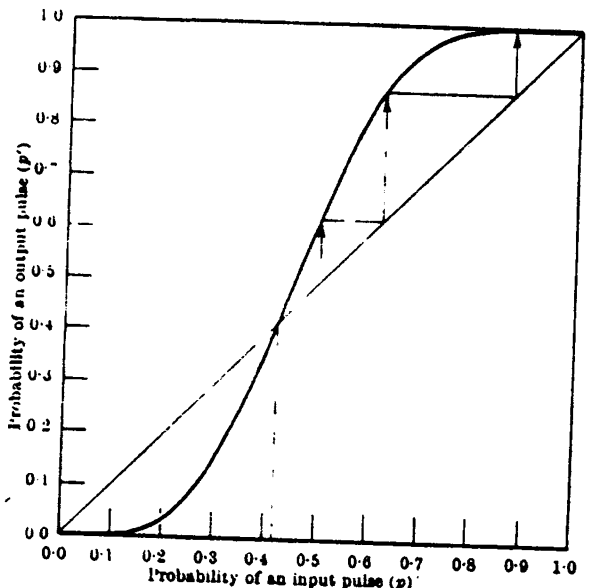


Fig. 1. Probability of an output pulse, as a function of the probability of an input pulse (threshold 5, 10 inputs)

The equilibrium will be stable at the activity p* if an increased activity at the input leads to a lesser activity at the output; that is, if:

$$\left|\left(\frac{\partial p'}{\partial p}\right)_{p^*}\right| < 1 \tag{5}$$

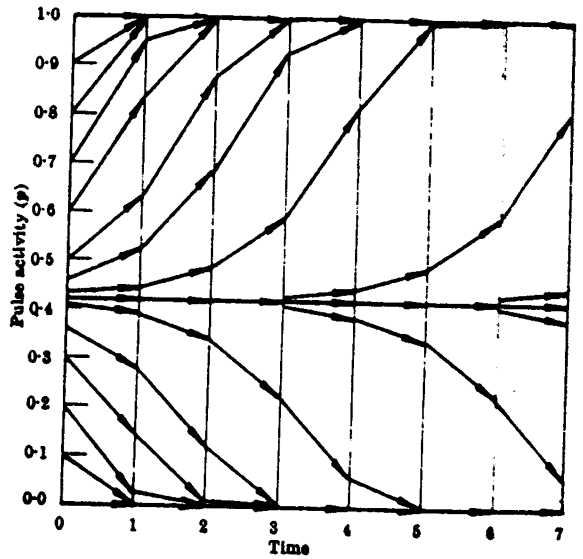


Fig. 2. How pulse activity changes with time, from various initial activities (threshold 5, 10 inputs)

However, differentiating expression (3) once and twice with respect to p shows that for all cases where solutions for (4) with $0 < p^* < 1$, and $n > 1$, exist, $p'(p)$ starts at the origin with zero slope and exhibits a single inflexion at the point:

$$p = \frac{\theta - 1}{n - 1} \tag{6}$$

Thus the stability criterion (5) is satisfied only where $p^* = 0$ or $p^* = 1$.

Fig. 1 illustrates the stiuation in a simple case where p' is plotted against p for $n = 10$ and $\theta = 5$, using (3). The equilibria p^* are the three points where $p'(p)$ intersects $p' = p$: 0, 0.42 and 1. Those at 0 and 1 are stable; that at 0.42 is unstable in the sense that the slightest perturbation from this value is followed by the system going to one of its extreme values. (In. Fig. 1 the change in p from an initial value of 0.5 is indicated by the stairway.) Fig. 2 shows how various perturbations provoke a runaway to an extreme value.

It seems, therefore, that the more richly organized regions of the brain offer us something of a paradox. They use threshold intensively, but usually transmit impulses at some moderate frequency, seldom passing in physiological conditions into total inactivity or maximal excitation. Evidently there must exist factors or mechanisms for stability which do not rely on fixed threshold alone.

One such mechanism that offers itself readily is the non-linear dependency of threshold on output that is sometimes referred to as 'inhibition by depletion'. If this dependency is of the form:

$$\theta = \mu^{-1/\lambda} (p')^m$$

with:

$$1 < \frac{1}{\lambda} \leqslant m$$

0.[2] Take ø's at random, and the system resulting is almost certain to show stability at preferred regions, within which it will behave in the goal-seeking way.

"Making a system that seeks a goal" is thus trivially easy: one forms a state-determined dynamic system at random (e.g., let it be specified by the spins of a coin). One is then almost certain to have a system that, like a guinea pig, will show that it is actively goal-seeking for some preferred state. It is true that the preferred state may well be meaningless or useless to the designer, but we should notice at this point that getting a machine to have *some* goal is no problem at all.

Achieving an Assigned Goal

Having disposed of this pseudoproblem, we can now consider the real, and difficult, problem. This arises when the designer not only wants the system to be goal-seeking but also wants it to seek some goal already specified. Air-traffic control systems are required to make collisions minimal, not maximal, and a physiological stabilizer of blood composition must have as its goal just those values that the human finds normal. Here the majority of stable states that might occur on random assembly are not acceptable.

When the system is small (designing a room thermostat, say), the designer needs no further general theory; he goes straight to the particular details. But when the system is of "cybernetic" size, he may still be uncertain of the next steps. I want to suggest in this paper that the general nature of the situation can be made much clearer if we apply what is already known in information theory.

The situation is most evident when a designer faces a heap of components from which he is going to construct his machine (but it is equally so when he faces a sheet of paper on which he will write a program). The point is that by his *selection* of the assembly or program he wants, from the set that includes both what he wants and what he does not, he *transmits a message* to the end product, and all the laws of

communication are applicable. To design a thermostat that will hold 72°F is to transmit the value "72" to the machine. Consider the less trivial example of the designer who wants to allot values −2, −1, 0, +1, or +2 to the four coefficients a, b, c, d in

$$\dot{x} = ax + by$$
$$\dot{y} = cx + dy$$

so that the system will be stable. In this case the quantity of information that must be transmitted from designer to system is calculable. 5^4 types are possible, of which 114 have the real parts of their latent roots both negative. In the worst case (if all values are equiprobable), the selection implies a transmission of $\log_2 (625/114)$ bits, i.e., just under 2½ bits. Thus, in this example, the channel represented by Fig. 1 *must* be able to transmit at least 2½ bits (per act of design).

Fig. 1.

The example is trivial: what matters is whether the principle is sound. If so, it will give us a deeper insight into problems that are anything but trivial.

Before we go further, however, we must notice a matter that might easily confuse us. Suppose some complex regulator accepts m inputs X_i ($i = 1, \ldots, m$), data about aircraft at an airport perhaps, and emits orders, values on n variables Y_j ($j = 1, \ldots, n$), to the aircraft. The designer is then asked to design it so as to be a "good" traffic controller. The basic situation may be represented as in Fig. 2. How the outputs

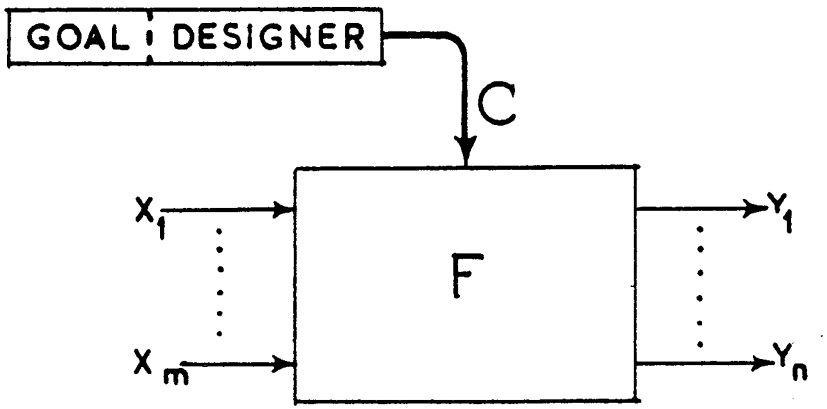


Fig. 2.

depend on the inputs is a relation F, the "transfer function" (in a general sense) of the system. The designer's task is to act so that the desired goal, selected from *all* the events (good and bad) that can occur at an airport, is transmitted to, and acts effectively as, a "good" F. This good F, it should be noted, comes from the set of all possible F's (not from the set of Y-values), so the transmission implied by the selection, along the channel C, is essentially independent of the transmission from X to Y. In this paper our attention will be focused on the quantities transmitted through such a channel as C.

The situation perhaps can be made intuitively more vivid if the designer's task be formulated as that of conveying to the heap of components what he means by the phrase "a successful airport," in the operational mode of getting the heap, when assembled, to separate the "good" set from its complement. Similarly, the problem of designing a pattern-recognizer (for genuine dollar bills, say) may be regarded generally as one of trying to tell the machine, in its *un*developed state, what is meant by "genuine, dollar, and bill." Again, this flow of design-information is essentially distinct from the flow that occurs later when the finalized machine scans an actual piece of paper and emits a verdict. Similarly, in information retrieval (getting from a library, say, the documents relevant

to "social feedbacks in universities"), the difficulty can be regarded as essentially one of getting across to the machine what is meant by "society, feedback, and university." In general, thinking of the design process as one of telling the machine what you want helps to make more evident the flow of information that is intrinsic to the act of designing.

Sponsor and Designer

If this thesis is granted, we are led inescapably to the deduction that exactly the same situation exists at the prior stage, represented in Fig. 3.

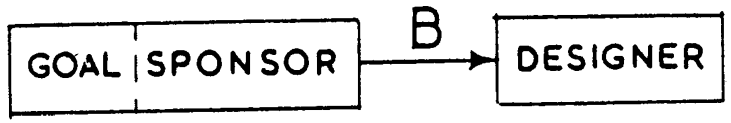


Fig. 3.

Whence came the particular goal that the designer transmitted to F? Rarely is the designer himself the originator; more commonly, as in the example of the air-traffic control system, the goal comes from what I will call generically the *sponsor*. Again I argue, as before, that effects must have causes. Selection among the F's can be attributed to prior selection among the designer's possible goals. And this selection must be attributed to prior selection among the *sponsor's* possible goals. Further, it is axiomatic through information theory (though seldom stated explicitly) that complex effects, as "messages received," require at least as much complex causation as "messages sent." By as much we mean either the number of bits or, more simply, the number of possible messages. So with this axiom we can assert that there must be sufficient transmission, from sponsor to designer, for the selection made by the sponsor (among *his* possible goals) to reach the designer.

Often this requirement is so obvious that these remarks would appear to be unnecessary. But it also often happens that the requirement is not, in fact, adequately met. It is more than likely that at this meeting there are some designers who have been given nothing like sufficient information (channel B) about the sponsor's goal, yet who are expected to achieve high and appropriate selection in the system F. Sponsors must learn enough elementary cybernetics to appreciate when they are asking the impossible. The sponsor must learn, in fact, that he is as subject to Conant's First Theorem[3] as every other would-be regulator. He is a regulator insofar as, with some goal in mind, he is trying to bring the designer, who might design all sorts of things, to accept his particular goal (for transmission to the system F). If the designer misunderstands and designs for G′ instead of G, the sponsor must correct him and keep acting until the designer has the sponsor's G as his guide. This is an act of regulation and so is wholly subject to Conant's Theorem, which asserts that transmission of the corresponding quantity of information (through B) is absolutely unavoidable. Armed with this theorem, the designer can *demand* that the sponsor transmit sufficiently.

The only worker in this field who seems to have clearly understood and stated these requirements was the elder von Moltke, who, in 1858, founded and organized the German General Staff. The key principle he gave it was that of the "directive." In this method the senior (corresponding to the sponsor here) gives no orders to the junior. Orders were replaced by the rule: the senior shall take all necessary time to explain to the junior what the senior wants *from the senior's point of view;* then he shall leave the junior free to use all personal initiative and local knowledge to achieve the goal. Von Moltke evidently had the clearest intuitive understanding of these quantitative flows.

Sometimes, however, the sponsor need not transmit everything. What is necessary is that *to* the designer shall come sufficient information (as determining factors) to enable him to select *his* goal adequately. But the determination need not

all come from the sponsor. A sponsor may well specify so much and then say "I delegate the rest to ————————." He may delegate it to a junior, who must then supply the rest. He may delegate it to the designer himself. The designer may himself delegate further, to the spins of a coin, perhaps. But in all cases *the total of determination coming to the designer must be not less than the quantity required for the selection of his goal.*

Such supplementation may take various forms. One well-known form occurs when the genes set the goal for the living organism. But, in the higher organisms, the goal is too complex to be transmitted through the genetic channel, so a part of the determination is delegated to the environment. Says the gene-structure to the kitten, "I have told you something about mice—now go out and get the finer details from the mice themselves." We call such supplementation "learning." And a learning machine is simply any machine that is specified only partly by its designer, who delegates the remainder of the specification to some "teaching" environment.

Quantities of Information

Once these general principles governing goals are clear, the rest is a matter of special techniques in special cases. But there is one aspect that I would like to mention, as I believe it to be of central importance. It concerns the case in which the goal is of very high complexity, as in artificial intelligence, high-order pattern recognition, and high-order regulation.

Without stopping to examine closely the idea of "complexity," I shall assume that a complex goal is one which has many parts and in which the required relations between the parts show high *conditionality* of part on part: when, that is, the whole goal is a nonreducible function of many variables. I want to stress that the difference in information content between the simple (reducible) and the complex (nonreducible) goal is enormous. The basic argument can be given clearly by using set theory. Thus: If each of n variables can take k dis-

with

(λ, μ and m being constants), stability criterion (5) is fulfilled if:

$$\left(\frac{\partial p'}{\partial p}\right)_{p^*} \leqslant \lambda < 1$$

This work was supported by the U.S. National Science Foundation, G17414.

1 Beurle, R.L., *Phil. Trans. Roy. Soc. London*, B,240, 55 (1956).

2 Ordnance Corps, *Tables of the Cumulative Binomial Probabilities*, ORDP 20-11 (Office of the Chief of Ordnance, Washington, DC 20025, 1952).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans

(Reprinted from Nature, Vol. 228, No. 5273, p. 784 only, November 21, 1970)

CONNECTANCE OF LARGE DYNAMIC (CYBERNETIC) SYSTEMS: CRITICAL VALUES FOR STABILITY

Many systems being studied today are dynamic, large and complex: traffic at an airport with 100 planes, slum areas with 10^4 persons or the human brain with 10^{10} neurones. In such systems, stability is of central importance, for instability usually appears as a self-generating catastrophe. Unfortunately, present theoretical knowledge of stability in large systems is meagre: the work described here was intended to add to it.

Most of these large systems, often biological or social, are grossly non-linear, which increases the difficulties associated with them. Here we consider linear systems merely as a first step towards a more general treatment.

We have attempted to answer: What is the chance that a large system will be stable? If a large system is assembled (connected) at random, or has grown haphazardly, should we expect it to be stable or unstable? And how does the expectation change as n, the number of variables, tends to infinity?

Monte Carlo-type evidence[1,2] had suggested that the probability of stability decreased rapidly as n was increased, in some cases perhaps as fast as 2^{-n}, an exponentially-fast vanishing of the chance that the system will be stable. This result, however, was for systems that were fully connected, where every variable had an immediate effect on every other variable. While this case is obviously important in theory, it is not the case in most large systems in real life: not every person in a slum has an immediate effect on every other person, and not every cell in the brain directly affects every other cell. The amount of connectedness ("connectance") is often far below 100 per cent. We have studied how much incomplete connectance affects the probability of a system's stability.

Let the linear system's state be represented by the vector x ($=x<x_1, \ldots, x_n>$, where each x_i is a variable, a function of time), and its changes in time by the matrix equation

$$x = Ax$$

To "join the variables at random" is to give the elements in A values taken from some specified distribution. "Non-connection from x_i to x_j" corresponds to giving the element a_{ji} the value zero. Thus, if the specified distribution has a peak at zero, sampling from it will give the equivalent of a dynamic system with many non-connections. The connectance, C, of the system can then be conveniently defined as the percentage of non-zero values in the distribution. Thus, if

tinct values, then the number of reducible (i.e., rectangular) relations between them is the number of rectangular subsets. In the space of k^n points, it is 2^{kn}. But the number of relations in general, not restricted to the reducible, is 2^{k^n}. Thus, specifying one of these relations (the subject of events acceptable as goal-achieving) requires, in the reducible case, kn bits, and in the nonreducible case, k^n bits.

The difference may be trifling to the printer, but numerically it is out of this world. Suppose, for instance, that the traffic at an airport involves only 100 variables (probably an underestimate) and that each variable need be distinguished in only 5 degrees (again, a very moderate demand). If the goal is reducible, specifying it may demand up to 500 bits; if it is *not* reducible, the quantity rises to 10^{70} bits, one bit for every atom in the universe! This fantastic leap is in no way exceptional; on the contrary, all that I have done in the last few years has shown that it is entirely typical. Allowing interaction commonly makes the informational content increase by vast orders of magnitude.

In these huge quantities we have a useful fixed point, and can keep some sense of proportion, by remembering Bremermann's limit.[4] Because of the quantal coarseness of matter, nothing made of it, machine or brain, can process information faster than about 10^{47} bits/p/sec. Take tons of computer and decades of time, and no feasible computation can handle more than about 10^{60} to 10^{70} bits. Thus, a generally complex goal, in the very moderate airport example just given, is already making demands quite beyond the achievable. I suggest that many of our troubles today, in our struggles with complex systems—especially in those researches that try to push into the really large and new—are basically ascribable to the fact that we are often attempting to handle quantities of information that are, by Bremermann's limit, actually unmanageable.

There are at least two ways in which the ideas and methods of information theory may help. First, even a rough approximation may suffice to warn us that we are attempting the impossible. Second, it may throw a quite unexpected light

on the various *strategic approaches* to a problem. Here is an example that I encountered recently.

Suppose, as in Fig. 2, that the designer has to select the right function F in this system with the m inputs $X_1, \ldots, X_m$. Now this system has an obvious transmission from the X's to the Y's, but transmission in information theory means more than the driving of electrons along a wire. Essentially, *information theory is the science concerned with deviations from statistical independence.* If the values of the Y's are not independent of those of the X's, then the ordinary "transmission through" occurs, but other deviations are possible. Thus the X_i's may show deviations from independence (correlations, say) among themselves; then a transmission may be defined and measured between the X_i's. In Ref. 5, $T(X_1 : \ldots : X_m)$ will not be zero. What effect will such correlation have on the quantity of information that must be handled by the designer?

Suppose $T(X_1 : \ldots : X_m) = 20$ bits. Another way of writing this fact follows.

$$2^{H(X_1, \ldots, X_m)} = 2^{\Sigma_i H(X_i)} \times 2^{-20}$$

Now "2 to an exponent entropy" is effectively the number of *independent* values, equivalent to the correlated. So the expression says that the transmission of twenty bits cuts the effective number of states at the input to the fraction 2^{-20}. Next, take the fact that the number of mappings of p input states into q output states (i.e., the number of F's from which the designer must select) is q^p. To select one F may demand up to $p \log_2 q$ bits. If q is fixed, the number of bits is proportional to p. But p has been cut by 2^{-20}, i.e., to one-millionth. Thus, the transmission, among the X_i's, of twenty bits does not just, for instance, subtract twenty bits from the designer's work: it cuts his work to a millionth part.

Having in mind this example and my other experiences of the last five years, I think it may reasonably be asserted that our most urgent need in artificial intelligence (and similar researches in highly complex systems) is that we be aware at

every moment of whether the information content is dependent upon a multiplier or upon an exponent. Crude and elementary as this distinction is, without it a worker may struggle to achieve a ten percent saving in efficiency, unaware that, because of a wrong basic strategy, he is working at a level a million times too high.

REFERENCES

1. W. McDougall. *Psychology*. New York: Holt, 1912.
2. W. R. Ashby. "The Set Theory of Mechanism and Homeostasis." In *Automaton Theory and Learning Systems*, ed. by D. J. Stewart. London: Academic Press, 1967, pp. 23-51.
3. R. C. Conant. "The Information Transfer Required in Regulatory Processes." *IEEE Transactions* SSC-5 (L969): 334-338.
4. H. J. Bremermann. "Quantal Noise and Information." *5th Berkeley Symposium on Mathematical Statistics and Probability* 4 (1967): 15-20.
5. W. R. Ashby. "Two Tables of Identities Governing Information Flows within Large Systems." *Communications of American Society for Cybernetics* 1 (1969): 3-8.

Information flows within co-ordinated systems*

W. ROSS ASHBY

University of Illinois, Urbana, Illinois, U.S.A.

Summary

Every coordinated activity, whether in the movements of a tight-rope walker's limbs, or in the traffic flows of a big city, requires an internal flow of information between the parts being co-ordinated. Once the co-ordination is well defined, the minimal quantity of internal information flow is determined numerically. An example is given to illustrate the principle.

This numerical quantity can be partitioned in various ways, corresponding to the various organizational ways for managing the co-ordination. So one can relate a proposed organization, in city or brain, to its resources for internal communication, to see if they are compatible.

Effects with delay ("memory") can be included in the formulation without essential change. Demands for memory, in co-ordinated activity, can be met in a variety of quantitatively different forms; so a designer can select among them for the most appropriate. An example is given as illustration.

We have brains primarily so that our bodily activities may be co-ordinated: so that our left hand shall act properly in conjunction with our right. Co-ordination and integration have long been recognized in physiology as the brain's highest functions, but cybernetics today is equally concerned with co-ordination in systems of other types. Big cities need co-ordination in their traffic flows; the prevention of smog requires that many preventive and

* The work on which this paper is based was partly supported by contracts AFOSR 70-1865 and OEC-1-7-071213-4557

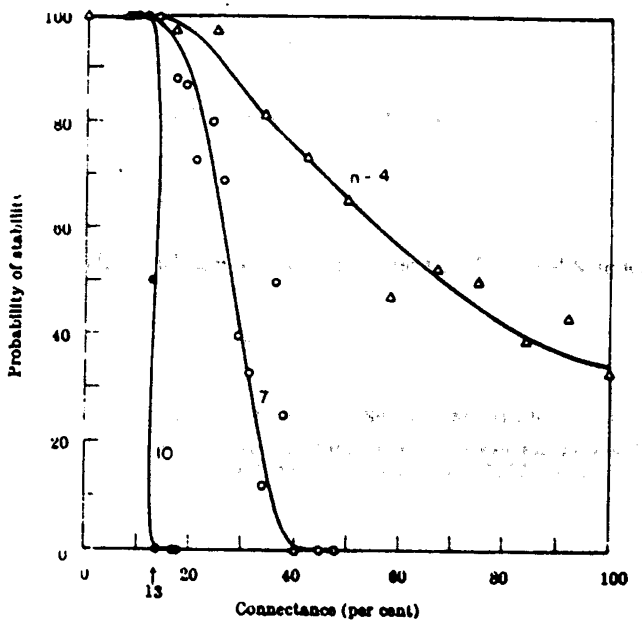


FIGURE 1. Variation of stability with connectance.

the coefficients are drawn from a distribution with 99 per cent zeros, and if n = 1,000, then each line of the equation would contain about ten non-zero coefficients, corresponding to a system in which each variable is directly affected by about ten other variables.

Because our work was essentially exploratory, we used the distribution in which the non-zero elements were distributed evenly between −1.0 and +1.0. The elements in the main diagonal, corresponding to the intrinsic stabilities of the parts, were all negative, distributed evenly between −1.0 and −0.1. Thus each sampled value of A corresponded to a system of individually stable parts, connected so that each part was affected directly by about C percent of the other parts.

On a digital computer, a value for n was given and a value for C. Random numbers appropriately distributed were then sampled to provide a matrix . Hurwitz's criterion was applied to test whether the real parts of A's latent roots were all negative (the stable case) and the result recorded. Further samples, giving further As, allowed the probability of stability (P) to be estimated. The probability was then re-estimated for another value of C, and so on, until the variation of P with C became clear.

The results showed the feature that we wish to report here. As the system was made larger, a new simplicity appeared. Fig. 1 shows a selection of the results, enough to illustrate the principal fact.

When n = 4, the probability that the system would be stable depended on C in a somewhat complex curve (which could perhaps be predicted exactly). But as n increases, the curve changes shape rapidly towards a step-function, so even when n is only 10, the shape might be so regarded, at least for some practical purposes. Thus, even at n = 10, questions of stability can be answered simply by asking whether the connectance is above or below 13 per cent: 2 per cent deviation either way being sufficient to convert the answer from "almost certainly stable" to "almost certainly unstable".

The matter is being investigated further, but it may be of general interest to notice that this work suggests that all large complex dynamic systems may be expected to show the property of being stable up to a critical level of connectance, and then, as the connectance increases, to go suddenly unstable.

This work was supported in part by the US Office of Scientific Research.

1 Ashby, W.R., ***Design for a Brain***, (Chapman and Hall, London, 1952).
2 Gardner, M.R., ***Critical Degenerateness in Linear Systems,*** Tech. Rep. No. 5.8 (Biological Computer Laboratory, University of Illinois, Urbana, Illinois 61801, 1968).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.

remedial actions be co-ordinated if one remedy is not to be nullified by another; and in social problems, too, the activities of welfare agencies need co-ordination. The purpose of this paper is to show that all co-ordinations require that information be transmitted within the system (a proposition that might be thought obvious), but particularly to show that the transmissions can be measured quantitatively. Every well-defined co-ordination specifies a basic total quantity of transmission, such that less than this quantity makes absolutely impossible the achievement of the co-ordination. It will also show that this total quantity can be analysed (partitioned) in various ways so that we can see how much is required between the components. In regulating traffic flow, for instance, it would show how much transmission is required between point and point. In the brain, it would show how much transmission is required between cell and cell, or between centre and centre.

Co-ordination is essentially a holistic phenomenon, discernible only over the whole. The method of information analysis presented here is also of this type. It does not say that between points X_i and X_j so much transmission must occur: it treats all the transmissions as a complex interrelated set, and allows the transmission between (say) X_i and X_j to take almost any value provided that suitable adjustments are made in the other internal transmissions.

The method could be presented in formal and abstract symbols, leaving the reader to find their application. I prefer to present one example, perhaps oversimplified, to show the method at work. The reader should have little difficulty in adapting the example to his own needs. The example is artificial; I would have preferred to have analysed real data, but it seems that no one has yet collected data on co-ordination with sufficient breadth to make this type of analysis possible. Perhaps when the existence of this method is better known, the experimenters will supply appropriate data.

THE TIGHT-ROPE WALKER

As example, let us consider that classic type of co-ordination shown by the tight-rope walker. The focal condition (Sommerhoff, 1950) is obviously that his four limbs must always have positions such that their centre of gravity lies vertically over the wire. (To keep the example simple I here ignore such complications as their angular momenta.) The unskilled person may well be able to move his limbs through just as wide a range as the expert, but the unskilled person will use combinations of position, all four limbs to the

left say, that the expert would avoid. Thus the contrast between the unskilled and the expert may be shown by the fact that the expert confines his actions to a particular subset of those anatomically possible.

The suggestion is not, of course, derived simply from tight-rope walking. As Sommerhoff (1950) showed extensively in biological examples, and as Ashby (1967) showed in the terms of set theory and binary relations, the identification of "co-ordination" with "deviations from statistical independence in an n-dimensional frequency table" is both broad and rigorous. Given any well-defined co-ordination between n variables, there is implied a frequency distribution over events in the n-dimensional space to which Shannon-type measures of information are applicable.

It is simpler now to proceed by example.

To avoid infinitesimals, suppose each of the four limbs can go to one of five places situated at distances -2, -1, -0, $+1$, $+2$ from the central plane. Thus, if the limbs L_1, L_2, L_3, L_4 are at (respectively) -1, $+2$, -1, $+1$, the average is $+0.25$, and their center of gravity is away from the central plane. If we allow two or more limbs to be at the same distance, there are 5^4 possible distributions (postures) of which only a subset would be used by the expert. It is easily verified that of these 625 postures, 85 have the zero average of the co-ordinated posture (61 in the 6 types of symmetrical distribution such as 00400, 01210, etc.; and 12 each in the asymmetrical 10120 and its reflection).

To obtain the necessary frequencies (or probabilities after dividing by the total) we may proceed on either of two assumptions (that lead, in fact, to just the same numerical results). One way is to assume that the 625 postures of the unskilled and the 85 of the expert are actually equiprobable, a very arbitrary assumption that may well be false when we consider actual people. The other way is to think of the case where the facilities for transmission will have to be provided, and to ask: suppose the worst happens—that without transmission all 625 postures occur, and that the expert (for other reasons) may be forced to produce all the 85: what is the least quantity of transmission facility that we must provide to be safe? This second form of the question seems to be free from objection (unlike the first), so I shall treat it as the question to be put.

With the frequencies now assumed to be equal (or the probabilities after dividing by 85) we can now find the basic entropies. These are defined in the usual way, by

$$H(X) = \sum_i p_i \log\left(\frac{1}{p_i}\right)$$

As we shall be using frequencies here, however, an arithmetically more convenient method, if the frequencies are $n_1, n_2, \ldots, n_i, \ldots$, summing to n, is to find $H(X)$ as

$$\frac{1}{n}\left(n \log n - \sum_i n_i \log n_i\right)$$

(With the n_i all integers, much interpolation may be avoided.)

In the general case, these entropies would be found by whatever method was appropriate. In this example we can soon find that L_i has the following frequency distribution, in the co-ordinated postures.

Value:	−2	−1	0	+1	+2	
Frequency:	15	18	19	18	15	Total: 85

So $H(L_i) = 2.315$ bits/posture. By symmetry, this is also the value of $H(L_2)$ etc.

$\langle L_1 L_2 \rangle$ has the following distribution, over its 25 values.

L_2:						
+2	5	4	3	2	1	15
+1	4	5	4	3	2	18
0	3	4	5	4	3	19
−1	2	3	4	5	4	18
−2	1	2	3	4	5	15
	−2	−1	0	1	2	

L_1

So $H(L_1, L_2) = 4.544$ bits/posture. All the 85 values of $\langle L_1 L_2 L_3 \rangle$ are different, so $H(L_1 L_2 L_3) = \log_2 85 = 6.409$. Similarly $H(L_1 L_2 L_3 L_4) = 6.409$ bits/posture. If a posture is significant over a time span of (say) 0.5 seconds, then twice these numbers would give the entropies in bits per second.

PARTITIONING THE INFORMATION FLOW

The further analysis uses the methods introduced by McGill (1954) and developed by Garner (1962) and Ashby (1965, 1969). The most important quantity required now is the total transmission, represented and defined by

$$T(L_1 : L_2 : L_3 : L_4) = H(L_1) + H(L_2) + H(L_3) + H(L_4) - H(L_1 L_2 L_3 L_4)$$

It measures the total deviation from statistical independence implied by the co-ordination (with the marginal distributions given). Here its value is

2.850 bits/posture. Its importance is due to the fact that with less than this total quantity of internal transmission the co-ordination cannot be ensured.

It is worth noticing that the total transmission required is not the obvious $\log_2 625 - \log_2 85$ (= 2.878) but a quantity smaller by 0.028. The reason is that the larger quantity would apply were each variable L_i to be distributed evenly over the five values. In fact the distribution (in the co-ordinated case) is not even. Thus, if the variable's distribution is changed from 17, 17, 17, 17, 17 to 15, 18, 19, 18, 15, the change would bring the conjoint 4-variable distribution nearer to the co-ordinated form *without any transmission being used between the variables*. Thus the algebraic and numerical analysis has already revealed a possibility for economy and efficiency that otherwise might have passed unnoticed. (In this example the gain is trivial; in other cases it might be of major importance.)

The total quantity of transmission required may be obtained by adding various components. One possible way is to use the fact that $T(L_1 : L_2 : L_3 : L_4)$ is identically equal to

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

Such a partition would be appropriate if the total co-ordination were achieved by mechanisms or channels that (1) achieved suitable co-ordination between L_1 and L_2 (between the arms, say) regardless of the positions of the legs, (2) achieved co-ordination between the legs regardless of the arms, and (3) co-ordinated arms and legs in a way not depending on the details of the relation *between* the arms (e.g. if the arm-pair has centre of gravity at +0.5 then the leg-pair must have centre of gravity at −0.5). The three quantities are found to be

0.086, 0.086, and 2.678

(respectively), summing to 2.850, of course.

Such numbers may be useful in various ways. Thus, suppose that only 2-bit channels were available. Instead of taking two such channels to achieve the 2.678, we could try another way of distributing the transmissions. Another way is represented by the partition (of the total) to

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

This partition would be appropriate if the co-ordination were achieved by first a constraint holding between L_1 and L_2; second, by the outcome of

this constraint (the vector $\langle L_1 L_2 \rangle$) acting to constrain L_3; and then the consequent $\langle L_1 L_2 L_3 \rangle$ acting to constrain L_4.

The quantities required are (respectively)

0.086, 0.449 and 2.315

still excessive in the last quantity. We have also, however, that this last quantity may be partitioned further:

$$T(L_1 L_2 L_3 : L_4) = T(L_1 L_2 : L_4) + T_{L_1 L_2}(L_3 : L_4)$$

$$2.315 = 0.449 + 1.866$$

Thus the requirement could be met by an extra channel of capacity 0.449, together with one whose average capacity is 1.866, linking L_3 and L_4; with the coding between them determined conjointly by L_1 and L_2.

Here I have written as a designer might see the matter, and use the equations to guide the design. The physiologist might use them if, say, he knew that no channel of more than 2.000 bits/posture was neuronically possible. Then the analysis would show decisively that any proposed neuronic arrangement using the first mode must be rejected: such a net could not achieve the observed co-ordination.

The coding question will not be treated in this paper as it is still being studied. By Shannon and Weaver's (1949) theorems, the necessary codings will certainly exist, but the theorems assume that successive acts of co-ordination (successive postures here) can borrow signalling capacity in order to make up an efficient code. If this cannot be done, then the actual capacities required may be somewhat higher than the numbers given here. Further consideration of coding can be given only when more details of the particular case are available.

MEMORY

In the co-ordination just described, it has been assumed that the variables specify the positions of the four limbs taken simultaneously. Exactly the same logic, and the same algebraic method hold good when the co-ordination occurs over time: when later events must be co-ordinated with earlier. $H(X, Y)$ may be the entropy of two distant events taken simultaneously, but it is equally possible that X and Y are separated only in time, so that

$X = Z(t)$ say, and $Y = Z(t + k)$. Now, if the system is to co-ordinate X and Y, it must have "memory", in some form, over the time span k. An example will show the method and something of the possibilities. Again it is artificial, for lack of presently existing real data.

Let us suppose that three unmanned vehicles will be landed on a planet, which has five places of interest. It is required that the three vehicles shall (1) at one time go to some three of the five places (no two vehicles to the same place), and (2) at another time, meet, all three, at a place other than those visited singly. (Events (1) and (2) may occur in either order.) And it is required that the co-ordination's demands on memory shall be minimal.

The computations are straightforward. We prepare for the worst case, where all events and distributions are equiprobable. Let the five places be $\{1, 2, 3, 4, 5\}$, and the three vehicles $\{A, B, C\}$. Let A, B, C represent their places on the first occasion in real time (regardless of whether event 1 or 2 is achieved), and A', B', C' their places on the later occasion. Thus the vector A, B, C, A', B', C' would show the defined co-ordination if its value were $\langle 4, 4, 4, 5, 2, 1 \rangle$ or $\langle 2, 5, 3, 1, 1, 1 \rangle$, and other similar combinations.

The basic entropies, in the "co-ordinated" case, are easily found.

(1) The 5 values of A occur all with frequency 48, so $H(A) = \log_2 5 = 2.322$. Similarly for $H(B), \ldots, H(C')$.

(2) The 20 permitted values of AA' occur all with frequency 12, so $H(A, A') = 4.322$; similarly for $H(B, B')$ and $H(C, C')$.

(3) Of ABC, 5 values (event 2) occur each with frequency 24, and 60 values (event 1) each occur twice. So $H(A, B, C) = 5.114 = H(A', B', C')$.

(4) The 240 permitted values of $ABCA'B'C'$ each occur once; so $H(A, B, C, A', B', C') = 7.907$.

The units are bits per double event.

One obvious way of organizing the system is to co-ordinate within each event at each of the two times, and also to co-ordinate between the two times. The total transmission required over *both* distributions of vehicles, is 6.025 bits, analysed to

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

$$1.851 \quad + \quad 1.851 \quad + \quad 2.322$$

Another method of organizing to achieve the co-ordination, would be to consider the "trajectory" (or transition) taken by each vehicle, as A might go $4 \rightarrow 5$, B $4 \rightarrow 2$, and C $4 \rightarrow 1$, and then co-ordinate the trajectories. This

ON TEMPORAL CHARACTERISTICS OF BEHAVIOR IN CERTAIN COMPLEX SYSTEMS

C.C. WALKER*, and W.R. ASHBY

Department of Psychology, University of California, Los Angeles, USA
Received January 17, 1966

Summary: Little is known about the behavior of dynamic systems with many intricately interacting parts, and about the factors which tend to affect their behavior in general, rather than detailed, ways. This paper describes a study of such systems built up out of unit elements which compute recursive logical functions.

Each element has two binary inputs and a binary internal state which is also the element's output state. (Output of elements can be branched.) Recursion is introduced by lettering the element's state at the next instant of system time $(t + 1)$ be a function of the present states of the two inputs as well as its internal state at the present system time (t). Hence, there are 256 different functions that can be computed, and a particular element's behavior is defined by the one function it computes.

One hundred identical elements connected at random constitute one system. Two hundred fifty six types of systems, corresponding to all the 256 logical functions, are studied by computer simulation, using five different sets of connections, starting the systems at ten randomly chosen initial system states. After being set at the initial state each system produces its behavior without further interference. We studied particularly the effects on these behaviors of those factors that might determine (i) how long a system would take to arrive at its terminal cycle and (ii) the size (periodicity) of the cycle shown terminally.

Among the facts elicited, the following seem especially notable:

1. Such systems tend to end in a complex cycle of behavior. The very short cycle is by no means the common ending.

2. The style of behavior, apart from details, is often strikingly independent of the pattern of connection.

3. One of the factors markedly affecting the length of time before the terminal cycle can be detected by an observer is the extent to which the elements act as informational transmitters.

4. A factor strongly affecting the tendency to terminate in a very short cycle is the number of conditions in which the elements' states will remain unchanged at the next instant of time.

5. The use of elements whose transitions are highly dependent on the element's preceding states encourages short initial periods before the system reaches long terminal cycles.

would demand the quantities

$$\underbrace{T(A:A') + T(B:B') + T(C:C')}_{0.322 \quad + \quad 0.322 \quad + \quad 0.322} + \underbrace{T(AA':BB':CC')}_{5.059}$$

$$0.966$$

The term $T(A:A')$ represents "memory" affecting only vehicle A, regardless of what the other vehicles do; and the same for vehicles B and C. What is striking is that the *three* "memories" of this type demand only 0.966 bits as compared with 2.322 bits demanded by the single, and more obvious, first type. The method thus enables different *functional* forms of "memory" to be examined for various characteristics.

One would, of course, also have to consider the physical method used to achieve the co-ordination between transitions, $T(AA':BB':CC')$. It is sufficient for us here to notice that these numerical analyses refer only to deviations from statistical independence in their quantities, not to any reasons, or physical causes, for the deviations. Thus any quantity T, here called "transmission", does not necessarily need an engineer's communication channel: suitably paired responses to a common signal may well provide the formal "transmission" demanded by these identities.

The coding problem remains, but I am content if I have shown that the fundamental concepts of co-ordination and integration can be measured, and that the measurements may give information about the system that is much deeper than can be obtained by simple intuition.

References

Ashby, W.R. (1965). "Measuring the internal informational exchange in a system". *Cybernetica*, **8**, 5–22.

Ashby, W.R. (1967). "The set theory of mechanism and homeostasis". In *Automaton Theory and Learning Systems* (Ed. D.J.Stewart. Academic Press, London. pp.23–51.)

Ashby, W.R. (1969). "Two tables of identities". *Bull. Am. Soc. Cybernetics.*

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

McGill, W.J. (1954). "Multivariate information transmission". *Psychometrica*, **19**, 97–116.

Shannon, C.E., and W.Weaver (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Sommerhoff, G. (1950). *Analytical Biology*. Oxford University Press, London.

Reprinted from *BioScience*, Vol. 18, No. 3, 1968, pp. 190-192

Information Processing in Everyday Human Activity

W. Ross Ashby

Information-processing in Man has so far been tested by finding his utmost capacity on some highly specialized task such as playing the piano or stenographing speech. How much is transmitted during his everyday life has been, so far as we are aware, not estimated. Yet this quantity must be basic in any study of his higher psychophysiology. We have therefore attempted to assess it.

We soon found that while the obtaining of numerical estimates was easy, these estimates differed so widely — by factors of a millionfold or more — as to make clear that the real problem was not the obtaining of numbers but the elucidation, and the logical justification, of the *method* to be used. Several possibilities were explored. In this article we report on, and confine ourselves to, what we now consider the essentials.

Especially instructive is consideration of the *minimal* quantity of transmission that must occur if some piece of everyday activity is to be accomplished successfully. Thus, if a man is to walk for even a few steps, the various movements at hip, knee, and ankle (with minor additions) must be coordinated; that is to say the various movements must *not* occur with statistical independence. Successful walking implies a major deviation from independence, and this deviation can be measured by Shannon's (1949) and McGill's (1954) measures of "transmission."

This "transmission," as b bits per second, does not imply that b bits must be sent *from* the hip's sensory endings *to* the knee's muscles of control; but it does imply that at least b bits per second must be transmitted somewhere in the system, in some appropriate manner, if the whole coordinated activity is to be produced by the normal processes of cause and effect.

To make the idea perfectly clear, since it is basic, let us consider the following simple example of coordination (not quite from everyday life): A pianist, during a passage, plays on the notes A, B, C, D, E, F, but only so as to produce chords of the interval of a third. If we form a frequency table showing how frequently the two fingers (X and Y) struck the various pairs, the only nonzero frequencies will be those in the asterisked cells of Table 1.

The author is a professor in the Department of Electrical Engineering, University of Illinois, Urbana.

TABLE 1

Note played by finger Y	A	B	C	D	E	F	Total
F	0	0	0	*	0	0	1
E	0	0	*	0	0	0	1
D	0	*	0	0	0	*	2
C	*	0	0	0	*	0	2
B	0	0	0	*	0	0	1
A	0	0	*	0	0	0	1
	A	B	C	D	E	F	8

Note played by finger X

If the nonzero frequencies occur equally (the most severe case), the entropy of Y ($=\Sigma\ P \log_2 P$) will be on the probabilities ⅛, ⅛, ¼, ¼, ⅛, ⅛, and will thus be 2.5 bits. $H(X)$ has the same value. $H(X, Y)$ will be of ⅛ repeated 8 times and will thus be 3.0 bits. The transmission implied by the restriction on the pairings is then 2.5+2.5−3.0, i.e., 2.0 bits per chord. Similarly, *every set of actions showing coordination between variables implies a minimal quantity of transmission between those variables.* It should be noticed that asking "how little can a man transmit?" is by no means as absurd as it may at first seem. This minimal quantity of transmission to *achieve a given coordination* may be regarded as analogous to the minimal quantity of work that a man (of known weight) must do to *climb a given height.* This latter quantity is fundamental in any question of energy: our quantity has a similar status in any question of coordination.

When the variables are more than two — $X_1, X_2, \ldots, X_n$ say — the measure of the total transmission necessary to achieve the given coordination will be given by $T(X_1 : X_2 : \ldots : X_n) = H(X_1) + \ldots + H(X_n) - H(X_1, \ldots, X_n)$ (McGill, 1954; Ashby, 1965).

With these ideas in mind, we took the following defined Action as basis for study and as reasonably typical of a piece of "everyday life."

(The human subject is given as being engaged in reading when he encounters an unfamiliar French word.)

Action: He walks across the room to his book shelf (avoiding a chair that is in his path), finds his French Dictionary (among 100 other books), finds the word, reads the English translation, and writes down the corresponding English word.

Now "information," as understood today, has meaning only when defined over some sample space (Shannon), or over a set of frequencies (McGill): the *multiplicity* of possibilities is essential. If therefore we think of this Action as having been performed by a particular person in a particular room on a particular day, then *this* event is unique in the universe, has no multiplicity, and makes any question about its informational properties merely improper. To bring this event into some relation to a measure of information, we must extend it to a *set* of Actions. It is this extension, in our opinion, which is the critical and essential step in the development of a logically defensible method.

This point of view (if disputed) may be made more plausible by considering a related question in the measurement of probabilities. Suppose we watched the door of Smith's store and ascertained the unique fact that the last person to step through the doorway before 12 noon was male. (This is the "particular event.") The question might then be raised: What is this event's probability?

Such a question demands a sample space: none has been defined. If the question of probability is to be pressed, a sample space must be provided. Clearly, the particular event may be embodied in many different sample spaces, each of which will give its own measure of the event's probability. Thus, we might extend the event to include all those people who passed through Smith's door at all hours of the day, or, keeping the time to noon, we might extend it to include the doors of all the stores in the street; and many other extensions are possible. Clearly, which extension is selected must depend on other criteria, depending on why the question was raised in the first place. Here all we need notice is that, so far as method is concerned, some sample space must be selected.

The following extensions of the unique Action seemed to us to be reasonably in accord with our restriction to "everyday life":

1) *Variations* that would occur even if the subject attempted at once to repeat his Action.

   a) Those due to inaccuracy of muscular movement, as in walking.
   b) At what pages the Dictionary falls open as the subject searches for the word.

2) *Variations* that might not occur on immediate repetition but would occur if similar Actions were taken on other days.

   a) The particular French word sought.
   b) The position of the obstructing chair.
   c) The position of the Dictionary among the other 100 books.

3) *Not varied* (in our study below) were all variables not in (1) or (2) above; in particular:

   a) The architectural features of the room.
   b) The subject's initial position in it.
   c) The French Dictionary.
   d) The other 100 books.
   e) The subject himself, his past experience, and his memories.

With the set of Actions well defined, we can now proceed to obtain well defined estimates of the necessary transmission between the variables if the coordinated and successful Action is to be achieved. (And as the average adult *does* perform the Action successfully, we can be sure that the average adult does transmit at least this quantity; should he transmit more, the excess measures his inefficiency.)

The computation, and its logic, can perhaps be made clearer if the following proposition be accepted as an axiom: *Once the sample space or set* (over which the transmission is to be computed) *has been defined, the computation proceeds in just the same way, and must arrive at the same number, whether the subject is an intelligent Homo or is a Robot designed to perform just that set of uctions and nothing more.* The approach through this axiom may reduce greatly one's initial intuitive estimate of what is necessary. In particular, it removes from our consideration all the activities *within* the nervous system, for these activities are neither described nor varied in the defined set of actions. (If the reader prefers to introduce neuronic variations into those listed above, his numerical answer would be different from ours: the *method*, however, would be the same. Essentially, he would be answering a different question.)

The Action falls naturally into about nine successive components that are sufficiently independent for their trans-

missions to be compounded by simple addition. The nine are given in Table 2, with our estimates of the transmission required in each component (details are given in the Appendix). Though many modifications might be made, our experience has suggested that such modifications are not likely to change the estimates by more than a factor of about 2. We are content that better estimates may be made later; in this paper our focus is essentially on the logic of method.

TABLE 2

(1) Walking 10 paces on two legs while maintaining normal verticality	30 bits
(2) Selecting a path to avoid collision with the chair	10 bits
(3) Finding the Dictionary among the 100 other books	7 bits
(4) Reaching out to the Dictionary, grasping it, and removing it from the shelf	22 bits
(5) (When the book falls open) Identifying how the opening is related to the wanted word	10 bits
(6) Repeating the opening by finger-movements till the page of the wanted word is reached	39 bits
(7) Reading the French word (to verify that the correct word has been reached)	6 bits
(8) Reading the corresponding English word (taking it to some central "cerebral" store)	14 bits
(9) Converting the stored word, through finger-movements, to a written word	31 bits
Total	169 bits

Discussion

The most surprising feature of the final result was, to us, the smallness of the number: 169 bits for about a minute's activity, or 3 bits per second. On further consideration, however, we concluded that the estimate may be essentially sound, for the following reason.

The question asks, in effect: If a robot is built to carry out the defined Action successfully, with the coordinations and corrective actions necessary, how much transmission *must* be provided? The answer cannot be far from our estimate, for either the machine will not be able to give a tolerable imitation of this Action (by being excessively clumsy), or it will demonstrably be using transmission wastefully. Yet even if it (or the human counterpart) were only 1% efficient, and used 300 bits per second, one would still want to know (say) why man's optic nerve, with about 500,000 fibers, offers at least that latter number of bits per second. We may well ask: Why do man's sense organs accept all the extra information?

A possible answer is suggested as soon as we realize that the two systems we are comparing are a Robot (or a man) performing the defined Action *and nothing more*, and the man of real life, who can perform not merely this Action (call it A_1) but who can also perform a great number of other Actions A_2, A_3, A_4, . . . Even while engaged in A_1, the normal man is able to respond to the intrusion of other variations — the ringing of the telephone, the discovery that the Dictionary is missing, the collapse of the bookshelves, and a host of variants not given in our list of "everyday variants" above. These choices *between* A_1, A_2, A_3, etc., will require a "higher level" activity with information-processing extra to that used *within* any particular A. Our estimate suggests that this "higher level" activity, not detectable while the Action is in progress, is, in fact, requiring much larger quantities of transmission than that used in the more obvious Action itself. One is reminded here of the modern computer, which differs from the older computer largely in the amount of organizational activity it undertakes, activity concerned not with direct computation but with *which* computation shall occur, and how and where.

Any estimate of the quantities of information-processing occurring in these higher levels requires, of course, consideration of the population of A's, and is beyond the scope of this study.

Finally, the fact that these estimates are acutely dependent on just which sample space is chosen (and chosen arbitrarily) may disturb the reader, for the freedom implies that the choser can make the estimate, here 169 bits, take any value he pleases. Can so arbitrary an estimate have any scientific value or use?

Here we would point out that a somewhat similar situation exists with "potential energy." The potential energy of a brick, say, can be given any arbitrary value, either by digging a suitably deep well under it (into which it can fall), or by bringing a suitably cold object near to it (to which it can give heat), or even by bringing up some antimatter! Yet the concept of potential energy in physics is obviously by no means useless. In practice, of course, we commonly use it to find its change, ΔE, which makes the arbitrary total value irrelevant. In addition, the quantity of potential energy is always discussed in relation to *the operations in which it takes part*. Perhaps one result of our work is to suggest that the "quantity of information" in a biological system would be better considered not as a measurement to be made for its own sake but in relation to a defined set of operations in which it is playing an active part.

Appendix

The Estimates

Component 1. Assuming a minimal analog for walking of 4 states of leg position and 4 states of each arm for balance, and assuming a 7-position plumb bob as a reference for balance orthogonal to the line of movement: for transmission between the bob and the left arm, 4 $\log_2$ 4, i.e., 8 bits; similarly for the right arm. For transmission between the legs, also 8 bits. Starting and stopping, each requires in addition that the organism must turn the plane of the bob and change the degree of freedom of both the arms, so an additional 6 bits will be needed.

Component 2. The subject can avoid collision if he can select to about ½ ft in a 10 ft width ($= \log_2 20$, i.e., 4.3 bits), and if he can select to ½ ft in 30 ft of progression ($= \log_2 60$, i.e., 5.9 bits). Thus the avoidance of collision does not require more than 10.2 bits.

Component 3. To select one object from 100 others requires $\log_2 100$, i.e., 6.6 bits. More may be necessary in practice because the difficulties of coding may lead to the use of an inefficient form of coding. But in any case there is no need to exceed 100 bits, for this would suffice for the very inefficient method (when both the Dictionary and other books are unvarying) of examining them one by one.

Component 4. To reach to a given point, shoulder, elbow, and wrist will have to be appropriately set to one of about 32, 16, and 8 positions respectively. The three are not wholly independent, so that the transmission necessary will be somewhat less than the sum of 12 bits. Once the hand is near the book, there may still be required a hooking of the index finger over the top of the spine — one of 8 positions at the middle joint and one of 4 at the terminal joint ($= 5$ bits), and similar movements at one other finger to grasp the book opposite to a rigid thumb.

Component 5. Since the book is opened (by component 6) as approximately a dichotomy, the decision whether the wanted French word lies before or after the place of opening requires no more than about 1 bit. However, a code using as

The significance of these facts for various applications in biological computers is discussed.

1. Introduction

A complex system is conveniently defined (Simon, 1962) as anything made up of many parts that interact in a non-simple way. At present there is little known about the behavior of such systems, even those with otherwise relatively elementary features. The work reported here examines temporal characteristics of behavior in a family of complex systems defined with reference to theoretically basic properties of the systems' parts and their interrelationship. The work was undertaken to provide a better grasp of how systems behave generally, that is, to sketch extremes, to find the behavior to be expected typically, and to examine the possibility that styles of system behavior may be related to simple characteristics of the parts' behavior.

The data are discussed from a point of view that will be most familiar to the biologically oriented reader. However, since control mechanisms in demand today are approaching biological complexity, and as the increased use of heuristic programs and Monte Carlo methods in conventional computing machines suggests that there may come to exist a need for probabilistic machinery, the results given here may be of interest to hardware-oriented readers as well.[1]

The parts composing the systems studied are simple electrical devices, here called *elements*, that can interact with one another. Each system is formed by taking many elements and joining them in an intricate arrangement; a family of systems is produced by systematically varying the elements' behavioral properties. The typical behaviors of individual systems of the entire family are the objects of experimental inquiry.

Given primary attention are three aspects of behavior: (1) the periodicity of the systems' terminal behavior, i.e., the length of time between repetitions in terminal behavior; (2) the duration of the temporary behavior, i.e., the length of time before the system settles into its terminal behavior; and (3) the activity in the system, i.e., the relative number of elements that change states, from one instant of time to the next.

---

1. A more detailed treatment of the work is found in: "A study of a family of complex systems, an approach to the investigation of organisms' behavior," by C.C. WALKER, AF Grant 7-64, Technical Report No. 5, June 1965, Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana, Illinois. Available on request through the Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, Illinois, U.S.A.

Previous results (Ashby, 1960; Fitzhugh, 1963) suggest that in intricately structured systems the production of very short cycles is related to the tendency of the parts to remain unchanged at the next instant of system time, and that systems which produce very short cycles reach activity levels of zero along an approximately exponential decay. Little else about such systems' behavior of relevance to the present study is known at present.

2. The Systems Studied; the Behavior Observed

2.1 The Systems Studied

The systems examined are:

a) Structurally intricate. A system's *structure*, by which is meant the network of connections that mediate its elements' interactions, has many loops and is without obvious regularity: the "circuitry" implied by the structure is very involved.

The process by which structures are actually determined is given in abstract terms below, and is interpreted operationally in Section 4.

b) Structurally rigid. The structure does not change thereafter, once a system is constructed.

c) Built of simple parts. The basic parts of which the systems are built, i.e., the elements, are functionally elementary. Elements are described fully below, but, to anticipate that discussion, they are devices just complicated enough to provide for structural complexity in the system, and to allow memory in the individual element.

d) Functionally homogeneous. The elements of any one system are identical.

e) Not influenced by factors outside the system. The activities of the systems are determined wholly from within, except for certain "start" and "stop" prerogatives reserved to the experimenter.

f) Clocked. Time, in the systems, is quantized.

Elements. So as to consider basic forms, the elements are taken to have two states, two independent inputs and any number of outputs; their inputs' states and their own states determine their succeeding states. The outputs of an element at any time carry the element's state at that time.

Transformations (Ts). Given the form of an element established above, a table such as Table 1 is a suitable representation of an element's behavior. Table 1 shows what the element's next state will be when its present state and its inputs' states take the values shown. The table gives the element's *transformation (T).*

little as this would be difficult to construct (identifying which letter of the alphabet is at the opening would require $\log_2$ 26, i.e., 4.7 bits). Over the whole Action, with 10 such decisions, 10 bits would be the minimum.

Component 6. A book of 1024 ($=2^{10}$) pages requires 10 dichotomies to arrive at a particular page. If each dichotomy is not accurate but is made within the middle one-fifth of the block, $\log_2$ 5, i.e., 2.3 bits, is needed for the selection of this fifth. Further, after each dichotomy, the decision whether to operate on the left or the right block demands 1 bit: 3.3 for both. Ten steps require 33 bits. Then, on the selected page, one word must be selected from (say) 50 words, a further need for 5.6 bits.

Component 7. Testing whether the located word in the Dictionary is the same as the sought word requires no more than deciding first whether the initial letter of each word is "same or different," a 1-bit discrimination, followed by 1 bit for each succeeding letter. An average word of 6 letters thus has a basic requirement of 6 bits. As the coding would have to be somewhat peculiar to achieve the minimum, the practical requirements will usually be somewhat greater.

Component 8. To obtain some state in the brain corresponding to some one word of 20.000 requires a transmission of $\log_2$ 20,000, i.e., 14.3 bits. ($H(X) = H(Y) = H(X,Y) = \log_2 20{,}000$.)

Component 9. Transcribing the stored word, with the subject experienced in writing whole letters, requires less than $\log_2$ 26 bits per letter (4.7), with (say) a repetition of 1 in 10 for error (of known location), requiring a 10% increase. A 6-letter word would thus require 6 x 4.7 x 1.1 = 31 bits.

Acknowledgment

Research for this article was sponsored in part by AF Grant 7-67, the Air Force Systems Engineering Group, and the National Aeronautics and Space Administration.

References

Ashby, W. Ross. 1965. Measuring the internal informational exchange in a system. *Cybernetica*. 8: 5-22.

McGill, W. J. 1954. Multivariate information transmission. *Psychometrika*, 19: 97-116.

Shannon, C. E., and W. Weaver. 1949. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Measuring the Internal Informational Exchange in a System

by W. Ross ASHBY (1) *(United Kingdom)*,

Professor at the University of Illinois, Urbana (U.S.A.)

If the researches in Cybernetics of the last ten years have shown anything, they have shown that really large systems — the living brain, an actual society, a big computer program, the biochemical processes of a cell, a nerve-network — have such excessively large complexities that the scientist, having only relatively small resources for their study and control, must necessarily simplify. He must focus his attention on some one aspect at a time, temporarily ignoring all others.

The force of this argument has recently become overwhelming. When large quantities of information-processing was demanded, one used to feel that all that was necessary was yet a little more speeding up of the electronic machinery, or perhaps a doubling or so of the size of the computer. BREMERMANN [1], however, has now demonstrated that no system made of matter as we know it today, and therefore subject to the mass-energy relation and to Heisenbergian uncertainty, can possibly process more than 10^{47} bits per gram per second. Take tons of computer and centuries of time, and one merely adds a few units to the exponent. Nothing is easier, however, than to make demands that go almost infinitely beyond this limit. Suppose, for instance, that a machine has 10.000 two-state relays : any thought of searching through its configurations instantly raises a demand for at least $2^{10.000}$, i.e. for $10^{3.000}$ operations. This number is physically impossible by a little less than 3.000 orders of magnitude. Cybernetics is today known to be

(1) Text of a lecture given at the 4th *International Congress on Cybernetics*, Namur, 19-23 october 1964.

The work on which this article is based was supported partly by the United States Air Force Office of Scientific Research, under Contract AFOSR 7-63

Table 1. *The Standard Form for an Element Transformation Table, here a General Transformation*

	Next State of Element		Present State of Element	
			0	1
	0	0	e_1	e_5
Present States of	0	1	e_2	e_6
Inputs	1	0	e_3	e_7
(L, R)	1	1	e_4	e_8

The states of an element are formalized as the integers 0 and 1. For simplicity, the inputs to a given element carry the states of the elements which output to the given element. The two inputs on an element are kept separate from one another and are arbitrarily designated its left (L) and right (R) inputs.

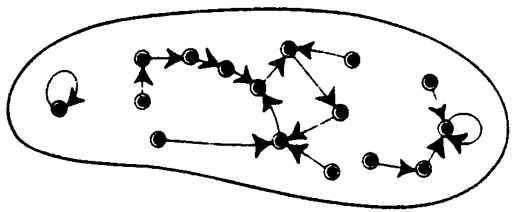


FIGURE 1.
An arbitrary field with sixteen states.

The marginal headings will be omitted in later transformations when no confusion is likely to result from their omission.

Each table entry $e_1, e_2, \ldots, e_8$ is either 0 or 1, and is assigned one of these values in any particular transformation. Since there are eight different entries and each may be 0 or 1, there exist 2^8 or 256 different transformations. We consider them all.

Structures (Ks). The structures used are generated by a process which can be visualized as follows: Let N elements be numbered from 1 to N. Place the numbers 1, 2, . . . , N in an urn. Take the first element's left input, draw a number at random from the urn and join that input to an output of the element designated by the drawn number. Replace the number in the urn. Repeat for the right input. Continue the process for each of the rest of the elements. The end of the process is a definite structure *(K)*. There is nothing variable or probabilistic about it once it is obtained. Note that self-affecting elements are permitted. Note also that each input is connected to some element in the system; the systems are therefore autonomous or input-free in the terminology of automata theory. As a further simplification, the structures are taken to be rigid. As was already mentioned, once a structure is defined, it does not change thereafter.

Time. Time in the systems is discrete, $t = 1, 2, 3, \ldots$, and the actions of the elements occur simultaneously. When an element assumes a state, its

outputs immediately communicate that state to those elements which it affects.

2.2 The Behavior of the Systems

The *state* (henceforth, when unqualified, this word refers to the state of the *system*) is the set of all element-states at a particular time, the elements taken in a fixed order (the order of their conventional numbering). For any given system its state at any time is a function of the preceding state, and nothing else. Denoting by S the state at time t, and by S' the state at time $t + 1$,

$$S' = f(S), \tag{1}$$

where f is itself a function of the transformation and the structure. Thus, the systems of this study are "independent mechanisms" in the sense of Murray (1955).

Fields. Each system of this study can be considered completely determined by the specification of (1) a transformation and (2) a structure. These two data determine not only the specific system in the physical sense, but its entire repertoire of behavior as well. A system's behavioral repertoire is conveniently visualized in the form of a diagram, e.g., Fig. 1, in which points represent states, between which are drawn arrows representing the system's state transitions determined by f [of Eq. (1)]. In this study such a diagram is called a system's *field.*

Cycles-Terminal Behavior. If the points in Fig. 1 represent all the states of some arbitrary system, then arrows suitably drawn between the points form a field. The systems of this study are independent mechanisms. Their fields, therefore, show connected arrows ending in sets of states which, if the systems are thought of as working with no limit in time, are continually reproduced. These states are the systems' terminal, or permanent behaviors. Such states are commonly said to be in *cycles.* The number of states in a cycle is here called the *cycle length*, which is the length of time between successive repetitions of any one state in the cycle. In Fig. 1, there are three cycles, with cycle lengths of one, four, and one states, respectively.

In the present study, very short cycles, those of length one, are given special attention. Cycles of length one are here called *states of equilibrium.* Such states are formally equivalent to what is common in the biologically important concepts: "is at rest," "has decided," "is adapted," "has learned the task," and so on.

Run-Ins — Temporary Behavior. In Fig. 1, note that there are states which are not in cycles. Once produced, such states do not recur. These states are the system's temporary behavior. A trajectory segment which immediately precedes a cycle is here called a *run-in.* The number of states in the run-in, from some initial state up to, but not including any cyclic states, is

urgently in need of methods that shall give us what we can actually get — that shall give us what we *really* want, not what we think we want. Finding methods for simplifying is thus the very core of today's problems in Cybernetics.

One such method is to study the system in its informational aspect. The aim of this paper is to show how this aspect can be treated mathematically, scientifically, and operationally.

Information theory was originally a study of two variables, the sender's state and the receiver's. McGILL [2] studied informational relations between three and four variables and indicated how the relations could be generalised to n. Here I want to consider the relations between n variables, especially when n is really large, as large, say, as the number of cells in the human brain — about 10^{10}.

The basic idea motivating this paper can be most readily seen by a practical example. Suppose a fleet, equipped with all modern signalling devices, finds just before it sails for war that a component used throughout the apparatus has proved defective, so that the fleet has to put to sea with only fifty old-fashioned hand-lamps for signalling from ship to ship. Clearly, the admiral may dispose of his fifty signallers in various ways over the ships, and there may be no manoevre of the whole fleet that is completely impossible; yet this lack must impose *some* characteristic on the fleet's manoevrings. After studying its manoevres for some time the enemy admiral might well say: *This fleet's ways of manoevring strongly suggest to me that it is seriously short of internal communications.*

With this idea in mind, I want to examine the question of how to measure the total internal exchanges of information within a system, especially within a dynamic system, such as a fleet or a brain. We might refer to this quantity as the total " turn-over " of information, or even of its informational " metabolism "; but as experience has shown that such picturesque ideas are apt to lead merely to a flood of verbiage of no definite import, a word may be advisable to make sure that our developing ideas have a completely clear and operational basis.

WHAT IS " INFORMATION THEORY " ?

In my opinion, Dr. R. B. BANERJI's [3] suggestion is right: information theory is basically just counting, and simply a branch of combinatorics. When a man says: *You can't get ten manœvres out of that satellite with only eight signals*, he is using the essence of information theory — he is counting the number of distinct causes, the number of distinct effects, and comparing the numbers. Where SHANNON [4] showed his skill was not in inventing a new philosophy or a new mystery but in showing how the counting could be extended into cases that would quite defeat the counting methods of the bank-teller — cases in which the causes were continuous (on a wave-form), cases in which the relevant causes were mixed with irrelevant causes (" noise "), and so on. Thus, if we study the " internal informational exchange " within a system we are really measuring the *quantities of cause and effect* working inside the system. The function $p \log p$ comes into the topic simply because, as SHANNON showed, this function, and it alone, gives numbers that remain proportional to the equivalent number of causes

GENERALISING TO n VARIABLES

The first steps, taken by McGILL [2] and later with GARNER [5] were natural and need not be justified at this stage. I shall show some of their consequences, and hope that these will justify their original decisions. (I do not exclude the possibility that other generalisations may be advisable in other developments.)

I assume that we have before us some well defined set $\mathcal{J}$ of variables — A, B,... I,... N — and that there are n of them. (I shall generally use capitals for sets or variables, and lower case for elements or values.) The n variables might be, for instance, n coordinates specifying the positions of the ships of a fleet, or the temperatures of the air at n places in a country, or the electric potentials at n points on a living brain. Each set of n values gives one *state*, a particular value of the vector or n-tuple $\langle a, b, \ldots i, \ldots n \rangle$. Many such states will provide a frequency table and, in the limit, probabilities. We thus start with an objective basis for computing the entropies, exactly in accordance with SHANNON's original definitions. Thus, H(A) will represent the entropy (scatter, uncertainty, variety, etc.) of variable A with all the other variables ignored, i.e. combined by summation. H(A, B, C) will similarly represent the entropy of the vector $\langle a, b, c \rangle$, and so on. H(A, ... N), which we can write briefly as H($\mathcal{J}$),is the entropy of the whole set of states, with every distinction preserved. $\mathcal{J}$-I will be used to represent the set with I omitted; similarly $\mathcal{J}$-IJ will represent the set with both I and J omitted. Subscripts will be used in SHANNON's way: to show variables that have been held constant (or are assumed known, or otherwise have zero entropy).

First we may notice that $H_{\mathcal{J}\text{-}I}(I)$, which can be found as $H(\mathcal{J})$-$H(\mathcal{J}\text{-}I)$, measures the amount by which the variable I varies (on the entropy scale) when all other variables in the system are held constant. It clearly measures, from the "causal" point of view, how much of I's changes cannot be allocated to any other variable, and that must therefore be allocated to the residue labelled "noise". Thus it measures in a precise sense the "intrinsic noisiness" of the variable I. For any system to be worth investigation, these entropies, for all of A, B, ...N, must be sufficiently small. All are measurable, of course, directly from the observed frequencies.

Following SHANNON and McGILL, we define the "transmission" between any two variables I and J, T(I : J), by

$$T(I : J) = H(I) + H(J) - H(I, J) \qquad (1)$$

This is the transmission with all other variables simply ignored, i.e. lost by being summed to form a two-way table showing only the frequencies of the values of I and J.

$$T(I, J, K : L) = H(I, J, K) + H(L) - H(I, J, K, L) \qquad (2)$$

This transmission is that between variable L and the vector IJK, treated as if it were one variable (of three components).

$$T(A : B : \ldots : N) = H(A) + H(B) + \ldots + H(N) - H(A, B, \ldots N) \qquad (3)$$

This transmission is the "total" transmission between all the variables. It is perhaps the most important quantity of the system, for it measures the total constraint holding over the system (the entropies of the individual variables being regarded as given). For this reason it measures the total quantity of relationships that exist in the system — the total quantity of law, as one might put it. Once an actual system has yielded a primary body of factual data, the "total transmission" computed from this data measures the total quantity of law that can be extracted from the data. Thus it is possible to measure how much law a given body of data contains *before* the particular details of the law (or laws) have been discovered.

The *direct* transmission between I and J

$$T_{\mathcal{J}\text{-}IJ}(I : J) = H_{\mathcal{J}\text{-}IJ}(I) + H_{\mathcal{J}\text{-}IJ}(J) - H_{\mathcal{J}\text{-}IJ}(I, J) \qquad (4)$$

is often interesting since it measures the transmission between I and J when all other variables are held constant. Thus it measures the direct transmission between them. (T(I : J) may include

relations holding between I and J because of their relationships with other, common, variables.)

The "interactions" Q are defined through the transmissions and through the interactions with fewer variables

$$Q(I, J, K) = T_I(J : K) - T(J : K) \qquad (5)$$

$$Q(I, J, \ldots M) = Q_I(J, \ldots M) - Q(J, \ldots M) \qquad (6)$$

Their properties will be discussed after some equations relating these quantities have been given.

All the equations that follow have either been given before (by SHANNON, McGILL, or GARNER), or are readily derivable from the basic definitions by elementary algebraic operations or by using McGILL and GARNER's rule that if an equation is true, it will remain true if every term in it has added to it the same subscript. A great number of equations can be developed; here I shall give only those of outstanding interest.

GROUP I:

$$H(A, B, \ldots N) = H(A) + H_A(B) + H_{AB}(C) + \ldots H_{AB\ldots M}(N) \qquad (7)$$

$$T(A : B : \ldots : N) = T(A : B) + T(A, B : C) + T(A, B, C : D) - \ldots T(A, \ldots M : N) \qquad (8)$$

$$T(A : B : \ldots : N) = \Sigma_{IJ}T(I : J) + \Sigma_{IJK}Q(I, J, K) + \ldots Q(A, B, \ldots N) \qquad (9)$$

$$T_{\mathcal{J}\text{-}AB}(A : B) = T(A : B) + \Sigma_I Q(A, B, I) + \Sigma_{IJ}Q(A, B, I, J) + \ldots Q(A, B, \ldots N) \qquad (10)$$

In these sums, the rule will be used throughout this paper that the sum is to be taken over only the *distinct* forms; those forms that are necessarily identical by symmetry are to be ignored and omitted. Thus, if $\mathcal{J} = \{A, B, C\}$, $\Sigma_{IJ}T(I : J)$ will represent the quantity

$$T(A : B) + T(B : C) + T(A : C) \qquad (11)$$

the identical terms T(B : A), T(C : B), and T(C : A) being omitted.

The equations or expansions (7) to (10) are all of direct interest. They apply (as contrasted with those to be given later) when every variable has essentially individual characteristics, so that the individuality must be sustained. They all show how some total quantity, characteristic of the *whole* system, is built up by the *additive* combination of quantities pertaining to the parts. Thus, equation (7), concerning H(A, B, ... N), shows how the total entropy is related

to entropies obtained by examining the variables one at a time (in some given or natural order). Thus the first contribution H(A), can be obtained experimentally by observation of A alone. $H_A(B)$ can be obtained by controlling only A and by observing only B. And so on.

Another way of looking at such an expansion is to notice that the quantity

$$H(A) + H_A(B) + \ldots + H_{AB\ldots M}(N) - H(A, B, \ldots N) \qquad (12)$$

is a known constant (in fact, zero), so the equation can be used whenever some of the components can be measured easily and others only with difficulty, or perhaps not at all. The easy measurements and the equation will then provide a method for estimating quantities that might otherwise remain unmeasurable. (The law of conservation of energy is used by physicists and engineers incessantly in just this way, the known energies being used to deduce the missing, unknown, energy.)

The next two equations (8) and (9) — relating to the total transmission T(A : B : ... : N) — show how this quantity characterising the whole system can be partitioned into quantities often of direct interest. Thus equation (8) shows how it might be analysed when the two variables A and B are pre-eminent, with the others falling into some natural sequence C, D, ... N. It analyses the whole transmission into that between A and B *plus* that between the subsystem AB and variable C *plus* that between subsystem ABC and variable D, and so on.

Interaction

Equation (9) analyses the total transmission into portions that can be related to different degrees of complexity in the system. First comes $\Sigma T(I : J)$, the sum of all the transmissions between pairs. The next portion, $\Sigma Q(I, J, K)$, is the sum of all the three-variable "interactions" (defined earlier in the paper). Their significance can be seen most clearly if the equation is written out with $n = 3$

$$T(A : B : C) = T(A : B) + T(B : C) + T(A : C) + Q(A, B, C) \qquad (13)$$

Here the interaction Q is clearly identified as that portion of the total transmission which cannot be ascribed to any of the variables acting in pairs. It represents, in other words, the amount of transmission (constraint, law, entropy) that is ascribable only to the

three variables acting as a unique triple. (An example is given below.) It thus measures the degree to which the system (here of three variables) is *irreducibly* complex, i.e. not to be treated by examination of the variables two at a time. Conversely, if Q(A, B, C) is zero, this fact at once tells us that this system's laws may be treated piecemeal, two at a time, and that the total constraint is just the sum of the constraints between the two variables of each pair.

Extension to more variables is now easy. Q(A, B, C, D) measures the degree to which the variables A, B, C, and D, regarded as causes, evoke effects (on one another) that cannot be ascribed to them three-at-a-time, but can be accounted for only by their acting as a unique quadruple.

The numerical values and the distributions of the interactions are so fundamental in the study of any complex system that some further discussion is advisable. As a first property we may notice that Q is a symmetrical function of its arguments, i.e. mere rearrangement of the letters within the parentheses does not alter its value. Q(A, B, C, D) must necessarily have the same numerical value as, say, Q(C, B, D, A), and there is really only one interaction involving the four variables, though the definitions of equations (5) and (6) might suggest the contrary. A simple proof of this statement is given by expressing Q(A, B, ... G) in terms of the basic entropies H

$$Q(A, B, \ldots G) = -H(A, B, \ldots G) + \Sigma H(g\text{-}1) - \Sigma H(g\text{-}2) + \ldots - \Sigma_{11} H(I, J) + \Sigma_1 H(I) \qquad (14)$$

where $\Sigma H(g\text{-}1)$ means the sum of all entropies with g-1 variables taken from the g variables A, B, ... G; and so on. The last two sums range over the same set. As it is known (Shannon [4]) that the H-functions are symmetrical, the function Q is evidently symmetrical also. In this respect, these "interactions" are closely related to those of Fisher's analysis of variance, a resemblance that has been traced in detail by McGill and Garner [5].

If, in equation (9), the system provides numbers that are large in the higher order interactions, then this "information analysis" says, in effect: This system is intrinsically complex, richly connected internally, and must be accepted as complex. Not infrequently, however, sytems are found to essentially simple; the fact will be reflected in the vanishing of the higher order interactions. Thus all who work with large systems, hoping to find among them some that are not as complex as they look, will be specially interested in those sytems that have higher order interactions all zero. What

can be said about such cases ? The subject deserves extensive treatment ; here I shall quote only a few selected facts to indicate the trends.

As a first example, consider the situation in which a hundred families, each of father, mother, son, and daughter, are vacationing at a resort. 400 variables are well defined : the 400 places at which the 400 people may be at any one moment. Let us suppose that within each family there is a good deal of correlation in movement : where Mr. X gives a good deal of information about where Mrs. X is, and where his son and daughter are. Suppose further that the different families are unknown to one another and are wholly independent in their movements. In such a 400-variable system, after many " states " have been observed and the entropies, transmissions, and interactions computed, it will be found that while some of the 4-variable interactions are non-zero (those whose four arguments refer to one family), all interactions of five or more variables are zero. Thus, though the " system " has 400 variables, the fact that it really consists of 100 independent subsystems of four variables each will be reflected in the fact that all interactions Q of five or more variables are zero.

The fact can be expressed more precisely in the following theorems, which are easily proved when one remembers that " independence " (between X and Y, say) corresponds to the quantitative relation

$$H(X) + H(Y) - H(X, Y) = 0$$

Theorem 1.

If the set of arguments of Q(A, B, ... G) can be partitioned into two sets such that all subsets of the one are probabilistically independent of all subsets of the other, then Q must be zero.

Theorem 2.

If a set $\mathcal{J}$ of variables A, B, ... N, is such that no subset of $\mathcal{J}$ can be increased in size beyond k variables without including at least one variable that is wholly independent of the rest of the subset, then all interactions between $k + 1$ or more variables will be zero.

The example illustrates this theorem with $k = 4$. Thus, " systems " that really consist of independent subsystems have zero higher interactions. With this fact we can begin to see something

of the relation between zero interactions and the possibility of breaking an apparently complex system down to simpler systems.

It must not, however, immediately be concluded that vanishing of the higher interactions proves that the system must consist of independent parts. One counter-example will be sufficient. Consider the system of three variables — A, B, and C, each of the two values 0 and 1 only — whose eight states occur with the probabilities

A	B	C	Probability
0	0	0	0
0	0	1	0
0	1	0	0.276906
0	1	1	0.169281
1	0	0	0.138453
1	0	1	0.346133
1	1	0	0.069227
1	1	1	0
			1.000000

Q(A, B, C) = 0 (to the sixth decimal) but a simple inspection shows that no variable can be picked out as independent of the other two.

To obtain further insight into the nature of the interaction functions Q we may next look at an example in which all the " transmission " (constraint) is in the interaction term only. The story goes that three suspected spies — Mr's X, Y, and Z — were kept under observation to see whether they were acting in collusion. Each was found to visit only Antwerp or Berlin or Copenhagen, and it was established that on any one day the three men were likely to be distributed equiprobably over the nine combinations (with A for Antwerp, etc.) :

Mr. X at	A	A	A	B	B	B	C	C	C
Mr. Y at	A	B	C	A	B	C	A	B	C
Mr. Z at	B	C	A	C	A	B	A	B	C

Are their movements really independent ?

If we examine them by pairs, testing for independence between X and Y, for instance, we find the frequencies of positions to be :

Mr. Y at				
	A	1/9	1/9	1/9
	B	1/9	1/9	1/9
	C	1/9	1/9	1/9
		A	B	C
		Mr. X at		

Outcome			*Probability*
A's	B's	C's	
H	H	H	0.405
H	H	T	045
H	T	H	005
H	T	T	045
T	H	H	045
T	H	T	005
T	T	H	045
T	T	T	405
			1.000

When the transmissions and interaction are analysed as before they are found to be

$$\begin{aligned} T(A:B:C) &= T(A:B) + T(B:C) + T(A:C) + Q(A, B, C) \\ 1.062 &= 0.531 + 0.531 + 0.320 - 0.320 \end{aligned} \qquad (16)$$

The interaction is here negative : an arithmetical possibility whose meaning has been discussed by McGILL [2]. We also notice that while A and C have transmissions with B amounting to 0.531 bits per event, due to the obvious effect that B's value has on A's and C's, there is an apparent " transmission " between A and C of 0.320 bits per event, due entirely to their common relation with B. The interaction Q amounts to just (the negative of) this quantity. Thus a three-variable interaction measures (among other things) the amount of *indirect* transmission between variables.

To sum up what has been said so far, it is clear that the " total transmission " over a system can be measured and that it can be related to the transmissions between the parts in various interesting ways. Examination of these relations in any particular case may give a valuable insight into the nature of the processes going on in the system.

UNIFORM VARIABLES

So far, in our treatment of the n variables, we have allowed each one to retain its full individuality. When the system becomes very large, however, the retention becomes unpractical, as would happen if we attempted to predict a society's history by considering every citizen as a person with certain special characteristics, every one of which was too significant to be passed over. Some large systems, however, have variables sufficiently like one another to make averaging and similar processes meaningful. This is certainly the

case when the system is a gas composed of identical molecules, but it may be sufficiently true of many other systems to be worth study and application.

We therefore turn from considering, say, T(A : B) to considering such symmetric functions as the sum of all the two-variable transmissions — $\Sigma_{IJ}T(I : J)$ — or their *mean* value, which will be indicated throughout by an upper bar : $\overline{T}(I : J)$. (The convention mentioned earlier is continued below : sums and means will be based on only those forms that are distinct.) Some of the more interesting equations and expansions, readily obtainable from Group I, are given now.

GROUP II :

$$\Sigma_I T(\mathcal{J}\text{-}I : I) = \Sigma_I H(I) - \Sigma H_{\mathcal{J}\text{-}I}(I) \tag{17}$$

$$\Sigma_{IJ} T_{\mathcal{J}\text{-}IJ}(I : J) = T(A : B : \ldots : N) + \frac{1.4}{2}\Sigma Q_3 + \frac{2.5}{2}\Sigma Q_4 + \ldots + \frac{(n\text{-}2)\ (n+1)}{2} Q_n \tag{18}$$

(where ΣQ_3 means the sum of all distinct three-variable interactions)

$$\Sigma_{IJ} T_{\mathcal{J}\text{-}IJ}(I : J) = \Sigma_{IJ} T(I : J) + \binom{3}{2}\Sigma Q_3 + \binom{4}{2}\Sigma Q_4 + \ldots + \binom{n}{2} Q_n \tag{19}$$

$$\Sigma_{IJ} T_{\mathcal{J}\text{-}IJ}(I : J) = \frac{1}{2}\Sigma_I T(\mathcal{J}\text{-}I : I) + \frac{1.3}{2}\Sigma Q_3 + \frac{2.4}{2}\Sigma Q_4 + \ldots + \frac{n(n\text{-}2)}{2} Q_n \tag{20}$$

The chief interest in these equations lies in the fact that the experimenter can use them so that easily observable variables give estimates of quantities that might be excessively difficult to measure directly. This method is, of course, used everywhere in the classic sciences where, for example, a planet's brightness (easily observable) is used to indicate its mass, or a bacterium's power to make bubbles of gas appear in a solution of lactose is used to indicate whether its body contains a certain enzyme. When some part of the brain is investigated, some variables (e.g. total oxygen consumption, amplitude of potential variation) may be readily observable, while others such as the communicational interchanges between the nerve cells may be most difficult. These equations provide relations by which the readily observable may be made to give (indirect) information about the inaccessible.

Equation (17), for example, says that the total transmissions occurring between variables and the rest of the system is equal to the sum of all the variables' entropies, diminished by all their intrinsic noise. Now the total amounts of transmissions may be difficult to measure ; on the other hand, the entropy of any one variable can often be obtained easily by direct observation of that one variable. If the variables are sufficiently uniform, the sum will be a mere multiple of this quantity. If the intrinsic noisiness is known, or can be assumed to be zero, the equation will at once give an estimate of the transmission represented, which may be of much profounder interest.

The next three equations, (18) to (20), are of special interest as showing some very attractive simplicities if the interactions should really be all zero. We would then have four quantities of outstanding importance all equal, so that knowledge of one would give knowledge of all

$$\begin{aligned} T(A : B : \ldots : N) &= \Sigma_{IJ} T_{\mathcal{J}\text{-}IJ}(I : J) \\ &= \Sigma_{IJ} T(I : J) \\ &= \frac{1}{2}\Sigma_I T(\mathcal{J}\text{-}I : I) \end{aligned} \tag{21}$$

Thus, to take equation (18) first, the total amount of transmission (constraint, law) would be equal to the sum of the direct transmissions between all pairs. In the example of the fleet, mentioned earlier, the direct communications between fifty signallers in pairs would have this relation to the total law, or orderliness, that can be imposed on the fleet as a system. (Any actual application of the equation to such a system would, of course, demand a much closer scrutiny of how the basic mathematical assumptions match the actual conditions holding in the real fleet.)

Equation (19), when the interactions are all zero, shows that the sum of the direct transmissions would equal the sum of the total (pairwise) transmissions. Now the latter can be estimated by direct observation of only two variables at a time, and is thus relatively simple. This measurement, with a suitable multiplier, would then enable one to estimate how much *direct* transmission was occurring — a quantity very difficult to measure in so tangled a system as the brain.

Conversely, if in a society, the activity of the telephone exchanges showed how much direct communication was occurring, equation (19), with zero interactions, would show the total pairwise transmission.

Mean transmissions

Equations relating the mean values cannot, of course, be obtained directly from Group II by just replacing sums by means, for the various sums, even in one equation, differ in their number of terms. Thus the equations for means are often somewhat different from those for sums, so that some relations appear most simply when expressed in means, some when expressed as sums. Group III contains the most interesting results. (The letters representing variables are, of course, all dummies, for they are assumed to run through all distinct combinations in forming the mean.)

Group III :

$$H(\mathcal{J}) = \overline{H}(I) + \overline{H}_I(I) + \overline{H}_{IJ}(I) + \dots + \overline{H}_{\mathcal{J}-I}(I) \qquad (22)$$

$$\overline{T}_{\mathcal{J}-IJ}(I:J) = \overline{T}(I:J) + \binom{n-2}{1}\overline{Q}(I,J,K) + \binom{n-2}{2}\overline{Q}(I,J,K.L) + \dots + Q(A,B,\dots N) \qquad (23)$$

$$T(A:B:\dots:N) = \overline{T}(I:J) + \overline{T}(I,J:K) + \overline{T}(I,J,K:L) + \dots + \overline{T}(\mathcal{J}-I:I) \qquad (24)$$

Let $\overline{T}(A:B:\dots:G)$, where the mean is taken over all possible distinct combinations of g elements drawn from $\mathcal{J}$, be written $\overline{T}_g$; then

$$\overline{T}_g = \binom{g}{1}\overline{T}_2 + \binom{g}{2}\overline{Q}_3 + \binom{g}{3}\overline{Q}_4 + \dots + \overline{Q}_g \qquad (25)$$

$$\Delta^\rho \overline{T}_g = \overline{Q}_\rho + \binom{g}{1}\overline{Q}_{\rho+1} + \binom{g}{2}\overline{Q}_{\rho+2} + \dots + \binom{g}{g-1}\overline{Q}_{\rho+g-1} + \overline{Q}_{\rho+g} \qquad (26)$$

where $\Delta T_i = T_{i+1} - T_i$ by definition.

Equation (22) shows how the entropy of the system as a whole $H(\mathcal{J})$, can be partitioned into components that correspond to the varying amounts of entropy shown, on the average, as 0, 1, 2, ..., n-1 of the variables of the system are fixed. Since these are means they will be estimatable anywhere in tne system if it is, as we are here assuming, everywhere uniform.

Equation (23) shows how the average amount of *direct* variable-to-variable transmission is related to the average total transmission between them (measurable by observations of only two at a time) and the amounts of interactions.

Equation (24), of fundamental importance, shows how the total transmission over the whole system can be partitioned into average effects :

— between variables in a pair,
— between pairs and a third,
— between triples and a fourth, and so on.

Equations (25) and (26) display a fact that should be capable of wide application, and that may give a deep insight into a system's intrinsic complexity. It can be stated formally.

Theorem 3.

If the k-th differences ($k \geq 3$) of the sequence 0, $\overline{T}_2$, $\overline{T}_3$, $\overline{T}_4$ ..., $\overline{T}_{n-1}$, $\overline{T}_n$ are all zero, and if $\overline{Q}_k = 0$, then all interactions involving more than k variables are zero.

The proof is readily obtained by first proving equation (25) and, from it, equation (26). Then put $g = 1, 2, 3, \dots$ in succession and get the equations (with $\overline{T}_1$ formally equal to 0)

$$\Delta^k \overline{T}_1 = 0 = \overline{Q}_k + \overline{Q}_{k+1}$$
$$\Delta^k \overline{T}_2 = 0 = \overline{Q}_k + 2\overline{Q}_{k+1} + \overline{Q}_{k+2}$$
$$\Delta^k \overline{T}_3 = 0 = \overline{Q}_k + 3\overline{Q}_{k+1} + 3\overline{Q}_{k+2} + \overline{Q}_{k+3}$$

et cetera.

From these, with $\overline{Q}_k = 0$, it follows, line by line, that $\overline{Q}_{k+1} = 0$; and so on. (Q. E. D.)

The converse follows directly from equation (26).

Theorem 4.

If all mean interactions involving k or more variables are zero, then the k-th differences of the sequence 0, $\overline{T}_2$, $\overline{T}_3$, ..., $\overline{T}_n$ will all be zero. In particular, the sequence will not increase faster than a polynomial of degree k-1.

Thus systems of small intrinsic complexity (no high order interactions) will be characterized by a relatively slow rate of increase

as the transmission is measured over larger and larger subsets of variables. Thus measurements made in ways selected to be technically simple may give information about communicational aspects that might be excessively difficult if attempted in the direct and obvious way.

Dynamic systems

It must be appreciated that though our examples in the preceding pages have often referred to dynamic systems — to systems actively changing in time, such as the fleet, the nervous system, a society — information theory and its theorems have no direct or natural relation to real time: the user of the theory is entirely free to say how his variables are related, if related at all, to events in real time. In spite of the theory's original application to events in real time, with messages from a sender arriving at some later time at a receiver, the ideas are based only on pairings or *correspondences* of events, with the user free to select the correspondence that shall suit his particular purpose. The transmission, for instance, is defined equal to

$$H(X) + H(Y) - H(X,Y)$$

but is wholly indifferent to when the events X and Y occurred in real time.

An obvious method for introducing real time is simply to let one of the variables, X say, become the real time (clock-reading) C; but this method is basically inappropriate: as a source of information, every signal emitted by the clock after its first two ticks is wholly redundant! A more promising method seems to be as follows.

Every sustained observation of a real dynamic system gives, at first, a primary protocol recording what values the variables took, at what times. Thus, if the variables are $X_1, X_2, \ldots X_n$, and the system, as n-tuple, was observed at times indicated by superscripts, the protocol will consist of an actual value given to each symbol in:

Time	State of system
0	$< x_1^0, \ldots x_n^0 >$
1	$< x_1^1, \ldots x_n^1 >$
...	.....
j	$< x_1^j, \ldots x_n^j >$
...	.....

A most important case occurs when the system is state-determined i.e. when the n-tuple x^{j+1} is the same function of x^j whatever the value of j. The protocol can then be represented equivalently by the single function f

$$x^{j+1} = f(x^j), \text{ or } x' = f(x)$$

When this is so, an important new set of variables, $2n$ in number,

$$< x_1, x_2, \ldots x_n, x_1', x_2', \ldots x_n' >$$

represents the transitions, i.e. the behaviour in real time, one *state* of the new system (of $2n$ variables) corresponding to one *transition* of the old.

Between these new $2n$ variables all the various measures of entropy, transmission, and interaction may be computed exactly as over any other set of variables, but they can now be interpreted by their relations with real time. Thus, $T(X_i : X_i')$ measures how much X_i's next value is dependent on its immediately preceding value. Again, $T(X_i : X_j')$ measures something very close to our naive concept of "cause and effect", for it measures how much the later value at X_j is dependent on the prior value at X_i. If the transmission between the same two is also found when all of $X_1, \ldots X_n$ are constant (except for X_i), then this new number measures the degree to which X_j is *directly* affected by X_i.

If larger sets are studied

$$< x_1^0, \ldots x_n^0, x_1^1, \ldots x_n^1, \ldots x_1^j, \ldots x_n^j >$$

such a transmission as $T(X_i^0 : X_i^j)$ would measure how much the variable X_i shows j steps later in time, the effect of its value earlier. The measure thus catches something essential in the concept of X_i's "memory". This method thus treats communications across gaps in space (between two of the n variables) and across gaps in time ("memory" effects) by wholly uniform concepts and methods.

The subject has yet to be extensively developed, but there seems to be good reason for believing that these measures may offer a method for getting a deep insight into such really complicated systems as the brain, the biochemistry of the cell, and the economics of modern society. The reader will have noticed that most of the labor is mere routine, and is thus eminently suitable for delegation to the modern computer.

SAMPLING VARIATIONS

It is clear that any use of these methods on actual data must be undertaken only when one is prepared with some knowledge of what may occur under the vagaries of random sampling. The subject has been discussed by MILLER [6].

CONCLUSION

Information theory started by studying the relation between two variables — sender and receiver — but it can readily be generalised to study the relations between any number of variables. Such a generalisation would be useful for studying the total internal exchanges of information between the parts of a large computer, or between the cells of a brain, or between the members of a large society.

The method is outlined and some basic equations given. When the systems are of mostly similar parts, average values become applicable and have special properties, some of which are tabulated.

A particularly attractive feature of the method is its ready separation of what is simple from what is intrinsically complicated in the system. Thus if the system has hidden simplicities, the method provides a possible way of finding them.

REFERENCES

[1] BREMERMANN, H. J., *Optimisation through evolution and re-combination* in *Self-organizing systems*. Editors M. C. Yovits, G. T. Jacobi and G. D. Goldstein, Spartan Books, Washington, 1962, pp. 93-106.

[2] MCGILL, W. J., *Multivariate information transmission*. Psychometrika, **19**, 97-116, 1954.

[3] BANERJI, R. B., (Personal communication).

[4] SHANNON, C. E. and WEAVER, W., *The mathematical theory of communication*. University of Illinois Press, Urbana, Illinois, 1949.

[5] GARNER, W. R. and MCGILL, W. J., *The relation between information and variance analyses*. Psychometrika, **21**, 219-228, 1956.

[6] MILLER, G. A., *On the bias of information estimates*, in *Information theory in psychology*. Editor H. Quastler, The Free Press, Glencoe, Illinois, 1956.

TWO TABLES OF IDENTITIES GOVERNING INFORMATION FLOWS WITHIN LARGE SYSTEMS

W. Ross ASHBY

Department of Electrical Engineering and Department of Biophysics

University of Illinois, Urbana, Illinois

Abstract

The study of information flows in large dynamic systems (during coordinated action, for instance) has led to the development of a number of identities, useful from many points of view. Two tables of those now available are presented in the hope that other workers in the subject may find them of use.

Shannon's original presentation of communication theory dealt with only two variables: Sender and Receiver, with Noise, perhaps, as an occasional third. Since then, however, following McGill [1], and Garner [2], the extension to n variables has been made by Ashby [3], Powers [4], and Conant [5]. The extension has been found especially useful in treating questions of the flows of information within a system when the system, large and complex, is undertaking activities that require, between the various parts, a major degree of coordination. One example would be the coordination of traffic during rush-hour, another would be the activities in the brain of a tight-rope walker.

Such studies frequently call for the manipulation of various identities that must hold over the various informational quantities – entropies, transmissions, interactions. As a number of such identities have now been developed, a collection of all the major forms is given, in the hope that the collection may be of use to other workers.

The sets of variables considered may, of course, be considered as existing separately; they may also be regarded as referring to the values taken, by one or more basic variables, at various *times* during a sequence. Conant [5] has considered this case, and has shown the importance of their average values as the sequence is extended indefinitely. He has proved that every identity in H, T, and Q (defined below) remains an identity in H^L, T^L, and Q^L (the limiting averages) of identical form. Thus the Tables given here give also, if the reader cares to imagine an L superscript throughout, an equal-sized set applicable to these limiting averages. Though different in interpretation, they seem hardly worth printing twice; so the extension is left to the reader.

Notation

It has been selected with care, to be as informative and suggestive as possible; but what suits one research well may conflict with another. A reader in the latter predicament may find a rewriting of the identities advantageous. The attempt to classify these identities has shown that every arrangement, no matter how obvious from one point of view, is absurd from some other. I have chosen the arrangement chiefly for ease of reference.

Capital letters have been used to represent variables, as each variable is essentially a *set* (of possible values). This approach emphasizes that the variables in these identities, though able to be the usual real numbers, are in no way restricted to them. They need not in fact have either metric or order. This freedom is specially desirable in mathematics having applications in biology [6], for many important variables in biology are naturally of classificatory type: which species of bacterium?–in which type of neuroglial cell?–oxidizing which substrate?–after giving which stimulus?

Definitions

(Given here to exclude any possibility of ambiguity) All start with either a set of probabilities or a table of observed frequencies (according to whether one is analyzing the information flow to be expected on some theory or analyzing events that have been observed in an experiment).

The work on which this report is based was supported by the Air Force Office of Scientific Research under Grant AF-AFOSR 7-67.

If the variable X takes its values $x_1, x_2, \ldots$ with probabilities $P_1, P_2, \ldots$ (summing to 1) then by definition:

$$H(X) = \Sigma_i P_i \log \frac{1}{P_i}.$$

If one works with observed frequencies, so that the values $x_1, x_2, \ldots$ occurred $k_1, k_2, \ldots$ times, (summing to k) then, by definition:

$$H(X) = \frac{1}{k}(k \log k - \Sigma_i k_i \log k_i).$$

(If the ratios k_1/k, etc., are regarded as corresponding to the probabilities P_1, etc., then the same numerical value for H(X) is given by both expressions.)

If the two variables X and Y take their conjoint values $< x_i, y_j >$ with probabilities P_{ij}, with $\Sigma_{ij}P_{ij} = 1$, then by definition:

$$H(X,Y) = \Sigma_{ij} P_{ij} \log \frac{1}{P_{ij}}.$$

With frequencies, if the event $< x_i, y_j >$ occurred k_{ij} times, with $\Sigma_{ij}k_{ij} = k$, then by definition:

$$H(X,Y) = \frac{1}{k}(k \log k - \Sigma_{ij} k_{ij} \log k_{ij}).$$

The extension to n variables A, B, . . . , N , is immediate. If the event $< a_i, b_j, \ldots, n_m >$ has probability $P_{ij\ldots m}$ (summing to 1), then by definition:

$$H(A, B, \ldots, N) = \Sigma_{ij\ldots m} P_{ij\ldots m} \log \frac{1}{P_{ij\ldots m}}.$$

If the same event occurred with frequency $k_{ij\ldots m}$, and $\Sigma_{ij\ldots m} k_{ij\ldots m} = k$, then, by definition:

$$H(A, B, \ldots, N) = \frac{1}{k}(k \log k - \Sigma_{ij\ldots m} k_{ij\ldots m} \log k_{ij\ldots m}).$$

If the logarithms are to base 2, the entropies will be in bits per single event.

From these basic quantities all the others may be defined:

$H_X(Y) = H(X,Y) - H(X)$

$H_{XYZ}(D,E,F,G) = H(X,Y,Z,D,E,F,G) - H(X,Y,Z)$ etc.

The transmissions T:

$T(X:Y) = H(X) + H(Y) - H(X,Y)$

$T(X:Y:Z) = H(X) + H(Y) + H(Z) - H(X,Y,Z)$

etc. (repeated for convenience in the Tables at (2).)

The conditional transmissions such as $T_Z(X:Y)$ - the average transmission between X and Y when Z is kept constant--are defined and similarly, using conditional entropies, e.g., by definition:

$T_Z(X:Y) = H_Z(X) + H_Z(Y) - H_Z(X,Y)$.

A variable such as X may itself be a vector, $X = < E,F >$ say; then one must keep distinct the transmissions $T(< E,F > :Y)$ and T(E:F:Y), for they have different properties and interpretations.

In the Tables I have used commas to separate the components of a vector, and colons otherwise (and often over-lining, $\overline{A,B,C}$, to make the distinction unmistakable. The entropies H do not require the distinction).

The interactions Q are defined in the Table at (23) and (26). Here, too, the notation must show whether two or more variables are entering as a vector or not. As with the T's, commas separate the components of a vector.

In my earlier work, I followed Garner /2/ in writing U for Uncertainty. However, I became so tired of writing U before every term that I fell into the more convenient habit of writing one U at the left of the line to do duty for all of them. This usage, however, showed that the letter was doing no work. I then changed to the more suggestive notation of using H for entropies (continuing Shannon's usage), with T for transmissions, and Q for interactions (as they are so different from the others). In five years I have found the notation to work well, so it is used here.

In regard to the details, the set of variables { A,B, . . . ,G } is represented as an unordered set by A. A will represent a vector, with components $< A,B, \ldots ,G >$ in that order, only when over-lined: $\overline{A}$. A has g variables. The set A-A has the g-1 variables { B, . . . ,G }, etc. I,J, . . are dummy variables used only for summing. α is any unordered subset of A; it is to be interpreted as a vector, e.g., $< \overline{B,C,G} >$, only when over-lined: $\overline{\alpha}$. The frequently wanted total transmission T(A:B: . . . :G) can unambiguously be represented by $T(A)$, – a transmission with one variable would be meaningless. Similarly for T(A-A), which is T(B:C: . . . :G), and similarly for Q.

Table of Identities

Entropies

1. $H(A,B, \ldots ,G) = H(A) + H_A(B) + H_{AB}(C) + \ldots + H_{A\ldots F}(G).$

Transmissions

2. $T(A:B: \ldots :G) = H(A) + \ldots + H(G) - H(A, \ldots ,G).$

3. $= T(A:B) + T(\overline{A,B}:C) + T(\overline{A,B,C}:D) + \ldots + T(\overline{A, \ldots ,F}:G).$

4. $= T(A:G) + T_A(B:G) + \ldots + T_{A\ldots D}(E:G) + T_{A\ldots E}(F:G) + T(A:F) + T_A(B:F) + \ldots + T_{A\ldots D}(E:F) + \ldots + T(A:C) + T_A(B:C) + T(A:B).$

5. $= T(F:G) + T_F(E:G) + \ldots + T_{C\ldots F}(B:G) + T_{B\ldots F}(A:G) + T(E:F) + T_E(D:F) + \ldots + T_{B\ldots E}(A:F) + \ldots + T(B:C) + T_B(A:C) + T(A:B),$ (with many other forms possible).

6. $= \sum_{IJ\in A} T(I:J) + \sum_{IJK\in A} Q(I:J:K) + \sum_{IJKL\in A} Q(I:J:K:L) + \ldots + Q(A: \ldots :G).$

7. $= \sum_{IJ\in A} T_{A-IJ}(I:J) - \frac{1.4}{2} \sum_{IJK\in A} Q(I:J:K) - \frac{2.5}{2} \sum_{IJKL\in A} Q(I:J:K:L) - \ldots - \frac{(g-1)(g+2)}{2} Q(A: \ldots :G).$

8. $T(A: \ldots :D:E: \ldots :G) = T(A: \ldots :D) + T(\overline{A, \ldots ,D}:E: \ldots :G).$

9. $T(A_1: \ldots :A_g:B_1: \ldots :B_g) = T(A_1: \ldots :A_g) + \Sigma_{i=1}^{g} T(\overline{A_1, \ldots ,A_g}:B_i) + T_{A_1, \ldots ,A_g}(B_1: \ldots :B_g)$

10. $T(A:B: \ldots :G:Z) = T(A:B: \ldots :G) + T(\overline{A, \ldots ,G}\ Z).$

11. $= T(A:Z) + \ldots + T(G:Z) + T_Z(A: \ldots :G).$

Transmission between vectors

12. $T(\overline{A_1, \ldots ,A_g}:\overline{B_1, \ldots ,B_h}:\overline{C_1, \ldots ,C_j}$: etc.)
$= \Sigma\, T(P_u:R_v)$ (all pairs, but no two with the same capital letter)
$+ \Sigma\, Q(L_u:M_v:N_w)$ (all triples, but no three all with the same capital letter)
$+ \ldots\ldots\ldots\ldots\ldots\ldots$
$+ Q(A_1: \ldots :A_g:B_1: \ldots :B_h:C_1: \ldots :C_j$: etc.).

As a special case:

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

14. $= \sum_{I\in A} T(I:Z) + Q(A:B:Z) + Q(\overline{A,B}:C:Z) + Q(\overline{A,B,C}:D:Z) + \ldots + Q(\overline{A, \ldots ,F}:G:Z).$

15. $= T(A:Z) + T_A(B:Z) + T_{AB}(C:Z) + \ldots + T_{A\ldots F}(G:Z).$

16. $= \sum_{I\in A} T_{A-I}(I:Z) - \sum_{IJ\in A} Q(I:J:Z) - 2 \sum_{IJK\in A} Q(I:J:K:Z) - \ldots - (g-1)Q(A: \ldots :G:Z).$

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

As a special case:

18. $T(\overline{A,B}:U: \ldots :Z) = T(U: \ldots :Z) + T(A:\overline{U, \ldots ,Z}) + T(B:\overline{U, \ldots ,Z}) + Q(A:B:\overline{U, \ldots ,Z}).$

19. $T(\overline{A_1, \ldots ,A_g}:\overline{B_1, \ldots ,B_h}) = \sum_{i=1}^{g} \sum_{j=1}^{h} T_{A_1, \ldots ,A_{i-1},B_1, \ldots ,B_{j-1}}(A_i:B_j).$

As a special case:

20. $T(\overline{A_1,A_2}:\overline{B_1,B_2}) = T(A_1:B_1) + T_{A_1}(A_2:B_1) + T_{B_1}(A_1:B_2) + T_{A_1B_1}(A_2:B_2).$

Conditional transmissions

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

22. $T_A(Y:Z) = T(Y:Z) + \sum_{I\in A} Q(I:Y:Z) + \sum_{IJ\in A} Q(I:J:Y:Z) + \ldots + Q(A: \ldots :G:Y:Z).$

Interactions

23. $Q(A:B:C) = T_A(B:C) - T(B:C)$ - as definition.

If the variable X takes its values $x_1, x_2, \ldots$ with probabilities $P_1, P_2, \ldots$ (summing to 1) then by definition:

$$H(X) = \Sigma_i P_i \log \frac{1}{P_i}.$$

If one works with observed frequencies, so that the values $x_1, x_2, \ldots$ occurred $k_1, k_2, \ldots$ times, (summing to k) then, by definition:

$$H(X) = \frac{1}{k}(k \log k - \Sigma_i k_i \log k_i).$$

(If the ratios k_1/k, etc., are regarded as corresponding to the probabilities P_1, etc., then the same numerical value for H(X) is given by both expressions.)

If the two variables X and Y take their conjoint values $< x_i, y_j >$ with probabilities P_{ij}, with $\Sigma_{ij} P_{ij} = 1$, then by definition:

$$H(X,Y) = \Sigma_{ij} P_{ij} \log \frac{1}{P_{ij}}.$$

With frequencies, if the event $< x_i, y_j >$ occurred k_{ij} times, with $\Sigma_{ij} k_{ij} = k$, then by definition:

$$H(X,Y) = \frac{1}{k}(k \log k - \Sigma_{ij} k_{ij} \log k_{ij}).$$

The extension to n variables A, B, . . . , N , is immediate. If the event $< a_i, b_j, \ldots, n_m >$ has probability $P_{ij \ldots m}$ (summing to 1), then by definition:

$$H(A, B, \ldots, N) = \Sigma_{ij \ldots m} P_{ij \ldots m} \log \frac{1}{P_{ij \ldots m}}.$$

If the same event occurred with frequency $k_{ij \ldots m}$, and $\Sigma_{ij \ldots m} k_{ij \ldots m} = k$, then, by definition:

$$H(A, B, \ldots, N) = \frac{1}{k}(k \log k - \Sigma_{ij \ldots m} k_{ij \ldots m} \log k_{ij \ldots m}).$$

If the logarithms are to base 2, the entropies will be in bits per single event.

From these basic quantities all the others may be defined:

$H_X(Y) = H(X,Y) - H(X)$

$H_{XYZ}(D,E,F,G) = H(X,Y,Z,D,E,F,G) - H(X,Y,Z)$ etc.

The transmissions T:

$T(X:Y) = H(X) + H(Y) - H(X,Y)$

$T(X:Y:Z) = H(X) + H(Y) + H(Z) - H(X,Y,Z)$

etc. (repeated for convenience in the Tables at (2).)

The condition of transmissions such as $T_Z(X:Y)$ - the average transmission between X and Y when Z is kept constant--are found similarly, using conditional entropies, e.g., by definition:

$T_Z(X:Y) = H_Z(X) + H_Z(Y) - H_Z(X,Y)$.

A variable such as X may itself be a vector, $X = < E,F >$ say; then one must keep distinct the transmissions $T(< E,F > : Y)$ and $T(E:F:Y)$, for they have different properties and interpretations.

In the Tables I have used commas to separate the components of a vector, and colons otherwise (and often over-lining, $\overline{A,B,C}$, to make the distinction unmistakable. The entropies H do not require the distinction).

The interactions Q are defined in the Table at (23) and (26). Here, too, the notation must show whether two or more variables are entering as a vector or not. As with the T's, commas separate the components of a vector.

In my earlier work, I followed Garner /2/ in writing U for Uncertainty. However, I became so tired of writing U before every term that I fell into the more convenient habit of writing one U at the left of the line to do duty for all of them. This usage, however, showed that the letter was doing no work. I then changed to the more suggestive notation of using H for entropies (continuing Shannon's usage), with T for transmissions, and Q for interactions (as they are so different from the others). In five years I have found the notation to work well, so it is used here.

In regard to the details, the set of variables { A,B, . . . ,G } is represented as an unordered set by A. A will represent a vector, with components $< A,B, \ldots, G >$ in that order, only when over-lined: $\overline{A}$. A has g variables. The set A-A has the g-1 variables { B, . . . ,G } , etc. I,J, . . . are dummy variables used only for summing. α is any unordered subset of A; it is to be interpretated as a vector, e.g., $< \overline{B,C,G} >$, only when over-lined: $\overline{\alpha}$. The frequently wanted total transmission $T(A:B: \ldots :G)$ can unambiguously be represented by $T(A)$, – a transmission with one variable would be meaningless. Similarly for $T(A\text{-}A)$, which is T(B:C: . . . :G), and similarly for Q.

Table of Identities

Entropies

1: $H(A,B, \ldots ,G) = H(A) + H_A(B) + H_{AB}(C) + \ldots + H_{A \ldots F}(G).$

Transmissions

2: $T(A:B: \ldots :G) = H(A) + \ldots + H(G) - H(A, \ldots ,G).$

3: $= T(A:B) + T(\overline{A,B}:C) + T(\overline{A,B,C}:D) + \ldots + T(\overline{A, \ldots ,F}:G).$

4: $= T(A:G) + T_A(B:G) + \ldots + T_{A \ldots D}(E:G) + T_{A \ldots E}(F:G) + T(A:F) + T_A(B:F) + \ldots + T_{A \ldots D}(E:F) + \ldots + T(A:C) + T_A(B:C) + T(A:B).$

5: $= T(F:G) + T_F(E:G) + \ldots + T_{C \ldots F}(B:G) + T_{B \ldots F}(A:G) + T(E:F) + T_E(D:F) + \ldots + T_{B \ldots E}(A:F) + \ldots + T(B:C) + T_B(A:C) + T(A:B),$ (with many other forms possible).

6: $= \sum_{IJ \in A} T(I:J) + \sum_{IJK \in A} Q(I:J:K) + \sum_{IJKL \in A} Q(I:J:K:L) + \ldots + Q(A: \ldots :G).$

7: $= \sum_{IJ \in A} T_{A\text{-}IJ}(I:J) - \frac{1.4}{2} \sum_{IJK \in A} Q(I:J:K) - \frac{2.5}{2} \sum_{IJKL \in A} Q(I:J:K:L) - \ldots - \frac{(g-1)(g+2)}{2} Q(A: \ldots :G).$

8: $T(A: \ldots :D:E: \ldots :G) = T(A: \ldots :D) + T(\overline{A, \ldots ,D} : E : \ldots : G).$

9: $T(A_1: \ldots :A_g:B_1: \ldots :B_g) = T(A_1: \ldots :A_g) + \sum_{i=1}^{g} T(\overline{A_1, \ldots ,A_g}.B_i) + T_{A_1, \ldots, A_g}(B_1: \ldots :B_g)$

10: $T(A:B: \ldots :G:Z) = T(A:B: \ldots :G) + T(\overline{A, \ldots ,G} \; Z).$

11: $= T(A:Z) + \ldots + T(G.Z) + T_Z(A: \ldots :G).$

Transmission between vectors

12: $T(\overline{A_1, \ldots, A_g} : \overline{B_1, \ldots, B_h} : \overline{C_1, \ldots, C_j} : \text{etc.})$
$= \Sigma\, T(P_u:R_v)$ (all pairs, but no two with the same capital letter)
$+ \Sigma\, Q(L_u:M_v.N_w)$ (all triples, but no three all with the same capital letter)
$+ \ldots\ldots\ldots\ldots\ldots\ldots$
$+ Q(A_1: \ldots :A_g:B_1: \ldots :B_h:C_1: \ldots :C_j: \text{etc.}).$

As a special case:

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

14: $= \sum_{I \in A} T(I:Z) + Q(A:B:Z) + Q(\overline{A,B}:C:Z) + Q(\overline{A,B,C}:D:Z) + \ldots + Q(\overline{A, \ldots ,F}:G:Z).$

15: $= T(A:Z) + T_A(B:Z) + T_{AB}(C:Z) + \ldots + T_{A \ldots F}(G:Z).$

16: $= \sum_{I \in A} T_{A\text{-}I}(I:Z) - \sum_{IJ \in A} Q(I:J:Z) - 2 \sum_{IJK \in A} Q(I:J:K:Z) - \ldots - (g-1)Q(A: \ldots :G:Z).$

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

As a special case:

18: $T(\overline{A,B}:U: \ldots :Z) = T(U: \ldots :Z) + T(A:\overline{U, \ldots ,Z}) + T(B:\overline{U, \ldots ,Z}) + Q(A:B:\overline{U, \ldots ,Z}).$

19: $T(\overline{A_1, \ldots ,A_g} : \overline{B_1, \ldots ,B_h}) = \sum_{i=1}^{g} \sum_{j=1}^{h} T_{A_1, \ldots, A_{i-1}, B_1, \ldots, B_{j-1}}(A_i:B_j).$

As a special case:

20: $T(\overline{A_1,A_2} : \overline{B_1,B_2}) = T(A_1:B_1) + T_{A_1}(A_2:B_1) + T_{B_1}(A_1:B_2) + T_{A_1B_1}(A_2:B_2)$

Conditional transmissions

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

22: $T_A(Y:Z) = T(Y:Z) + \sum_{I \in A} Q(I:Y:Z) + \sum_{IJ \in A} Q(I:J:Y:Z) + \ldots + Q(A: \ldots :G:Y:Z).$

Interactions

23: $Q(A:B:C) = T_A(B:C) - T(B:C)$ - as definition.

24: $= T(A:B:C) - T(A:B) - T(B:C) - T(C:A)$
(in symmetrical form).

25: $Q(A:B:C:D) = T(\overline{A,B:C,D}) - T_A(B:D) - T_B(A:C) - T_C(A:D) - T_D(B:C)$
(and other symmetrical forms).

26: $Q(A:\ldots:F:G) = Q_G(A:\ldots:F) - Q(A:\ldots:F)$,
as definition,

27: $= -H(A) + \sum_{I\in A} H(A\text{-}I) - \sum_{IJ\in A} H(A\text{-}IJ) + \ldots - \ldots + (-1)^{s-1} \sum_{IJ\in A} H(I,J) + (-1)^s \sum_{I\in A} H(I).$

28: $= T(A) - \sum_{I\in A} T(A\text{-}I) + \sum_{IJ\in A} T(A\text{-}IJ) - \ldots + \ldots + (-1)^s \sum_{IJ\in A} T(I:J).$

29: $Q(A:\ldots:G:Z) = -H_A(Z) + \sum_{I\in A} H_{A\text{-}I}(Z) - \sum_{IJ\in A} H_{A\text{-}IJ}(Z) + \ldots - \ldots + (-1)^{s+1} H(Z).$

30: $= T(\bar{A}:Z) - \sum_{I\in A} T(\overline{A\text{-}I}:Z) + \sum_{IJ\in A} T(\overline{A\text{-}IJ}:Z) - \ldots + (-1)^{s-1} \sum_{I\in A} T(I:Z)$, a special case of:

31: $Q(A:\ldots:G:U:\ldots:Z) = \sum_{\substack{\alpha\subseteq A \\ \zeta\subseteq Z}} (-1)^k T(\bar{\alpha}:\bar{\zeta})$, ($\subseteq$ is a subset of)

where $Z = \{U, \ldots, Z\}$, and $\bar{\alpha}$ and $\bar{\zeta}$ are all subsets of A and Z, as vectors. k is the number of elements of $\{A, \ldots, G, U, \ldots, Z\}$ not occurring in T's parentheses.

32: $= Q_A(U:\ldots:Z) - \sum_{I\in A} Q_{A\text{-}I}(U:\ldots:Z) + \sum_{IJ\in A} Q_{A\text{-}IJ}(U:\ldots:Z) - \ldots + (-1)^{s-1} \sum_{I\in A} Q_I(U:\ldots:Z) + (-1)^s Q(U:\ldots:Z).$

33: $Q(A:\ldots:G:Y:Z) = T_A(Y:Z) - \sum_{I\in A} T_{A\text{-}I}(Y:Z) + \ldots - \ldots + \ldots + (-1)^s T(Y:Z).$

34: $= Q(\bar{A}:Y:Z) - \sum_{I\in A} Q(\overline{A\text{-}I}:Y:Z) + \sum_{IJ\in A} Q(\overline{A\text{-}IJ}:Y:Z) - \ldots + (-1)^{s-1} \sum_{I\in A} Q(I:Y:Z).$

Interaction with vectors

35: $Q(\overline{A,B}:Y:Z) = Q(A:B:Y:Z) + Q(A:Y:Z) + Q(B:Y:Z)$,

a special case of:

36: $Q(\overline{A,\ldots,G}:Y:Z) = \sum_{I\in A} Q(I:Y:Z) + \sum_{IJ\in A} Q(I:J:Y:Z) + \ldots + Q(A:\ldots:G:Y:Z).$

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

Conditional interactions

38: $Q_A(U:\ldots:Z) = Q(U:\ldots:Z) + \sum_{I\in A} Q(I:U:\ldots:Z) + \sum_{IJ\in A} Q(I:J:U:\ldots:Z) + \ldots + Q(A:\ldots:G:U:\ldots:Z).$

Compact Forms

A number of the equations can be written in a form suggesting a useful algebra. α takes as its values all the 2^s subsets of A, though with some values ignored as giving obviously meaningless terms. $Q(\alpha)$ and $Q(\alpha{:}Z)$, when reduced to two variables (by α having two or one element respectively), may be interpreted as $T(\alpha)$ or $T(\alpha{:}Z)$ respectively. When α becomes the empty set $\{\}$, the interpretations of such expressions as $H\{\}(X)$ and $Q(\{\}:Y:Z)$ are most simply found by comparison with the form written explicitly in its fewest variables.

m below is the number of variables in $A - \alpha$, i.e., the number missing from A.

Many of these forms clearly fall into pairs, e.g.,

6c $\quad T(A) = \Sigma_\alpha Q(\alpha)$

28c $\quad Q(A) = \Sigma_\alpha (-1)^m T(\alpha)$

Similarly paired are 13 and 30, 36 and 34, 22 and 33, 38 and 32. By treating (e.g., in 6c) $\sum_{I\in A} T(A\text{-}I)$ as analogous to $\binom{s}{1} T^{s-1}$, $\sum_{IJ\in A} T(A\text{-}IJ)$ as analogous to $\binom{s}{2} T^{s-2}$, a Blissard-type calculus (e.g., Riordan/7/)is possible, enabling one to use the well known rule in the calculus of finite differences relating powers of $\Delta + 1$ and of E-1, and the inversion of each to give the other. This calculus may give interesting developments in the future; here it is useful chiefly as a check on the accuracy of the Tables, where the identities were first found independently.)

Table of Compact Forms

6c: $T(A) = \sum_{\alpha\subseteq A} Q(\alpha)$. ($\subseteq$ is a subset of)

13c: $T(\bar{A}:Z) = \sum_{\alpha\subseteq A} Q(\alpha:Z)$.

17c: $T(\bar{A}:Z) = T(Z) + \sum_{\alpha\subseteq A} Q(\alpha:\bar{Z})$. ($\underline{\subseteq}$ ital., a set)

22c: $T_A(Y:Z) = \sum_{\alpha\subseteq A} Q(\alpha:Y:Z)$.

27c: $Q(A) = \sum_{\alpha\subseteq A} (-1)^{m+1} H(\alpha)$.

28c: $= \sum_{\alpha\subseteq A} (-1)^m T(\alpha)$.

29c: $Q(A:Z) = \sum_{\alpha\subseteq A} (-1)^{m+1} H_\alpha(Z)$.

30c: $= \sum_{\alpha\subseteq A} (-1)^m T(\bar{\alpha}:Z)$.

31c: $Q(A:Z) = \sum_{\substack{\alpha\subseteq A \\ \zeta\subseteq Z}} (-1)^k T(\bar{\alpha}:\bar{\zeta})$. ($\underline{\subseteq}$ ital., a set)

32c: $= \sum_{\alpha\subseteq A} (-1)^m Q_\alpha(Z)$. (ditto)

33c: $Q(A:Y:Z) = \sum_{\alpha\subseteq A} (-1)^m T_\alpha(Y:Z)$.

34c: $= \sum_{\alpha\subseteq A} (-1)^m Q(\bar{\alpha}:Y:Z)$.

36c: $Q(\bar{A}:Y:Z) = \sum_{\alpha\subseteq A} Q(\alpha:Y:Z)$.

38c: $Q(A:Z) = \sum_{\alpha\subseteq A} Q(\alpha:Z)$. ($\underline{\subseteq}$ ital., a set)

Proofs

The indications given below will be sufficient. Many of the identities are derivable from the others: the routes of deduction have been examined to make sure that no circularities have been included.

1: By repeated use of the definition of $H_X(Y)$.
2: By definition.
3: By repeated use of (10).
4: By (3) and then (15).
5: As (4), using another sequence. Yet further sequences are, of course, possible.
6: Given by McGill /1/ and also by Fano /8/. Or it can be proved by reduction of both sides to H's through (2) and (27).
7: In (6), replace each T(I:J) by T(I:J) subscripted, using (22).
8: By reduction to H's through (2).
9: By reduction to H's through (2).
10: A special case of (8).
11: By reduction to H's through (2).
12: By induction from (13).
13: By induction from $T(\overline{X,Y}:Z) = T(X:Z) + T(Y:Z) + Q(X:Y:Z)$ (which can be verified by reduction to H's). Then make X a vector and use (36).
14: Expand the terms of (15) by using (23) repeatedly with A a vector.
15: By reduction to H's through (2), adding subscripts throughout where necessary.
16: Expand the terms of (13) by (23) as $T(I:Z) = T_{A\text{-}I}(I:Z) - Q(\overline{A\text{-}I}:I:Z)$; then use (36) on each of these Q's.
17: By (10) the left side equals $T(\overline{A,\ldots,G}:\overline{U,\ldots,Z}) + T(U:\ldots:Z)$, with (10)'s Z here representing $\overline{A,\ldots,G}$; then use (13) to expand the first term.
18: A special case of (17).
19: By using (15) repeatedly.
20: A special case of (19).
21: From (6) subscripted throughout with Z, subtract (6) unsubscripted.
22: It is $T(Y:Z) + Q(\bar{A}:Y:Z)$. Then use (36).
23: By definition.
24: By reduction to H's through (2).
25: By (26) and then reduction to H's through (2).
26: By definition.
27: By reduction to H's, through (26) (subscripted as required), (23), and (2).
28: By induction from (24).
29: By reduction to H's, using $H_X(Y)$'s definition.
30: A special case of (31).
31: By reduction to H's through (27).
32: By induction on (26).
33: From (23), re-written as $Q(A:Y:Z)$, subtract (23) subscripted with B. Repeat with C, etc.
34: From (33), subtract the zero quantity $(1-1)^s T(Y:Z)$, written out as a binomial expansion, and use (23).
35: By reduction to H's through (27).
36: By induction on (35).
37: (35) and (26) give $Q(\overline{A,B}:Y:Z) = Q(A:Y:Z) + Q_A(B:Y:Z)$. Then use induction.
38: By induction from (26), written as $Q_G(A:\ldots:F) = Q(A:\ldots:F) + Q(A:\ldots:F:G)$.

References

1. McGill, W. J., "Multivariate Information Transmission," *Psychometrika*, 19, 97-116 (1954).

2. Garner, W. R., *Uncertainty and Structure as Psychological Concepts*, John Wiley & Sons, New York (1962).

3. Ashby, W. R., "Measuring the Internal Informational Exchange in a System," *Cybernetica*, 8, 5-22 (1965).

4. Powers, S. G., *Uncertainty Analysis in Dynamic Systems*, B.C.L. Report 8.0, Biological Computer Laboratory, University of Illinois, Urbana (1967).

5. Conant, R. C., *Information Transfer in Complex Systems, with Applications to Regulation*, Doctoral Thesis, University of Illinois, Urbana (1968).

6. Ashby, W. R., "The Set Theory of Mechanism and Homeostasis," *Automation Theory and Learning Systems*, D. J. Stewart (ed.), Academic Press, London, 23-51 (1967).

7. Riordan, J., *Combinatorial Identities*, John Wiley & Sons, New York (1968).

8. Fano, R. M., *Transmission of Information*, John Wiley & Sons, New York (1961).

III.

INFORMATIONAL LIMITS

INFORMATIONAL LIMITS

INTRODUCTION

One kind of informational limit is mentioned briefly in many of Ashby's papers and most explicitly in "Some Consequences of Bremermann's Limit..." It is the limit on practical computability imposed by physical laws, which has as a consequence that question-answering procedures requiring more than about 10^{70} bits are in fact unanswerable. Ashby shows by illustration that apparently simple and harmless questions, usually of a combinatorial nature, can far exceed this limit. More important, he shows that the limit has philosophical implications - for one, "our achieved science will always be one of the world in its simpler interactions. If there are complex natural laws, we shall never know them."

Especially in later papers Ashby stressed this theme repeatedly. There is, however, a common interpretation of Bremermann's Limit which is unnecessarily pessimistic, to wit: "If a question involves selection of an element out of a set of $2^{10^{70}}$ elements or more, then to answer it requires 10^{70} bits of information and thus Bremermann's Limit tells us that the question is unanswerable." This interpretation is wrong [105]. It is indeed impossible if the method used is 10^{70} dichotomies (at 1 bit apiece), but there may be other methods to make the selection.

A second kind of informational limit is that imposed on a decision-maker who is limited in the amount of available information. In "Chance Favors the Mind Prepared" (a letter to the editor of Science magazine), and more elaborately in "Computers and Decision Making," Ashby points out forcefully that the process of selection is limited by the information available. This is one statement of his famous Law of Requisite Variety, but the articles are included in this section rather than the next because they so clearly show how the decision-maker is constrained by an informational limit. The basic rule, says Ashby, is: Use what you know to narrow the field as much as possible; then do as you please. When the limit of information has been reached, chance is as rational as any other method for making decisions.

2: An artificial retina has a million sensitive units, each of which can only be excited or not-excited. It acts through a net that produces, as output, only a 1-bit move or not-move. Suppose we ask "what is the relation between input and output?" The question asks, essentially, for the mapping from the set of input states ($2^{1,000,000}$ in number) to the set of output states (2 in number). The number of mappings is the output number raised to the power of the input number. So the selection of a particular mapping from the $2^{(2^{1,000,000})}$ demands (unless other restrictions intervene) no less than $10^{300,000}$ bits. Again, an apparently simple question has demanded a quantity of information-processing that goes far beyond the limit.

These examples may suffice to show how easily we may ask questions, or demand computational processes, that go far beyond Bremermann's limit. Instead of being a remote and almost imaginary curiosity, it stands right in our way as soon as we attempt the more advanced forms of information processing.

The consequences of this limit are various. Here I shall mention only a few that seem to me to be outstanding in the context of bionics.

As "regulation and control" are of the highest practical importance, let us first apply the limit here. A simple example may help to get the basic ideas clear. Let us suppose that a fleet, just as it is about to leave port on active service, discovers that its communication devices for ship-to-ship coordination have failed; as a result, it has to put to sea with only some human signallers equipped with hand-operated flash lamps. Here we have a dynamic system, with a goal clearly defined (by current naval strategy), and that is subject to a limit on the amount of communication that can occur internally, determined by the capacity of the signallers. Now it is clear that the admiral may dispose of his signallers in various ways, and there may be no single maneuver of the whole fleet that can be declared impossible, yet common sense tells us that this fleet's maneuvers are going to show special features that, for instance, the enemy admiral may well notice after a time.

"Achieving coordination in maneuver" means that the total set of all possible combinations of movement (including those that lead at once to collision) are to be restricted to a special subset of the combinations (those combinations approved by naval strategy.) Achieving the restriction *demands* the corresponding quantity of transmission (by Shannon's tenth theorem or by the law of requisite variety.) Thus, to be

more definite, suppose that there are 100 ships, that the only requirement in maneuver is that all ships shall turn in the same direction, and that the signaller's total capacity as a channel provides 200 bits per course-setting. Such a fleet can coordinate its directions to the degree of choosing between port, starboard, and ahead (for $99 \log_2 3$ is less than 200) but no distribution of signallers or arrangement of coding can refine the selection of direction to adding half-to-port and half-to-starboard (for these would require $99 \log_2 5$ bits, which is greater than the 200 bits available). Thus, *the existence of a limit on the total quantity of information transmissible puts an absolute limit to the amount of regulation or control achievable.*

The arithmetic of this example shows that Bremermann's limit is no immediate threat in the case of regulations that are direct. A million ships, all having to move correctly to one part in a million would only demand $10^6 \log_2 10^6$ bits per course-setting, i.e. about 2×10^7 bits—nowhere near the limit. But this smallness does not mean that the limit can be forgotten when we change to the bionic sciences. Here the regulation and control is often directed at some *complexly patterned* event, with strong interactions between the parts (or all statements highly conditional). In such cases the quantities of information tend to increase (when the number of components is increased) at the explosive exponential rate rather than at the moderate multiplicative.

A well-known example of the effect of a complex goal is given by the mechanical chess-player. The goal ("achieve mate") looks simple, but at the present time the only sure method known for specifying what this means in individual plays is to write out all possible plays and to label each as "good" or "bad". Since the number of plays is at least 10^{120}, Bremermann's limit is an impassable barrier. Since the game of chess is simpler than the battle of life, we may expect that his limit, far from being a mere numerical curiosity, will impose itself frequently in real and practical situations.

The reason for the sudden jump from the moderate quantity of information used by the fleet to the immoderate quantity demanded by chess is, of course, due to the *combinatorial* quality of chess. Whether a piece's position is good or bad *is conditional on* where the other pieces are. The conditionality makes the variety grow combinatorially (often exponentially), where the simpler forms grow only additively or at a simple multiplicative rate. Since in bionics, and in advanced computing, we are

specially concerned with these combinatorial processes, it is particularly in our science that we are likely to encounter the limit early in our work. The topics that are specially likely to imply a major degree of interaction between the parts are especially those involving the concepts of:

System	Order
Organization	Subset
Pattern	Property
Net	Relation
Automaton	Constraint

all of them highly relevant to "advanced information processing" and "the mechanical brain".

We are thus likely to encounter the limit at an early stage in our researches, especially in bionics and artificial intelligence. But the theme has much wider implications in philosophy, at which I would like to glance.

The most obvious fact is that we, and our brains, are themselves made of matter, and are thus absolutely subject to the limit. Not only are we subject as individuals, but the whole cooperative organization of World Science is also made of matter, and is therefore subject to it. Thus both the total information that I can use personally, and the information that World Science can use, are limited, on any ordinary scale, to about 10^{80} bits. Whatever our science will become in the future, all will lie below this ceiling.

We cannot claim any special advantage because of our pre-eminent position in the world of organisms. We have been shaped, and selected to be what we are, by the process of natural selection. As a selection, this process can be measured by an information-measure: it is therefore subject to its limits. In any type of selection, under any planetary conditions, a planetary surface made of matter cannot produce adaptation faster than the rate of the limit. However good we may think we are, 10^{80} measures something that we do not exceed. The science of the future will be built by brains that cannot have had more than 10^{80} bits used in their preparations, and they themselves will advance only by something short of 10^{80}. This is our informational universe: what lies beyond is unknowable.

We can see something of what will be unknowable. Sometimes nature's laws have a simple informational structure. The law of gravity, for instance, has been found to relate the attractions between two particles, i and j say; and this relation is *unconditional* on the positions of other particles, k, l, m, ... etc. This unconditionality means that the complexities go up, as more particles are added, in a more or less additive way (the potentials do, in fact, combine simply by addition). Contrast this case with (say) a social system, in which the relation between two variables i and j may itself depend on other variables. This would be as though, in gravity, the law of attraction between i and j were altered by the position of k. Here the complexity goes up in some manner approximately exponentially. Thus the existence of the limit tells us that our achieved science will always be one of the world in its simpler interactions. If there are complex natural laws, we shall never know them.

The limit is thus likely to be specially obstructive in the sciences of the complex. One of these is sociology, just referred to as an example. The other is our own science of bionics, especially when we attack the problems of artificial intelligence. What should we do?

One reaction to the limit is simply to ignore it, noticing it only when we must. But the history of science has shown repeatedly that when an awkward limitation appears the science tends to become sterile until it has actually made the limitation a part of its working conceptual structure. The early microscopists, for instance, treated the limitations imposed by light's finite wave-length as a mere nuisance. Seeing was believing, until Abbe and Helmholtz developed the new microscopy, in which the wave features of diffraction and interference became *intrinsic* working parts of the theory. Atomic physics, too, ran into evergrowing troubles until it recast its basic ideas and constructed a new theory with the basic limitations, due to quantum restrictions and indeterminancy, built into it. Thus, there seems good reason to suggest that our best way, in the face of this limit, is to study it and to make it an integral part of our working ideas.

How is this integration to be achieved? I can here offer only a slight suggestion, in the hope that it will be found useful. Most of this work lies in the researches of the future.

First, we know that the mathematicians and engineers have derived great advantage from their development of the "linear" processes: matrix algebra, the Laplace transform, etc. With these processes they can work extensively in the linear world without the danger of breaking, at each operation, into the far more complex world of the non-linear.

This example shows that an extensive set of operations can be developed such that a great deal of worthwhile work can be done within the set, with the operations themselves *automatically* preventing the worker from wandering into the "forbidden" regions. Bremermann's limit specifies just such a region.

Now Minsky (1963) has summarized the essence of the problem of "artificial intelligence" in words with which I entirely agree: "The real problem is to find methods which significantly delay the apparently inevitable exponential growth of search trees." So far as the system studied is genuinely combinatorial, so far is the exponential growth inevitable, and Bremermann's limit acts with maximal intensity. But a large proportion of our problems in bionics are in fact subject to strong internal constraints, (most of them derived ultimately from the intense redundancy and repetitiveness shown at the atomic level.) One of the most general and wide spread constraints is that the system is to some degree reducible, i.e. capable of being studied piecemeal. When it is so, a system that seems to demand excessive information-processing may in fact allow its study to be achieved with less. (The essential reason is that if a quantity that increases exponentially, as a^n, can be treated in k stages, the branches fall to the order of $ka^{n/k}$. When n is large, the effect of k on the exponent, by dividing it, is far more powerful than its effect as a multiplier.) The method "consider the problem a piece at a time" is so widespread and so powerful that it may well be worthwhile to attempt the development of *all those operations that do not destroy reducibility*. When we know the set, the operations in it will form a calculus like those of the linear systems—such that we may do what we like within the set without fear of converting the problem from one solvable under the limit to one no longer solvable under it. A start in this direction has been made by the formulation of "cylindrance", (Ashby, 1965) which measures, for any relation between n variables, the degree to which it can be treated as if made of sub-relations, each on only some subset of the variables. It treats not only the fairly obvious case in which the relation consists of k wholly independent sub-relations but also the much more interesting case in which the whole relation has something of the simplicity of a k-fold division while being in fact still connected. (An elementary example is given by a country's telephonic communications, in that although all subscribers are joined potentially to all, the actual communications are almost all by pairs.)

The limit (of about 10^{80} bits) implies that we can never study the fully general relation between more than about 270 variables. (10^{80} bits allows us to pick one arbitrary subset from 10^{80} elements; 270 binary variables provide this number.) Since the cylindrance (a measure of intrinsic complexity) cannot exceed the number of variables, the limit implies that we can never study the fully general relation whose intrinsic complexity (if measured by cylindrance) exceeds 270.

If therefore we intend to study a system (a living brain perhaps) in which the relations do not have a cylindrance exceeding 270 we have a system that is potentially studiable. If now we unwisely ask questions or perform operations that raise the cylindrance above this number our very method of study has rendered it unstudiable. It is now known that cylindrance is safe under the operation of intersection (when the relations are treated as subsets of a product space) but that it may readily be raised by union.

This work is still in progress, but it already shows that there may exist methods, specially suited to the study of the complex system, whose operations do not lead us to the humiliating situation in which we discover that it is our own methods that have turned a potentially studiable system into one that, under the limit, is now essentially unstudiable.

SUMMARY

That nothing made of matter can transmit or process information faster than 10^{47} bits per g per sec may seem of small practical importance. In fact, many of the processes that have been proposed for machines with artifical intelligence require transmissions far in excess of this limit. Examples are given to show that large-scale processes of combinatorial richness run into the limit only too easily.

Not only are our machines so restricted, but the scientist's brain, made of matter, is also so restricted. Thus our personal knowledges, our philosophies, and our science are also limited to the same degree.

Some of its consequences in science are discussed. If our science is to be realistic, our theories must be structured so that this limit becomes an integral part of them. A suggestion is made of one way in which this incorporation might be achieved.

REFERENCES

Ashby, W. Ross (1965). Constraint analysis of many-dimensional relations. In: *Progress in biocybernetics*, edited by N. Wiener and J. P. Schade, Elsevier Publishing Co., Amsterdam; pp. 10–18.

Bremermann, H. J. (1962). Optimization through evolution and recombination. In: *Self-organizing systems* 1962, edited M. C. Yovits *et al.*, Spartan Books, Washington, D. C., pp. 93–106.

Idem (1965). Quantum noise and information. *5th Berkeley Symposium on Mathematical Statistics and Probability*; Univ. of California Press, Berkeley, California.

Minsky, M. (1963). Steps towards artificial intelligence. In: *Computers and Thought*, edited E. A. Feigenbaum and J. Feldman, McGraw-Hill Book Co., New York; pp. 406–450.

LETTERS

Chance Favors the Mind Prepared

. . . Let us assume that the problem is essentially one of selection: of a few students from many applicants, of a draft from a much larger body, or, very generally, of a good decision from a great number of possible decisions. The fundamental discovery of the last 20 years is that all such selection processes are subject to the laws of information theory. The first is that appropriate selection can be based only on information in the requisite quantity, and the second is that information is measurable and finite. It follows that in any real life situation the amount of appropriate selection that can be achieved is also finite. At any given moment, a would-be selection will have available a certain quantity of information and no more. With this quantity he can execute a corresponding quantity of rational, appropriate, meaningful selection. When the information is exhausted, no further rational grounds exist.

Selection, then, to be rational and defensible, must be based on information. But it often happens in real life that the quantity of information available falls short of the necessary. A thousand students may rationally be reduced to 500 by the information that the college accepts men only, but what are we to do if the college can accept only 50? One would not forget, of course, that more information may be available, perhaps sufficient for the whole selection to be rational; but what if the required information is either not available or could be obtained only at cost that is prohibitive? The fundamental principle of decision on a finite quantity of information may be expressed thus: *Use all that you know to shrink the range of possibilities to their minimum; after that, do as you please.*

With this rule in mind, we can see why the editorial was unsatisfactory. Its very title: 'Chance, or Human Judgment?" tended to set the reader thinking of the two competitors, mutually exclusive, while the truth is that they are natural complements. In arriving at a decision, human judgment first should prevail; then chance should be used as the necessary supplement to bring the decision to uniqueness . . . Modern methods of decision-making use both, chance and human judgment. From this point of view the use of chance is in no way a "denial of rationality." On the contrary, chance *is* the intelligent man's method of selection when he knows that the quantity of information available to him as selector is less than the quantity of selection demanded from him.

Computers and Decision Making

Sir, -- The recent correspondence on this topic has shown that there are still many misunderstandings current, and some failures to keep abreast of modern knowledge. Since today we have a clear and coherent theory of the matter, I would like to help the growth of clarity and simplicity by sketching its essentials.

The processes that interest us, and about which dispute has raged, are those (whether carried out by brain or digital computer) in which the end-product shows evidence of high selection. For instance, of all the ways in which a bookful of letters might be arranged, one set was actually produced by one Shakespeare showing evidence of very intense selection. And a computer has emitted a string of digits corresponding exactly with the first thousand digits of π. Most practical activities show this aspect of selection as an essential component (as Sommerhoff has shown extensively in his Analytical Biology, 1950). Roughly, getting right answers implies selection.

We now arrive at the simple postulate, valid for all systems, living or mechanical: Any system that achieves appropriate selection (to a degree better than chance) does so as a consequence of information received.

For what is the alternative? Are we to accept as natural the examination candidate who starts giving the appropriate answers before he has been told the questions? Or the man who sends off his claim to the insurance company before the fire has broken out? Or the computer that starts printing the appropriate answer before the programme tape has been run in? Science knows nothing of such things; until such a phenomenon is clearly demonstrated the postulate must stand.

The argument for the postulate can be given deeper and more rigorous formulation. The law of requisite variety [1] expresses the theme rigorously. It is closely related to Shannon's Tenth Theorem [2], which says that the selection by which various "noisy" versions of a message are reduced to the correct version cannot be achieved unless a correction channel (or whatever agent performs the correction) transmits a certain quantity of information to the site of correction. Human beings and computers alike are bound by the fact that if they would achieve appropriate selections they must work either subject to the postulate -- or by pure magic.

Once the postulate has been accepted, the strategy for decision-making follows inescapably. In simple and general terms it is as follows:

(1) The would-be selector, whether living or mechanical, must first receive some quantity of information. This information is then to be used to narrow the field of uncertainty (among the various possible answers or outputs) to its minimum. The amount of narrowing is bounded by the amount of information.

(2) After the information has been used up in reducing the field to its minimum, what remains is the "field of ignorance". Lacking further

information, no further selection is justifiable. No arbitrary selection within it can claim superiority over any other method ("the random is as good as any other").

In other words, the basic formula for decision-making is: Use what you know to narrow the field as far as possible; after that, do as you please.

Sometimes it happens that there is still a demand for selection even within the field of ignorance. Some selection can always be performed (for example, by using a table of random numbers as determinant), but such selection has no better than a chance possibility of being appropriate. Sometimes the selection can be carried further by the provision of more information, and sometimes this new information can be obtained by the process of "making trials"; for a "trial" is not merely a shot at Success -- it may be a process that progressively wins more information, and so makes possible a further appropriate selection.

Thus (still under the iron rule of the postulate) it may happen that the selection is achieved in stages: first the primary information is used to narrow the field of ignorance; then extra information is won by trials until the total information has reached the quantity necessary for the complete selection.

The "despair" mentioned by Mr. C. Strachey (Letters, 3 March) can now be seen in its proper proportion. Within the field of ignorance it is justified, for (by hypothesis) the primary information is exhausted and everything must be tried. But the "everything" is only "everything within the field of ignorance", and this may be only a small, perhaps an exceedingly small, fraction of the whole.

The principles given above hold over both brain and computer, and over both the simple and complex cases. The complex case often breaks up into a sequence of selections, over each of which the postulate holds. A part of the selection now often becomes a selection of an appropriate "way of breaking up"; the postulate holds with equal force over this particular case [3].

In conclusion it may be of interest to glance at the reason why these simple principles have so long eluded us. I think the reason is that we have hitherto quite mis-estimated the quantitites of information that go respectively to the computer and to the human being before they start their selective processes. In programming a computer we are acutely aware of how much labour it costs us, and we think the amount of information is very large; in fact it is small. The programmer, as a human being, however, is apt to be almost unaware of how much processing the human being has gone through -- in evolution and in childhood -- and he is apt to think the amount is small; in fact it is extremely large. After two thousand million years of evolutionary selection, followed by all the experiences of childhood and later training, he has accumulated a great store of information; with it he can perform feats of appropriate selection that far surpass those of today's computers, provided the problem is of a type for which his information is relevant. When this is so he can show to advantage.

Before he plays chess, for instance, he learns a lot about three-dimensional geometry by just moving about in the world; rows, columns, and diagonals can be indicated to him on the chessboard with a flick of the finger. The computer, however, must have this particular three-dimensional geometry specified in detail. But let the chess become five-dimensional, say, so that both are equally void of primary information, and the human being's thought processes become as laborious and detailed as the computer's. In the same way, the average human being has accumulated a great deal of information about "continuity"; so, if the problem has this peculiarity, he has a flying start over the computer. The facts thus suggest that the human being comes to his work with a far greater store of information (as "pre-programming") than the computer.

If this difference be taken fully into account, both their activities, successes and failures, will be found to be in accord with the basic postulate. And so both are bound to follow the same basic strategy for decision making.

W. Ross Ashby.
Burden Neurological Institute,
Bristol.

1. An Introduction to Cybernetics. By W. Ross Ashby. (Chapman and Hall, London, 1956.)
2. The Mathematical Theory of Communication. by C. E. Shannon and W. Weaver. (University of Illinois Press, Urbana, 1949.)
3. These more complex cases are discussed in Chapters 17 and 18 of the new edition of Design for a Brain. (Chapman and Hall, London. (1960.)

IV.

REGULATION AND CONTROL, AND THEIR RELATION TO INFORMATION

REGULATION AND CONTROL,

AND THEIR RELATION TO INFORMATION

INTRODUCTION

The Law of Requisite Variety is probably the result for which Ashby is best known. It is discussed in his Introduction to Cybernetics but is included here in "Requisite Variety and its Implications..." for completeness and because of the very interesting comments in the discussion section of the paper. The Law itself states that in the attempt to force desirable outcomes in a situation over which it has only partial control, any regulator is numerically limited by the information it has available. Additionally, the capacity of a regulator to regulate is bounded by its capacity as an information channel. There are some important qualifications (often overlooked) on the Law, but even in its simplest form it has been influential in illuminating the deep relation between information and regulation. Since regulation is a form of appropriate selection (of the "right" action out of the set of possible actions), and since Ashby argues that intelligence is basically appropriate selection, he would have been interested in recent research in which it has been found that human IQ's are correlated with the speed at which the humans can perform certain information-transfer tasks (such as quickly hitting a key when a light comes on). These results seem to verify Ashby's implied suggestion of a link between channel capacity and intelligence.

It should be pointed out that the term "channel capacity" used here and elsewhere by Ashby is not identical with the term as used by Shannon and others in the more technical version of information theory. In that context it has a somewhat elaborate definition involving limits and restrictive assumptions. Ashby uses it to mean the maximum amount of information which can be conveyed from input to output in one step, unconstrained by and without regard for what has gone before. This is a technical point which need not bother most readers.

The brief letter, "The Brain as Regulator," emphasizes a point implicit in the derivation of the Law of Requisite Variety but often overlooked there, that the best regulator is necessarily deterministic. The final paper, "Every Good Regulator of a System is a Model of the System," carries this conclusion still further by showing that under certain broad conditions, any regulator which is both maximally successful and simple must be isomorphic with the system being regulated (or homomorphic under especially favorable conditions.) The regulator can be viewed as succeeding by using an internal model of the system to "predict" what the system will do, then deterministically selecting an appropriately matched "countermove." If the regulator successfully develops its strategy by trial and error, the process will evolve an equivalent of a model of the system. Ashby was quite keen on this paper; he viewed the theorem in it as the basis of a "theoretical neurology," since it seems to show how the brain must act in its exercise of appropriate selection.

REQUISITE VARIETY AND ITS IMPLICATIONS FOR THE CONTROL OF COMPLEX SYSTEMS

By W. ROSS ASHBY,
Director of Research, Barnwood House (Gloucester)

Recent work on the fundamental processes of regulation in biology (Ashby, 1956) has shown the importance of a certain quantitative relation called the law of requisite variety*. After this relation has been found, we appreciated that it was related to a theorem in a world far removed from the biological – that of Shannon on the quantity of noise or error that could be removed through a correction-channel (Shannon and Weaver, 1959; theorem 10). In this paper I propose to show the relationship between the two theorems, and to indicate something of their implications for regulation, in the cybernetic sense, when the system to be regulated is extremely complex.

Since the law of requisite variety uses concepts more primitive than those used by entropy, I will start by giving an account of that law.

I

Variety.

Given a set of elements, its ***variety*** is the number of elements that can be distinguished. Thus the set

$$\{g\ b\ c\ g\ g\ c\}$$

has a variety of three letters. (If two observers differ in the distinctions they can make, then they will differ in their estimates of the variety. Thus if the set it

$$\{b\ c\ a\ a\ C\ a\ B\ a\}$$

its variety in shapes is five, but its variety in letters is three. We shall not, however, have to treat this complication.)

For many purposes the variety may more conveniently be measured by the logarithm of this number. If the logarithm is taken to base 2, the unit is the *bit*. The context will make clear whether the number or its logarithm is being used as measure.

Regulation and the pay-off matrix.

Regulation achieves a "goal" against a set of disturbances. The

* Traduction francaise: loi de la variete indispensable.

disturbances may be actively hostile, as are those coming from an enemy, or merely irregular, as are those coming from the weather. The relations may be shown in the most general way by the formalism that is already well known in the theory of games (Neumann and Morgenstern, 1947).

A set D of disturbances d_i can be met by a set R of responses r_j. The outcomes provide a table or matrix

		R			
		r_1	r_2	r_3	...
D	d_1	z_{11}	z_{12}	z_{13}	...
	d_2	z_{21}	z_{22}	z_{23}	...
	d_3	z_{31}	z_{32}	z_{33}	...
	d_4	z_{41}	z_{42}	z_{43}	...
	...	...	...	...	...

in which each cell shows an element z_{ij} from the set Z of possible outcomes.

It is not implied that the elements must be numbers (though the possibility is not excluded). The form is thus general enough to include the case in which the events d_i and r_j are themselves vectors, and have a complex internal structure. Thus the disturbances D might be all the attacks that can be made by a hostile army, and the responses R all the counter-measures that might be taken. What is required at this stage is that the sets are sufficiently well defined so that the facts determine a single-valued mapping of the product set $D \times R$ into the set Z of possible outcomes. (I use here the concepts as defined by Bourbaki, 1951).

The "outcomes" so far are simple events, without any implication of desirability. In any real regulation, for the benefit of some defined person or organism or organization, the facts usually determine a further mapping of the set Z of outcomes into a set E of values. E may be as simple as the 2-element set (good, bad), and is commonly an ordered set, representing the preferences of the organism. Some subset of E is then defined as the "goal". The set of values, with perhaps a scale of preference, is often obvious in human affairs; but in the biological world, and in the logic of the subject, it must have explicit mention. Thus if the outcome is "get into deep water", the valuation is uncertain until we know whether the organism is a cat or a fish.

In the living organisms, the scale of values is usually related to their

"essential variables" – those fundamental variables that must be kept within certain "physiological" limits if the organism is to survive. Other organizations also often have their essential variables: in an economic system, a firm's profits is of this nature, for only if this variable keeps positive can the firm survive.

Given the goal – the "good" or "acceptable" elements in E – the inverse mapping of this subset will define, over Z, the subset of "acceptable outcomes". Their occurrence in the body of the table or matrix will thus mark a subset of the product set $D \times R$. Thus is defined a binary relation S between D and R in which "the elements d_i and r_j have the relation S" is equivalent to "r_j as response to d_i gives an acceptable outcome".

Control.

In this formulation we have considered the case in which the regulator acts so as to limit the outcome to a particular subset, or to keep some variables within certain limits, or even to hold some variables constant. This reduction to constancy must be understood to include all those cases, much more numerous, that can be reduced to this form. Thus if a gun is to follow a moving target, the regulation implied by accuracy of aim may be represented by a keeping at zero of the difference between the gun's aim and the target's position. The same remark is clearly applicable to all cases where an unchanging (constant) relation is to be maintained between one variable that is independent and another variable that is controlled by the regulator.

Thus, as a special instance, if a variable y (which may be a vector) is to be controlled by a variable a, and if distrubance D has access to the system so that y is a function of both the control a and the value of disturbance D, then a suitable regulator that has access to the disturbance may be able to counter its effects, remove its effect from y, and thus leave y wholly under the control of a. In this case, successful regulation by R is the necessary and sufficient condition for successful control by a.

Requisite variety.

Consider now the case which, given the table of outcomes (the pay-off matrix), the regulator R has the best opportunities for success. (The other cases occur as degenerate forms of this case, and need not be considered now in detail.)

Given the table, R's opportunity is best if R can respond knowing D's value. Thus, suppose that D must first declare his (or its) selection d_i; a particular row in the table is thereby selected. When this has been done, and knowing D's selection, R selects a value r_j, and thus selects a particular column. The

outcome is the value of Z at the intersection. Such a table might be:

		R		
		r_1	r_2	r_3
D	d_1	c	a	d
	d_2	b	d	a
	d_3	c	d	c
	d_4	a	a	b
	d_5	d	b	b

If outcomes a, b count as Good, and c, d as Bad, then if D selects d_1, R must select r_2; for only thus can R score Good. If D selects d_2, R may choose r_1 or r_3. If D selects d_3, then R cannot avoid a Bad outcome; and so on.

Nature, and other sources of such tables, provides them in many forms, ranging from the extreme at which every one of R's responses results in Good (these are distinctly rare!), to those hopeless situations in which every one of R's responses leads to Bad. Let us set aside these less interesting cases, and consider the case, of central importance, in which each column has all its elements different. (Nothing is assumed here about the relation between the contents of one column and those of another.) What this implies is that if the set D had a certain variety, the outcomes in any one column will have the same variety. In this case, if R is inactive in responding to D (i.e., if R adheres to one value r, for all values of D), then the variety in the outcomes will be as large as that in D. Thus in this case, and if R stays constant, D can be said to be exerting full control over the outcomes.

R, however, aims at confining the actual outcomes to some subset of the possible outcomes Z. It is necessary, therefore, that R acts so as to lessen the variety in the outcomes. If R does so act, then there is a quantitative relation between the variety in D, the variety in R, and the smallest variety that can be achieved in the set of actual outcomes; namely, *the latter cannot be less than the quotient of the number of rows divided by the number of columns* (Ashby, 1956; S.11/5).

If the varieties are measured logarithmically, this means that if the varieties of D, R, and actual outcomes are respectively V_d, V_r, and V_o then the minimal value of V_o is $V_d - V_r$. If now V_d is given, V_o's minimum can be lessened *only by a corresponding increase in* V_r. This is the law of requisite variety. What it means is that restriction of the outcomes to the subset that is valued as Good demands a certain variety in R.

We can see the relation from another point of view. R, by depending on D for its value, can be regarded as a channel of communication between D and the outcomes (though R, by acting as a regulator, is using its variety subtractively from that of D). The law of requisite variety says that R's *capacity as a regulator cannot exceed its capacity as a channel for variety.*

The functional dependencies can be represented as in Fig. 1. (This diagram is necessary for comparison with Figs. 2 and 3.)

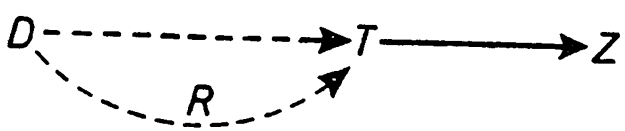


FIG. 1

The value at D threatens to transmit, via the table T to the outcomes Z, the full variety that occurs at D. For regulation, another channel goes through R, which takes a value so paired to that of D that T gives values at Z with reduced variety.

Nature of the limitation.

The statement that some limit cannot be exceeded may seem rash, for Nature is full of surprises. What, then, would we say if a case were demonstrated in which objective measurements shows that the limit was being exceeded? Here we would be facing the case in which appropriate effects were occurring without the occurrence of the corresponding causes. We would face the case of the examination candidate who gives the appropriate answers before he has been given the corresponding questions! When such things have happened in the past we have always looked for, and found, a channel of communication which has accounted for the phenomenon, and which has shown that the normal laws of cause and effect do apply. We may leave the future to deal similarly with such cases if they arise. Meanwhile, few doubt that we may proceed on the assumption that genuine overstepping of the limitation does not occur.

Examples in biology.

In the biological world, examples that approximate to this form are

correspond with mathematical precision. This inexactness of correspondence does not matter in our present context, for we shall not be concerned with questions involving high accuracy, but only with the existence of this particular limitation.

An approximate example occurs when an organism is subject to attacks by bacteria (of species d_j) so that, if the organism is to survive, it must produce the appropriate anti-toxin r_j. If the bacterial species are all different, and if each species demands a different anti-toxin, then clearly the organism, for survival, must have at least as many anti-toxins in its repertoire of responses as there are bacterial species.

Again, if a fencer faces an opponent who has various modes of attack available, the fencer must be provided with at least an equal number of modes of defence if the outcome is to have the single value: attack parried.

Analysis of Sommerhoff.

Sommerhoff (1950) has conducted an analysis in these matters that bears closely on the present topic. He did not develop the quantitative relation between the varieties, but he described the basic phenomenon of regulation in biological systems with a penetrating insight and with a wealth of examples.

He recognizes that the concept of "regulation" demands variety in the disturbances *D*. His "coenetic variable" is whatever is responsible for the values of *D*. He also considers the environmental conditions that the organism must take into account (but as, in his words, these are "epistemically dependent" on the values of the coenetic variable, our symbol *D* can represent both, since his two do not vary independently). His work shows, irrefutably in my opinion, how the concepts of coordination, integration, and regulation are properly represented in abstract form by a relation between the coenetic variable and the response, such that the outcome of the two is the achievement of some "focal condition" (referred to as "goal" here). From our point of view, what is important is the recognition that without the regulatory response the values at the focal condition would be more widely scattered.

Sommerhoff's diagram (Fig. 2) is clearly similar. (I have modified it slightly, so as to make it uniform with Figs. 1 and 3).

His analysis is valuable as he takes a great many biological examples and shows how, in each case, his abstract formulation exactly catches what is essential while omitting the irrelevant and merely special. Unfortunately, in stating the thesis, he did what I did in 1952 – used the mathematical language of analysis and continuous functions. This language now seems unnecessarily clumsy and artificial; for it has been found (Ashby, 1956) that the concepts

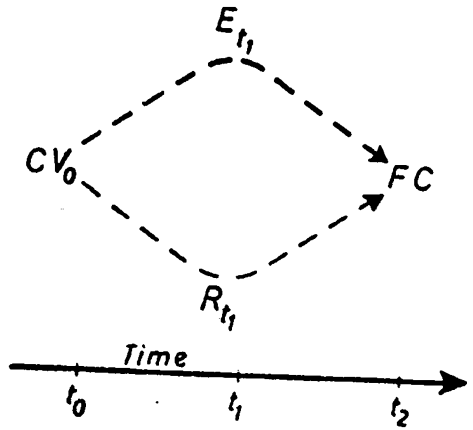


FIG 2

of set theory, especially as expounded by Bourbaki (1952) are incomparably clearer and simpler, while losing nothing in rigor. By the change to set theory, nothing in fact is lost, for nothing prevents the elements in a set from being numbers, or the functions from being continuous, and the gain in generality is tremendous. The gain is specially marked with biological material, in which non-numerical states and discontinuous functions are ubiquitous.

Let me summarise what has been said about "regulation". The concept of regulation is applicable when there is a set *D* of disturbances, to which the organism has a set *R* of responses, of which on any occasion it produces some one, r_j say. The physico-chemical or other nature of the whole system then determines the outcome. This will have some value for the organism, either Good or Bad say. If the organism is well adapted, or has the know-how, its response r_j, as a variable, will be such a function of the disturbance d_i that the outcome will always lie in the subset marked as Good. The law of requisite variety then says that such regulation cannot be achieved unless the regulator *R*, as a channel of communication, has more than a certain capacity. Thus, if *D* threatens to introduce a variety of 10 bits into the outcomes, and if survival demands that the outcomes be restricted to 2 bits, then at each action *R* must provide variety of at least 8 bits.

Ergodicity.

Before these ideas can be related to those of the communication theory of Shannon, we must notice that the concepts used so far have not assumed ergodicity, and have not even used the concept of probability.

The fact that communication theory, during the past decade, has tended to specialize in the ergodic case is not surprising when we consider that its application has been chiefly to telephonic and other communications in which the processes go on incessantly and are usually stationary statistically. This fact should not, however, blind us to the fact that many important communications are non-ergodic, their occurrence being specially frequent in the biological world. Thus we frequently study a complex biological system by isolating it, giving it a stimulus, and then observing the complex trajectory that results. Thus the entomologist takes an ant colony, places a piece of meat nearby, and then observes what happens over the next 24 hours, without disturbing it further. Or the social psychologist observes how a gang of juvenile criminals forms, becomes active, and then breaks up. In such cases, even a single trajectory can provide abundant information by the comparison of part with part, but the only ergodic portion of the trajectory is that which occurs ultimately, when the whole has arrived at some equilibrium, in which nothing further of interest is happening. Thus the ergodic part is degenerate. It is to be hoped that the extension of the basic concepts of Shannon and Wiener to the non-ergodic case will be as fruitful in biology as the ergodic case has been in commercial communication. It seems likely that the more primitive concept of "variety" will have to be used, instead of probability; for in the biological cases, systems are seldom isolated long enough, or completely enough, for the relative frequencies to have a stationary limit.

Among the ergodic cases there is one, however, that is obviously related to the law of requisite variety. It is as follows.

Let D, R, and E be three variables, such that we may properly observe or calculate certain entropies over them. Our first assumption is that if R is constant, all the entropy at D will be transmitted to, and appear at, E. This is equivalent to

$$H_r(E) = H_r(D) \qquad (1)$$

By writing $H(D,R)$ in two forms we have

$$H(D) + H_d(R) = H(R) + H_r(D)$$

Use of (1) gives

$$\begin{aligned} H(D) + H_d(R) &= H(R) + H_r(E) \\ &= H(R,E) \\ &\leqslant H(R) + H(E) \end{aligned}$$

i.e. $$H(E) \geqslant H(D) + H_d(R) - H(R) \qquad (2)$$

The entrophy of E thus has a certain minimum – the expression on the right of (2). If $H(E)$ is the entropy of the actual outcomes, then, for regulation, it may have to be reduced to a certain value. Equation (2) shows what can reduce it; it can be reduced:

(i) by making $H_d(R) = 0$, i.e. by making R a determinate function of D,

(ii) by making $H(R)$ larger.

If $H_d(R) = 0$, and $H(R)$ the only variable on the right of (2), then a decrease in $H(E)$ demands at least an equal increase in $H(R)$. This conclusion is clearly similar to that of the law of requisite variety.

A simple generalization has been given (Ashby, 1956) in which, when R remains constant, only a certain fraction of D's variety or entropy shows in the outcomes or in $H(E)$. The result is still that each decrease in $H(E)$ demands at least an equal increase in $H(R)$.

With this purely algebraic result we can now see exactly how these ideas join on to Shannon's. His theorem 10 uses a diagram which can be modified to Figure 3 (to match the two preceding Figures).

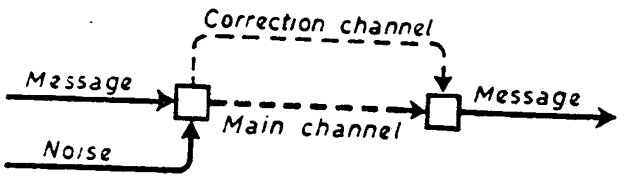


FIG. 3

Our "disturbance D", which threatens to get through to the outcome, clearly corresponds to the noise; and his theorem says that the amount of noise that can be prevented from appearing in the outcomes is limited to the entropy that can be transmitted through the correction channel.

The message of zero entropy.

What of the "message"? In regulation, the "message" to be transmitted is a constant, i.e. has zero entropy. Since this matter is fundamental, let us consider some examples. The ordinary thermostat is set at, say, 70°F. "Noise", in the form of various disturbances, providing heat or cold, threatens to drive the output from this value. If the thermostat is completely efficient, this variation will be completely removed, and an observer who watches the temperature will see continuously only the value that the initial controller has set. The "message" is here the constant value 70.

Similarly, the homeostatic mechanism that keeps our bodies, in health, about 98°F is set at birth to maintain this value. The control comes from the gene-pattern and has zero entropy, for the selected value is unchanging.

The same argument applies similarly to all the regulations that occur in other systems, such as the sociological and economic. Thus an attempt to stabilize the selling price of wheat is an attempt to transmit, to the farmers, a "message" of zero entropy; for this is what the farmer would receive if he were to ask daily "what is the price of wheat today"? The stabilization, so far as it is successful, frees the message from the effects of those factors that might drive the price from the selected value.

Thus, all acts of regulation can be related to the concepts of communication theory by our noticing that the "goal" is a message of zero entropy; and that the "disturbances" correspond to noise.

The error-controlled regulator.

A case in which this limitation acts with peculiar force is the very common one in which the regulator is "error-controlled". In this case the regulator's channel for information about the disturbances has to pass through a variable (the "error") which is kept as constant as possible (at zero) by the regulator *R* itself. Because of this route for the information, the more successful the regulator, the less will be the range of the error, and therefore the less will be the capacity of the channel from *D* to *R*. To go to the extreme: if the regulator is totally successful, the error will be zero unvaryingly, and the regulator will thus be cut off totally from the information (about *D*'s value) that alone can make it successful – which is absurd. The error-controlled regulator is thus fundamentally incapable of being 100 percent efficient.

Living organisms encountered this fact long ago, and natural selection and evolution have since forced the development of channels of information, through eyes and ears for instance, that supply them with information about *D* before the chain of cause and effect goes so far as to cause actual error. At the present time, control by error is widely used in industry, in servomechanisms and elsewhere, as a means to regulation. Some of these regulations by error-control are quite difficult to achieve. Immersed in the intricacies of Nyquist's theorem, transfer functions, and other technical details, the design engineer may sometimes forget that there is another way to regulation. May I suggest that he would do well to bear in mind what has been found so advantageous in the biological world, and to consider whether a regulation which is excessively difficult to design when it is controlled by error may not be easier to design if it is controlled not by the error but by what gives rise to

the error.

This is a first application to cybernetics of the law of requisite variety and Shannon's theorem 10.

It is not my purpose in this paper, however, to explore in detail how the limitation affects simple regulators. Rather I want to consider its effect in matters that have so far, I think, received insufficient attention. I want to indicate, at least in outline, how this limitation also implies a fundamental limitation on the human intellect, especially as that intellect is used in scientific work. And I want to indicate, in the briefest way, how we scientists will sometimes have to readjust our ways because of it.

II

The limitations of the scientist.

In saying that the human intellect is limited, I am not referring to those of its activities for which there is no generally agreed valuation – I am not referring for instance, to the production of pictures that please some and displease others – for without an agreed valuation the concept of regulation does not exist. I refer rather to those activities in which the valuation is generally agreed on, and in which the person shows his capacity by whether he succeeds or fails in getting an acceptable outcome. Such is the surgeon, whose patient lives or dies; such is the mathematician, given a problem, which he does or does not solve; such is the manager whose business prospers or fails; such is the economist who can or cannot control an inflationary spiral.

Not only are these practical activities covered by the theorem and so subject to limitation, but also subject to it are those activities by which Man shows his "intelligence". "Intelligence" today is defined by the method used for its measurement; if the tests used are examined they will be found to be all of the type: from a set of possibilities, indicate one of the appropriate few. Thus all measure intelligence by the *power of appropriate selection* (of the right answers from the wrong). The tests thus use the same operation as is used in the theorem on requisite variety, and must therefore be subject to the same limitation. (*D*, of course, is here the set of possible questions, and *R* is the set of possible answers.) Thus what we understand as a man's "intelligence" is subject to the fundamental limitation: it cannot exceed his capacity as a transducer. (To be exact, "capacity" must here be defined on a per-second or a per-question basis, according to the type of test.)

The team as regulator.

It should be noticed that the limitation on "the capacity of Man" is grossly ambiguous, according to whether we refer to a single person, to a team, or to the whole of organized society. Obviously, that one man has a limited capacity does not impose a limitation on a team of *n* men, if *n* may be increased without limit. Thus the limitation that holds over a team of *n* men may be much higher, possibly *n* times as high, as that holding over the individual man.

To make use of the higher limitation, however, the team must be efficiently organized; and until recently our understanding of organization has been pitifully small. Consider, for instance, the repeated attempts that used to be made (especially in the last century) in which some large Chess Club played the World Champion. Usually the Club had no better way of using its combined intellectual resources than either to take a simple majority vote on what move to make next (which gave a game both planless and mediocre), or to follow the recommendation of the Club's best player (which left all members but one practically useless). Both these methods are grossly inefficient. Today we know a good deal more about organization, and the higher degrees of efficiency should soon become readily accessible. But I do not want to consider this question now. I want to emphasize the limitation. Let us therefore consider the would-be regulator, of some capacity that cannot be increased, facing a system of great complexity. Such is the psychologist, facing a mentally sick person who is a complexly interacting mass of hopes, fears, memories, loves, hates, endocrines, and so on. Such is the sociologist, facing a society of mixed races, religions, trades, traditions, and so on. I want to ask: given his limitation, and the complexity of the system to be regulated, what scientific strategies should he use?

In such a case, the scientist should beware of accepting the classical methods without scrutiny. The classical methods have come to us chiefly from physics and chemistry, and these branches of science, far from being all-embracing, are actually much specialized and by no means typical. They have two peculiarities. The first is that their systems are composed of parts that show an extreme degree of homogeneity: contrast the similarity between atoms of carbon with the dissimilarity between persons. The second is that the systems studied by the physicist and chemist have nothing like the richness of internal interaction that have the systems studied by the sociologist and psychologist.

Or take the case of the scientist who would study the brain. Here again is a system of high complexity, with much heterogeneity in the parts, and great richness of connection and internal interaction. Here too the quantities of information involved may well go beyond the capacity of the scientist as a transducer.

Both of these qualities of the complex system – heterogeneity in the parts, and richness of interaction between them – have the same implication: the quantities of information that flow, either from system to observer or from part to part, are much larger than those that flow when the scientist is physicist or chemist. And it is because the quantities are large that the limitation is likely to become dominant in the selection of the appropriate scientific strategy.

As I have said, we must beware of taking our strategies slavishly from physics and chemistry. They gained their triumphs chiefly against systems whose parts are homogeneous and interacting only slightly. Because their systems were so specialized, they have developed specialized strategies. We who face the complex system must beware of accepting their strategies as universally valid. It is instructive to notice that their strategies have already broken down in one case, which is worth a moment's attention. Until about 1925, the rule "vary only one factor at a time" was regarded as the very touchstone of the scientific method. Then R.A. Fisher, experimenting with the yields of crops from argicultural soils, realized that the system he faced was so dynamic, so alive, that any alteration of one variable would lead to changes in an uncountable number of other variables long before the crop was harvested and the experiment finished. So he proposed formally to vary whole sets of variables simultaneously – not without peril to his scientific reputation. At first his method was ridiculed, but he insisted that his method was the truly scientific and appropriate one. Today we realize that the rule "vary only one factor at a time" is appropriate only to certain special types of systems, not valid universally. Thus we have already taken one step in breaking away from the classical methods.

Another strategy that deserves scrutiny is that of collecting facts "in case they should come in useful sometime" – the collecting of truth "for truth's sake". This method may be efficient in the systems of physics and chemistry, in which the truth is often invariant with time; but it may be quite inappropriate in the systems of sociology and economics, whose surrounding conditions are usually undergoing secular changes, so that the parameters to the system are undergoing changes – which is equivalent to saying that the systems are undergoing secular changes. Thus, it may be worth while finding the density of pure hafnium, for if the value is wanted years later it will not be changed. But of what use today, to a sociologist studying juvenile delinquency, would a survey be that was conducted, however carefully, a century ago? It *might* be relevant and helpful; but we could know whether it was relevant or not only *after* a comparison of it with the facts of today; and when we know these, there would be no need for the old knowledge. Thus the rule "collect truth

for truth's sake" may be justified when the truth is unchanging; but when the system is not completely isolated from its surroundings, and is undergoing secular changes, the collection of truth is futile, for it will not keep.

There is little doubt, then, that when the system is complex, the scientist should beware of taking, without question, the time-honored strategies that have come to him from physics and chemistry, for the systems commonly treated there are specialized, not typical of those that face him when they are complex.

Another common aim that will have to be given up is that of attempting to "understand" the complex system; for if "understanding" a system means having available a model that is isomorphic with it, perhaps in one's head, then when the complexity of the system exceeds the finite capacity of the scientist, the scientist can no longer understand the system – not in the sense in which he understands, say, the plumbing of his house, or some of the simple models that used to be described in elementary economics.

Operational research.

It will now be obvious that the strategies appropriate to the complex system are those already getting well known under the title of "operational research". Scientists, guided doubtless by an intuitive sense of what is reasonable, are already breaking away from the classical methods, and are developing methods specially suitable for the complex system. Let me review briefly the chief characteristics of "operational research".

Its first characteristic is that its ultimate aim is not understanding but the purely practical one of control. If a system is too complex to be understood, it may nevertheless still be controllable. For to achieve this, all that the controller wants to find is some action that gives an acceptable result; he is concerned only with what happens, not with why it happens. Often, no matter how complex the system, what the controller wants is comparatively simple: has the patient recovered? – have the profits gone up or down? – has the number of strikes gone up or down?

A second characteristic of operational research is that it does not collect more information than is necessary for the job. It does not attempt to trace the whole chain of causes and effects in all its richness, but attempts only to relate controllable causes with ultimate effects.

A third characteristic is that it does not assume the system to be absolutely unchanging. The research solves the problems of today, and does not assume that its solutions are valid for all time. It accepts frankly that its solutions are valid merely until such times as they become obsolete.

The philosopher of science is apt to look somewhat askance at such methods, but the practical scientist knows that they often achieve success when the classical methods bog down in complexities. How to make edible bread, for instance, was not found by the methods of classical science – had we waited for that we still would not have an edible loaf – but by methods analogous to those of operational research: if a variation works, exploit it further; ask not *why* it works, only *if* it works. We must be careful, in fact, not to exaggerate the part played by classical science in present-day civilization and technology. Consider, for instance, how much empirical and purely practical knowledge plays a part in our knowledge of metallurgy, of lubricants, of house-building, of pottery, and so on.

What I suggest is that measurement of the quantity of information, even if it can be done only approximately, will tell the investigator where a complex system falls in relation to his limitation. If it is well below the limit, the classic methods may be appropriate; but should it be above the limit, then if his work is to be realistic and successful, he must alter his strategy to one more like that of operational research.

My emphasis on the investigator's limitation may seem merely depressing. That is not at all my intention. The law of requisite variety, and Shannon's theorem 10, in setting a limit to what can be done, may mark this era as the law of conservation of energy marked its era a century ago. When the law of conservation of energy was first pronounced, it seemed at first to be merely negative, merely an obstruction; it seemed to say only that certain things, such as getting perpetual motion, could not be done. Nevertheless, the recognition of that limitation was of the greatest value to engineers and physicists, and it has not yet exhausted its usefulness. I suggest that recognition of the limitation implied by the laws of requisite variety may, in time, also prove useful, by ensuring that our scientific strategies for the complex system shall be, not slavish and inappropriate copies of the strategies used in physics and chemistry, but new strategies, genuinely adapted to the special peculiarities of the complex system.

REFERENCES

ASHBY, W. ROSS, ***Design for a brain.*** 2nd. imp. Chapman & Hall, London, 1954.

ASHBY, W. ROSS, ***An introduction to cybernetics.*** Chapman & Hall, London, 1956.

BOURBAKI, N., ***Theorie des ensembles. Fascicule de resultats.*** A.S.E.I. No. 1141. Hermann et Cie, Paris, 1951.

NEUMANN, J. (von) and MORGENSTERN, O., ***Theory of games and economic behavior.*** Princeton, 1947.

SHANNON, C.E. and WEAVER, W., ***The mathematical theory of communication.*** University of Illinois Press, Urbana, 1949.

SOMMERHOFF, G., ***Analytical biology.*** Oxford, University Press, London, 1950.

THE BRAIN AS REGULATOR

There has been some debate whether the brain is determinate or probabilistic in its behavior. As a contribution to this question, the following argument shows that from one point of view the determinate system is demonstrably of greater efficiency than the probabilistic. The physiologist may therefore expect to find that natural selection has made the brain as determinate as possible.

If the brain is regarded as basically a means to survival, then a necessary condition for survival is that against the (moderately well-defined) set of disturbances that threaten the organism's existence the brain must so respond that the outcome of the combined action of disturbance and response keeps the organism's essential variables within normal limits[1,2,3]. The brain must, in other words, act as a regulator, homeostatic in the general sense. The sequence of disturbances that comes to the organism can then often be treated (algebraically at least) as an information source (in Shannon's sense[4]) having an entropy, $H(D)$ say. Similarly, the sequence of responses will have an entropy $H(R)$ and so will the sequence of values at the essential variables, $H(E)$. It is necessary for survival (though no sufficient) that $H(E)$ be kept small.

The most interesting case is that in which, if the organism does nothing, *all* the changes at D produce changes at E (that is, the organism is passively buffeted to full degree). In this case $H_r(E) = H_r(D)$, and it then follows[2] that $H(E)$'s minimum is given by $H(D) + H_d(R) - H(R)$. So far as $H_d(R)$ is concerned, the expression will be least if $H_d(R) = 0$. This relation gives a necessary condition that the brain must satisfy if it is to have maximal efficiency in generalized homeostasis.

The meaning of $H_d(R) = 0$ is readily specified. Of any source, A, $H(A) = 0$ implies that the Markov chain must be such that when it reaches equilibrium the transitions that still occur are determinate (that is, have probabilities all 0 or 1). $H_d(R) = 0$ implies that this tendency to determinatedness must hold for each value of D.

Thus, if the brain is survival-promoting by acts of regulation, it has maximal efficiency (other things being equal) if, under constant external conditions, it tends towards the deterministic way of behaving.

REFERENCES

1 ASHBY, W. ROSS, ***Design for a Brain,*** 2nd ed., 64 (Chapman and Hall, London, 1960).

2 ASHBY, W. Ross, ***Introduction to Cybernetics",*** 197, 208 (Chapman and Hall, London, 1956).

3 SOMMERHOFF, G., ***Analytical Biology*** (Oxford University Press, London, 1950).

4 SHANNON, C.E., and WEAVER, W., ***The Mathematical Theory of Communication*** (University of Illinois Press, Urbana, 1949).

INT. J. SYSTEMS SCI., 1970, VOL. 1, NO. 2, 89-97

EVERY GOOD REGULATOR OF A SYSTEM MUST BE A MODEL OF THAT SYSTEM*

ROGER C. CONANT
Department of Information Engineering, University of Illinois
and
W. ROSS ASHBY
Biological Computers Laboratory, University of Illinois**

[Received 3 June 1970]

The design of a complex regulator often includes the making of a model of the system to be regulated. The making of such a model has hitherto been regarded as optional, as merely one of many possible ways.

In this paper a theorem is presented which shows, under very broad conditions, that any regulator that is maximally both successful and simple *must* be isomorphic with the system being regulated. (The exact assumptions are given.) Making a model is thus necessary.

The theorem has the interesting corollary that the living brain, so far as it is to be successful and efficient as a regulator for survival, *must* proceed, in learning, by the formation of a model (or models) of its environment.

1. Introduction

Today, as a step towards the control of complex dynamic systems, models are being used ubiquitously. Being modelled, for instance, are the air traffic flows around New York, the endocrine balances of the pregnant sheep, and the flows of money among the banking centers.

So far, these models have been made mostly with the idea that the model might help, but the possibility remained that the cybernetician (or the Sponsor) might think that some other way was better, and that making a model (whether digital, analogue, mathematical, or other) was a waste of time. Recent work (Conant 1969), however, has suggested that the relation between regulation and modelling might be much closer, that modelling might in fact be a *necessary* part of regulation. In this article we address ourselves to this question.

---

* Communicated by Dr. W. Ross Ashby. This work was in part supported by the Air Force Office of Scientific Research under Grant AF-OSR 70-1865.

** Now at University College, Cardiff, Wales.

The answer is likely to be of interest in several ways. First, there is the would-be designer of a regulator (of traffic round an airport say) who is building, as a first stage, a model of the flows and other events around the airport. If making a model is *necessary,* he may proceed relieved of the nagging fear that at any moment his work will be judged useless. Similarly, before any design is started, the question: How shall we start? may be answered by: A model *will* be needed; let's build one.

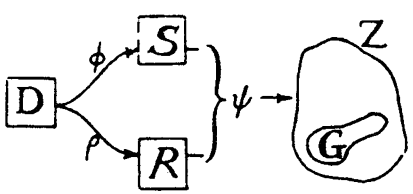
Quite another way in which the answer would be of interest is in the brain and its relation to behavior. The suggestion has been made many times that *perhaps* the brain operates by building a model (or models) of its environment; but the suggestion has (so far as we know) been offered only as a possibility. A proof that model-making is necessary would give neurophysiology a theoretical basis, and would predict modes of brain operation that the experimenter could seek. The proof would tell us what the brain, as a complex regulator for its owner's survival, *must* do. We could have the basis for a theoretical neurology.

The title will already have told this paper's conclusion, but to it some qualifications are essential. To make these clear, and to avoid vaguenesses and ambiguities (only too ready to occur in a paper with our range of subject) we propose to consider exactly what is required for the proof, and just how the general ideas of regulation, model, and system are to be made both rigorous and objective.

2. Regulation

Several approaches are possible. Perhaps the most general is that given by Sommerhoff (1950) who specifies five variables (each a vector or n-tuple perhaps) that must be identified by the part they play in the whole process.

Fig. 1



(1) There is the total set Z of events that may occur, the regulated and the unregulated; e.g. all the possible events at an airport, good and bad. (Set Z in Ashby's (1967) reformulation in terms of set theory.)

(2) The set G, a sub-set of Z, consisting of the "good" events, those ensured by effective regulation.

(3) The set R of events in the regulator R; (e.g. in the control tower). [We have found clarity helped by distinguishing the regulator as an object from the set of events, the values of the variables that compose the regulator. Here we use italic and Roman capitals respectively.]

(4) The set S of events in the rest of the system S (e.g. positions of aircraft, amounts of fuel left in their tanks) [with italic and Roman capitals similarly].

(5) The set D of primary disturbers (Sommerhoff's "coenetic" variable); those that, by causing the events in the system S, tend to drive the outcomes out of G; (e.g. snow, varying demands, mechanical emergencies).

(Figure 1 may help to clarify the relations, but the arrows are to be understood for the moment as merely suggestive.) A typical act of regulation would be given by a hunter firing at a pheasant that flies past. D would consist of all those factors that introduce disturbance by the bird's coming sometimes at one angle, sometimes another; by the hunter being, at the moment, in various postures; by the local wind blowing in various directions; by the lighting being from various directions. S consists of all those variables concerned in the dynamics of bird and gun other than those in the hunter's brain. R would be those variables in his brain. G would be the set of events in which shot does hit bird. R is now a "good" regulator (is achieving "regulation") if and only if, for all values of D, R is so related to S that their *interaction* gives an event in G.

This formulation has withstood 20 years' scrutiny and undoubtedly covers the great majority of cases of accepted regulation. That it is also rigorous may be shown (Ashby 1967) by the fact that if we represent the three mappings by which each value (Fig. 1) evokes the next:

$$\phi: D \to S$$

$$\rho: D \to R$$

$$\psi: S \times R \to Z$$

then "R is a good regulator (for goal G, given D, etc., ϕ and ψ)" is equivalent to

$$\rho \subset [\psi^{-1}(G)] : \phi. \qquad (1)$$

to which we must add the obvious condiction that

$$\rho\rho^{-1} \subset 1 \subset \rho^{-1}\rho,$$

to ensure that ρ is an actual mapping, and not, say, the empty set! (We represent composition by adjacency, by a dot, or by parenthesis according to which best gives the meaning.)

It should be noticed that in this formulation there is no restriction to linearity, to continuity, or even to the existence of a metric for the sets, though these are in no way excluded. The variables, too, may be partly functions of earlier real time; so the formulation is equally valid for regulations that

involve "memory", provided the sets D, etc., are defined suitably.

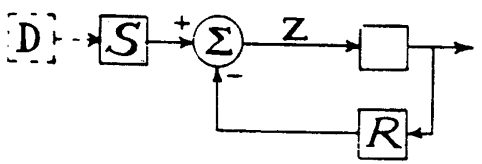
Any concept of "regulation" must include such entities as the regulator *R*, the regulated system *S*, and the set of possible outcomes *Z*. Sometimes, however, the criterion of success is not whether the outcome, after each interaction of S and R, is within a goal-set G, but is whether the outcomes, on some numerical scale, have a root-mean-square sufficiently small.

A third criterion for success is to consider whether the entropy *H(Z)* is sufficiently small. When Z can be measured on an additive scale they tend to be similar: complete constancy of outcome ⇔ *H(Z)* = 0 ⇔ r.m.s = 0, (though the mathematician can devise examples to show that they are essentially independent). But the entropy measure of scatter has the advantage that it can be applied when the outcome can only be classified, not measured (e.g. species of fish caught in trawling, amino-acid chain produced by a ribosome). In this paper we shall use the last measure, *H(Z)*, and we define "successful regulation" as equivalent to "H(Z) is minimal".

3. Error-, and cause-, controlled regulation

The reader may be wondering why error-controlled regulation has been omitted, but there has been no omission. Everything said so far is equally true of this case; for if the cause-effect linkages are as in Fig. 2, *R* is still receiving

Fig. 2



information about D's values, as in Fig. 1, but is receiving it after a coding through *S*. The matter has been discussed fully by Conant (1969). There he showed that the general formulation of Fig. 1 (which represents only that *R* must receive information from D by *some* route) falls into two essentially distinct classes according to whether the flow of information from D to Z is conserved or lossy. Regulation by error-control is essentially information-conserving, and the entropy of Z cannot fall to zero (there must be some residual variation). When, however, the regulator *R* draws its information directly from D (the cause of the disturbance) there need be no residual variation: the regulation may, in principle, be made perfect.

The distinction may be illustrated by a simple example. The cow is homeostatic for blood-temperature, and in its brain is an error-controlled center that, if the blood-temperature falls, increases the generation of heat in the muscles and liver – but the blood-temperature must fall first. If, however, a sensitive temperature-recorder be inserted in the brain and then a stream of ice-cold air driven past the animal, the temperature rises without any preliminary fall. The error-controlled reflex acts, in fact, only as reserve: ordinarily, the nervous system senses, at the skin, that the cause of a fall has occurred, and reacts to regulate before the "error" actually occurs. Error-controlled regulation is in fact a primitive and demonstrably inferior method of regulation. It is inferior because with it the entropy of the outcomes Z cannot be reduced to zero; its success can only be partial. The regulations used by the higher organisms evolve progressively to types more effective in using information about the causes (at D) as the source and determiner of their regulatory actions. *From here on, in this paper, we shall consider "regulation" of this more advanced, cause-controlled type* (though much of what we say will still be true of the error-controlled).

4. Models

Defining "regulation", as we have seen, is easy in that one is led rapidly to one of a few forms, closely related and easily distinguished in practical use. The attempt to define a "model", however, leads to no such focus. We shall obtain a definition suitable for this paper, but first let us notice what happens when one attempts precision. We can start with such an unexceptional "model" as a table-top replica of Chartres cathedral. The transformation is of the type, in three dimensions:

$$y_1 = k x_1$$
$$y_2 = k x_2$$
$$y_3 = k x_3$$

with k about 10^{-2}. But this example, so clear and simple, can be modified a little at a time to forms that are very different. A model of Switzerland, for instance, might well have the vertical heights exaggerated (so that the three k's are no longer equal). In two dimensions, a (proportional) photograph from the air may be followed by a Mercator's projection with distortion, that no longer leaves the variables separable. So we can go through a map of a subway system, with only the points of connection valid, to "maps" of a type describable only mathematically.

In dynamic systems, if the transformation converts the real time t to a model time t' also in real time we have a "working" model. An unquestionable "model" here would be a flow of electrons through a net of conducting sheets that accurately models, in real time, the flow of underground water in Arizona. But the model sailing-boat no longer behaves proportionately, so that a complex

relation is necessary to relate the model and the full-sized boat. Thus, in the working models, as in the static, we can readily obtain examples that deviate more and more from the obvious model to the most extreme types of transform, without the appearance of any natural boundary dividing model from non-model.

Can we follow the mathematician and use the concept of "isomorphism"? It seems that we cannot. The reason is that though the concept of isomorphism is unique in the branch where it started (in the finite groups) its extension to other branches leads to so many new meanings that the unicity is lost.

As example, suppose we attempt to apply it to the universe of binary relations. R, a subset of E x E, and S, a subset of F x F, are naturally regarded as "isomorphic" if there exists a one-one mapping σ of E into F such that $S = \sigma R \sigma^{-1}$ (Riguet 1948, 1951, Bourbaki 1958). But S and R are still closely related, and able to claim some "model" relationship if the definition is weakened to

$$\exists \sigma, \tau : S = \sigma R \tau^{-1}$$

(with τ also one–one). Then it can be weakened further by allowing ϕ (and τ) to be a mapping generally or even a binary relation. The sign of equality similarly can be weakened to "is contained in". We have now arrived at the relation given earlier (1) under "regulation"):

$$\rho \subset A \cdot \phi$$

which evidently implies some "-morphic" relation between ρ and ϕ (with A assumed given).

In this paper we shall be concerned chiefly with isomorphism between two dynamic systems (S and R in Fig. 1). We can therefore try using the modern abstract definition of "machine with input" as a rigorous basis.

To discuss iso-, and homo-, morphism of machines, it is convenient first to obtain a standard representation of these ideas in the theory of groups, where they originated. The relation can be stated thus:

Let the two groups be, one of the set E of elements e_i, with group operation (multiplication) δ, so that $\delta(e_i, e_{j'}) = e_{k'}$ and other similarly of δ' on elements F. Then the second is a homomorph of the first if and only if there exists a mapping h, from E to F, so that, for all $e_{i'}, e_j \in$ E:

$$\delta'[h(e_i), h(e_j)] = h[\delta(e_i, e_j)]. \tag{2}$$

If h is one–one onto F, they are isomorphic. This basic equation form will enable us to relate the other possible definitions.

Hartmanis and Stearns' (1966) definition of Machine M′ being a homomorphism of M follows naturally. Let machine M have a set S of internal states,

a set I of input-values (symbols), a set O of output-values (symbols), and let it operate according to δ, a mapping of S X I to S, and λ, a mapping of S X I to O. Let machine M′ be represented similarly by S′, I′, O′, δ', λ'. Then M′ is a homomorphism of M if and only if there exist three mappings:

h_1, of S to S′
h_2, of I to I′
h_3, of O to O′

such that, for all $s \in$ S and $i \in$ I:

$$\left.\begin{aligned} h_1[\delta(s,i)] &= \delta'[h_1(s), h_2(i)], \\ h_3[\lambda(s,i)] &= \lambda'[h_1(s), h_2(i)]. \end{aligned}\right\} \tag{3}$$

This definition corresponds to the natural case in which corresponding inputs (to the two machines) will lead, through corresponding internal states, to corresponding outputs. But, unfortunately for our present purpose, there are many variations, some trivial and some gross, that also represent some sort of "similarity". Thus, a more general form, representing a more complex form of relation, would be given if the mappings

$$h_1, \text{ of S to S', and } h_2, \text{ of I to I',}$$

were replaced by one mapping

$$h_4, \text{ of I} \times \text{S to I'} \times \text{S'}.$$

(More general because h_4 may or may not be separable into h_1 and h_2). Then the criterion would be,

$$\forall i, s : \delta'[h_4(s,i)] = h_4[\delta(s,i)], \tag{4}$$

a form not identical with that at (3).

There are yet more. The "Black Box" case ignores the internal states S, and treats two Black Boxes as identical if equal inputs given equal outputs. Formally, if μ and μ' are the mappings from input to output, then the second Box is a homomorphism of the first if and only if there exists a mapping h, of I to I′, such that:

$$\forall i \in \text{I} : \mu'[h(i)] = h[\mu(i)]. \tag{5}$$

Here it should be rembered that equality of outputs is only a special case of correspondence. Also closely related are two Black Boxes such that the second is "de-coder" to the first: the second, given the first's output, will take this as input and emit the original input:

$$\forall i \in \text{I} : \mu'\mu(i) = i. \tag{6}$$

This is an isomorphism. In the homomorphic relation, the input i and the final output $\mu'\mu(i)$ would both be mapped by h to the same class:

$$\forall i \in I : \; h\mu'\mu\,(i) = h(i). \tag{7}$$

These examples may be sufficient to show the wide range of abstract "similarities" that might claim to be "isomorphisms". There seem, in short, to be as many definitions possible to isomorphism as to model. It might seem that one could make practically any assertion one likes (such as that in our title) and then ensure its truth simply by adjusting the definitions. We believe, however, that we can mark out one case that is sufficiently a whole to be worth special statement.

We consider the regulatory situation described earlier, in which the set of regulatory events R and the set of events S in the rest of the system (i.e., in the "reguland" *S*, which we view as *R*'s opponent) jointly determine, thought a mapping ψ, the outcome events *Z*. By an optimal regulator we will mean a regulator which produces regulatory events in such a way that *H(Z)* is minimal. Then under very broad conditions stated in the proof below, the following theorem holds:

Theorem: The simplest optimal regulator *R* of a reguland *S* produces events R which are related to the events S by a mapping *h:* S→R.

Restated somewhat less rigorously, the theorem says that the best regulator of a system is one which is a model of that system in the sense that the regulator's actions are merely the system's actions as seen through a mapping *h*. The type of isomorphism here is that expressed (in the form used above) by

$$\exists h : \; \forall i : \; \rho(i) = h[\sigma(i)] \tag{8}$$

where ρ and σ are the mappings that *R* and *S* impose on their common input I. This form is essentially that of (5) above.

Proof: The sets R, S, and Z are the mapping ψ: R X S → Z are presumed given. We will assume that over the set S there exists a probability distribution $p(S)$ which gives the relative frequencies of the events in S. We will further assume that the behavior of any particular regulator *R* is specified by a conditional distribution $p(R|S)$ giving, for each event in S, a distribution on the regulatory events in R. Now $p(S)$ and $p(R|S)$ jointly determine $p(R,S)$ and hence $p(S)$ and *H(Z)*, the entropy in the set of outcomes. ($H(Z) \equiv -\sum p(z_k) \log p(z_k)$.) With $p(S)$ fixed, the class of optimal regulators therefore corresponds to the class of optimal distributions $p(R \mid S)$ for which $H(Z)$ is minimal. We will call this class of optimal distributions π.

It is possible for there to be very different distributions $p(Z)$ all having the same minimal entropy $H(Z)$. To consider that possibility would merely complicate this proof without affecting it in any essential way, so we will suppose that every $p(R|S)$ in π determines, with $p(S)$ and ψ, the same (unique) $p(Z)$.

We now select for examination an arbitrary $p(R|S)$ from π.

The heart of the proof is the following lemma:

Lemma: $\forall s_j \in S$, the set $\{ \psi(r_i, s_j) > 0 \}$ has only one element. That is, for every s_j in S, $p(R|s_j)$ is such that all r_i with positive probability map, with s_j under ψ, to the same z_k in Z.

Proof of lemma: Suppose, to the contrary, that $p(r_1|s_j) > 0$, $p(r_2|s_j) > 0$, $\psi(r_1, s_j) = z_1$, and $\psi(r_2, s_j) = z_2 \neq z_1$. Now $p(r_1, s_j)$ and $p(r_2, s_j)$ contribute to $p(z_1)$ and $p(z_2)$ respectively, and by varying these probabilities (by subtracting Λ from $p(r_1, s_j)$ and adding Δ to $p(r_2, s_j)$ we could vary $p(z_1)$ and $p(z_2)$ and thereby vary $H(Z)$. We could make Δ either positive or negative, whichever would make $p(z_1)$ and $P(z_1)$ more unequal. One of the useful and fundamental properties of the entropy function is that any such increase in imbalance in $p(Z)$ necessarily increases $H(Z)$. Consequently, we could start with a $p(R|S)$ resulting in a lower $H(Z)$; this contradiction proves the lemma.

Returning to the proof of the theorem, we see that for any number of π and any s_j in S, the values of R for which $p(R|s_j)$ is positive all give the same z_k. Without affecting $H(Z)$, we can arbitrarily select one of those values of R and set its conditional probability to unity and the others to zero. When this process is repeated for all s_j in S, the result must be a member of π with $p(R|S)$ consisting entirely of ones and zeroes. In an obvious sense this is the *simplest* optimal $p(R|S)$ since it is in fact a mapping *h* from S into R. Given the correspondence between optimal distributions $p(R|S)$ and optimal regulators *r*, this proves the theorem.

The Theorem calls for several comments. First, it leaves open the possibility that there are regulators which are just as successful (just as "optimal") as the simplest optimal regulator(s) but which are unnecessarily complex. In this regard, the theorem can be interpreted as saying that although not all optimal regulators are models of their regulands, the ones which are not are all unnecessarily complex.

Second, it shows clearly that the search for the best regulator is essentially a search among the mappings from S into R; only regulators for which there is such a mapping need be considered.

Third, the proof of the theorem, by avoiding all mention of the inputs to the regulator *R* and its opponent *S*, leaves open the question of how *R, S* and *Z* are interrelated. The theorem applies equally well to the configurations of Fig. 1 and Fig. 2, the chief difference being that in Fig. 2 *R* is a model of *S* in the sense that the events R are mapped versions of the events S, whereas in Fig. 1 the modelling is stronger; *R* must be a homo- or isomorph of *S* (since it has the same input as *S* and a mapping-related output).

Last, the assumption that $p(S)$ must exist (and be constant) can be

weakened; if the statistics of S change slowly with time, the theorem holds over any period throughout which $p(S)$ is essentially constant. As $p(S)$ changes, the mapping h will change appropriately, so that the best regulator in such a situation will still be a model of the reguland, but a time-varying model will be needed to regulate the time-varying reguland.

5. Discussion

The first effect of this theorem is to change the status of model-making from optional to compulsory. As we said earlier, model-making has hitherto largely been suggested (for regulating complex dynamic systems) as a possibility: the theorem shows that, in a very wide class (specified in the proof of the theorem), success in regulation implies that a sufficiently similar model must have been built, whether it was done explicitly, or simply developed as the regulator was improved. Thus the would-be model-maker now has a rigorous theorem to justify his work.

To those who study the brain, the theorem founds a "theoretical neurology". For centuries, the study of the brain has been guided by the idea that as the brain is the organ of thinking, whatever it does is right. But this was the view held two centuries ago about the human heart as a pump; today's hydraulic engineers know too much about pumping to follow the heart's method slavishly: they know what the heart ought to do, and they measure its efficiency. The developing knowledge of regulation, information-processing, and control is building similar criteria for the brain. Now that we know that any regulator (if it conforms to the qualifications given) must model what it regulates, we can proceed to measure how efficiently the brain carries out this process. There can no longer be question about *whether* the brain models its environment: it must.

REFERENCES

ASHBY, W. ROSS, 1967, ***Automaton Theory and Learning Systems,*** edited by D.J. Stewart (London: Academic Press), p. 23-51.

BOURBAKI, N., 1958, ***Theorie des Ensembles; Fascicule de Resultats,*** 3rd edition (Paris: Hermann).

CONANT, ROGER C., 1969, ***I.E.E.E. Trans. Systems Sci.,*** 5, 334.

HARTMANIS, J., and STEARNS, R.E., 1966, ***Algebraic Structure Theory of Sequential Machines*** (New York: Prentice-Hall).

RIGUET, J., 1948, ***Bull. Soc. math. Fr.,*** 76, 114; 1951, These de Paris.

SOMMERHOFF, G., 1950, ***Analytical Biology*** (Oxford University Press).

V.

THE ANALYSIS OF CONSTRAINTS

THE ANALYSIS OF CONSTRAINTS

INTRODUCTION

By now the reader must be fully aware of the importance Ashby placed on constraints. He pointed out that they are the essense of organization and hence of structure in multivariable systems; he showed that they can be measured with information theory; further, he showed that when a dynamic deterministic system is allowed to run to equilibrium, constraints will appear which ultimately explain such phenomena as adaptive behavior and intelligence. In this section are several key papers concerned with the analysis of constraints. The first paper, "General Systems Theory as a New Discipline," is included here because it discusses, along with other matters of interest, the deduction of the structure of a "black box," i.e. a mechanism whose laws are hidden from the investigator. The only course open to the investigator is to search for constraints in the behavior and responses of the box, and this search is nicely characterized by Ashby as a communication process between box and investigator. Being a process of information transfer it is of course subject to the laws of information mentioned in the previous chapters.

"The Constraint Analysis of Many-dimensional Relations" has been an influential paper in that it inspired the subsequent interest in structure modelling or reconstructability analysis. It is based on the observation that most systems of many variables contain a hidden simplicity, in that they are not "fully connected" (each variable interacting with every other) but are "locally connected" (each variable interacting with only a few others). Indeed from other Ashby works the inference may be drawn that fully connected systems must be extremely rare, because of their tendency to be unstable and hence self-destructive. Where a constraint exists (in this case, a limit on connections), some advantage may generally be taken of it, and the discovery of this type of simplicity in an apparently complex system is extremely valuable and important in system studies.

The paper introduces the notions of cylindrical closure and of cylindrance of a relation. If an n-variable relation R has a cylindrance of p, it may be expressed as an aggregate of the C_p^n implicit relations (projections of R) each involving exactly p variables, and if p is much smaller than n this can be an immensely important simplification. Since the writing of the paper subsequent research [105] has generalized Ashby's result, first by expressing an n-variable relationship as an aggregate of lower-order subrelations not necessarily all of the same order p, and secondly by applying similar methods to probabilistic relationships, but Ashby's paper is the intellectual ancestor of all such efforts.

The most mathematically sophisticated paper in this collection is "The Identification of Many-dimensional Relations," written with Robert Madden and to a large extent representing Madden's doctoral work carried out under Ashby's direction. It deals with the problem of identifying an n-dimensional relation R from the collection of its p-dimensional projections and thus is relevant to the question of determining a relation from partial information, reminiscent of the fable of the seven blind men and the elephant. A major conclusion of the paper is that in general there is not enough information in the lower-order projections to determine R unambiguously.

GENERAL SYSTEMS THEORY AS A NEW DISCIPLINE*

The emergency of general system theory is symptomatic of a new movement that has been developing in science during the past decade: Science is at last giving serious attention to systems that are intrinsically complex. This statement may seem somewhat surprising. Are not chemical molecules complex? Is not the living organism complex? And has not science studied them from its earliest days? Let me explain what I mean.

Science has, of course, long been interested in the living organism; but for 200 years it has tried primarily to find, within the organism, whatever is *simple.* Thus, from the whole complexity of spinal action, Sherrington isolated the stretch reflex, a small portion of the whole, simple within itself and capable of being studied in functional isolation. From the whole complexity of digestion, the biochemist distinguished the action of pepsin or protein, which could be studied in isolation. And avoiding the whole complexity of cerebral action, Pavlov investigated the salivary conditioned reflex – an essential simple function, only a fragment of the whole, that could be studied in isolation.

The same strategy – of looking for the simple part – has been used incessantly in physics and chemistry. Their triumphs have been chiefly those of identifying the *units* out of which the complex structures are made. The triumph has been in analysis, not in synthesis. Thus today the biochemist knows more about the amino-acids of which egg-protein is composed than he does about the white of egg from which they have been obtained. And the physiologist knows more about the individual nerve cell in the brain than he does about the action of the great mass of them in integration.

Thus until recently the strategy of the sciences has been largely that of analysis. The units have been found, their properties studied, and then, somewhat as an after-thought, some attempt has been made to study them in combined action. But this study of synthesis has often made little progress and does not usually occupy a prominent place in scientific knowledge.

Even when a study of synthesis seems to be made, the synthesis is often found, on closer examination, to be that in which the interaction between the parts is as slight as possible. We notice for instance how often the combinations that are treated in physics and chemistry occur under the operation of simple addition. Thus two masses in the pan of a balance have a mass that is the simple

---

* Based on an address presented to the meeting of the Society for General Systems Research at Atlanta, Georgia, December 27, 1955.

sum of the separate masses. Similarly two wave forms in an electrical network are usually studied in the *linear* case – the case in which the two patterns combine by simple addition.

Now combination by simple addition is the very next thing to no combination at all. Thus one penny combines with one penny to give just two, precisely because pennies do not in fact interact to any appreciable extent. Contrast this merely nominal combination with what happens when, say, acid is brought together with alkali, or rabbit is brought together with rabbit. Here there is real interaction, and the outcome cannot be represented as a simple sum. Thus, for a century or more, science has advanced chiefly by analyzing complex wholes into simple parts. Synthesis has, on the whole, been neglected.

The rule "analyze into parts, and study them one at a time" was so widely followed that there was some danger of its degenerating into a dogma; and the rule was often regarded as the touchstone of what was properly scientific.

Perhaps the first worker to face squarely up to the fact that not all systems allow this analysis into single parts was Sir Ronald Fisher. His problem was to get information about how the complex system of soil and plants would react to fertilizers by giving crops. One method of study is to analyze plant and soil into a host of little physical and chemical subsystems, get to know each subsystem individually, and then predict how the combined whole would respond. He decided that this method would be far too slow, and that the information he wanted could be obtained by treating soil and plant as a complex whole. So he proceeded to conduct experiments in which the variables were not altered one at a time.

At first, scientists were shocked; but second thoughts have convinced us that his methods were sound. Thus Fisher initiated a new scientific strategy. Faced with a system of great complexity, he accepted the complexity as an essential, a non-ignorable property; and showed how worthwhile information could be obtained from it. He also showed that this could be done only if the worker accepted the need for a new scientific strategy.

What I have said is, of course, equivalent to saying that whereas physics and chemistry, given a system, promptly breaks it to pieces in order to study the parts, there is arising a new discipline that studies the system without breaking it to pieces. The internal interactions are left intact, and the system is, in the well known words, studied as a whole. What methods are there for the study of such *intact* systems? What general methods, in other words, can general system theory follow?

Two main lines are readily distinguished. One, already well developed in the hands of von Bertalanffy and his co-workers, takes the world as we find it, examines the various systems that occur in it – zoological, physiological, and

so on – and then draws up statements about the regularities that have been observed to hold. This method is essentially empirical.

The second method is to start at the other end. Instead of studying first one system, then a second, then a third, and so on, it goes to the other extreme, considers the set of "all conceivable systems" and then reduces the set to a more reasonable size. This is the method I have recently followed. Since it may seem, at first sight to be somewhat recklessly speculative, I would like to consider briefly its justification.

The method of considering *all* possible systems, regardless of whether they actually exist in the real world, has already been used, and shown its value, in many well established sciences. Crystallography, for instance, studies on the one hand those crystals that actually occur in nature; and it also studies, in its mathematical branch, all forms that are conceptually possible. It has been found that the set of all conceivable crystals must still obey certain laws; and mathematical crystallography can make confident predictions about what will be found in certain cases.

Whence come these laws? Dare one dogmatize about what nature may do? In this case one can, and the reason is that we have the option of saying what *we* mean by a "crystal". When we define it as something that shows certain properties of symmetry, we can go on to say that it must also show certain other properties of symmetry because the latter are necessarily implied by the former – they are, as it were, the same properties seen from another viewpoint.

Mathematical crystallography thus forms a background or framework, more comprehensive than the empirical material, on which the empirical material – the real crystals – can find their natural places and be appropriately related to one another. Few will deny the value of the mathematical theory, for without it the study would be a chaos of special cases.

The method of considering more than the actual has also long been used, to advantage, in physics. Much of its theory is concerned with objects that do not exist and never have existed: particles with mass but no volume, pulleys with no friction, springs with no mass, and so on. But to say that these entities do not exist does not mean that mathematical physics is mere fantasy. The massless spring, though it does not exist, in the real world, is a most important concept; and a physicist who understands its theory is better equipped to deal with, say the balance of a watch than one who has not mastered the theory.

I would suggest that a similar logical framework would be desirable as a part of general system theory. The forms occurring in the real world are seldom an orderly or a complete set. If they are to be related to one another, and higher relations and laws investigated, a rigorous logic of systems must be developed, forming a structure on which all the real forms may find their natural places

and their natural relations.

Can such a structure be developed? Can one reasonably start by considering the class of "all conceivable systems"? I suggest one can.

The first objection to be met is that the class is ridiculously wide. It includes for instance the "system" that consists of the three variables: the temperature of this room, its humidity, and the price of dollars in Singapore. Most people will agree that this set of variables, as a "system", is not reasonable, though it certainly exists. How then do we proceed?

Consideration of many typical examples shows that the scientist is, in fact, highly selective in his choice of systems for study. Large numbers of possible aggregations of variables are dismissed by him as "not suitable for study". The criteria he uses are often well known to him intuitively, though seldom stated explicitly. What is often also not recognized explicitly is the intensity of the selection used. Eddington tells the story of the empirical scientist who threw a net into the sea, examined the catch, and then announced the empirical law "all sea creatures are more than two inches long." In studying systems we do not, one hopes, proceed quite as naively as this; but that there *are* subtle laws that have an epistemological, rather than an empirical, basis, can hardly be doubted. Thus while little can be said about "all possible" systems, a good deal can be said about the very special sub-set of those systems that are accepted by the scientist as being "suitable for study". I shall give some examples a little later.

What is the criterion that the scientist applies when he decides whether a proposed set of variables does or does not form a "natural" system? We can see something of what is necessary by first thinking of the parallel case in energetics. For a system to be suitable for study by the physicist no energy must enter or leave it except as the experimenter directs. Such a system is usually described as "closed" to energy, but the adjective is not well chosen, as often an important part of the investigation is the addition of, say, a measured quantity of heat to it to provoke changes. I shall refer to such a system as "energy-tight".

In the same way, the systems suitable for study in the biological world, while freely open to energy, must be closed to all sources of disturbance, or variation, or entropy (in Shannon's sense) except as directed by the experimenter. They must be, in the technical sense, "information-, or noise-tight". This is the net that catches the systems that come to the empirical scientist for study. It imposes a considerable degree of selection from the set of all conceivable systems. And the selection imposes a number of special properties on the systems that conform to it. Some of the properties are obvious – we need not bother with them; but some are subtle and appear only in disguised form: they have to be discovered, and their true origin identified. Thus we can now ask: how can we identify those properties of a system that are direct consequences of the scientist's insistence that it shall be information-tight?

The Black Box

To answer this question there is, in my opinion, no finer approach than that given by the so-called Problem of the Black Box. It arises in electrical engineering, but its range is really far greater – perhaps as great as the range of science itself.

We imagine that the Investigator has before him a Black Box that, for any reason, cannot be opened. It has various inputs – switches that he may move up or down, terminals to which he may apply potentials, photoelectric cells on to which he may shine lights, and so on. Also available are various outputs – terminals on which a potential may be measured, lights that may flash, pointers that may move over a graduated scale, and so on. The Investigator's problem is do what he pleases to the inputs, and to make such observations on the outputs as he pleases, and to *deduce* what he can of the Box's contents.

In its original, specifically electrical, form, the problem was to deduce the contents in terms of known elementary components. Our problem however is somewhat wider. The questions we are interested in, in general system theory, are such matters as:

What *general* rules of strategy should guide the exploration, when the Box is not limited to the electrical but may be of any nature whatever?

When the raw data have been obtained from the outputs, what operations should *in general* be applied to the data if the deductions made are to be logically permissible?

Finally, the most basic question of all:

What can in principle be deduced from the Box's behavior, and what is fundamentally not deducible? That is: given that the Investigator has certain finite resources for exploration and observation, what limitations does this finiteness impose on his knowledge of the Black Box?

At first, the questions may seem to be too general to be answerable, but it is now clear that this is not so. The modern development of communication theory can give substantial guidance in this matter, for what *we* are considering can be viewed as a compound system, composed of Box and Investigator. He acts on the Box when he stimulates it, and the Box acts on him when it gives him a dial-reading as observation. Thus each acts on the other. The interactions that occur between them are as subject to the laws of communication as any other interaction between two sub-systems. (I should make it clear

here that the communication theory involved is not the theory which is restricted to the ergodic case.)

In practice, there are no absolute bounds given in relation to any particular Black Box. By however many ways we have tested it there are always further ways at least conceivable. By however many senses or instruments we have observed it, there are always further ways. For however long we have observed it, we could always go on longer. Eventually however the time will always come, for practical reasons, when the exploration and observation must stop – at least for the time being, when the scientist stops to *think* about the Box, and to draw deductions from his data. I shall assume that from now on the investigation has reached this stage. Thus certain definite inputs have been used, certain variables observed, and a protocol of finite length recorded.

It is now axiomatic that whatever the interaction, it will eventually appear as a protocol of events, stating in general the succession of states taken by each part as they occurred in time. This protocol can now be regarded as a message that contains information about the Box's nature. It is axiomatic that: the Investigator's knowledge about the Box, in any of its aspects, must be essentially a re-coding of what is in the protocol; he may not claim more.

From this point of view, to discover something about the Black Box – something that has permanence – is to discover a constraint in the protocol. The study of a system can thus be summed up in a few words: to discover the constraints, the statistical structure, in the protocol. Should the Investigator find none – should the protocol, as an information source, show maximal entropy, and therefore no redundancy – then he will say, simply, "I can make *nothing* of its behavior; it is totally chaotic." Thus, *any deduction about the nature of the Black Box must be essentially a re-coding of the redundancies (or constraints) in its protocol.*

Suppose now that the Investigator announces that he has discovered a "property" of the Box – some characteristic of its behavior that holds all the way through the protocol. If he describes the property in a suitably compact statement, we can see that he is carrying out precisely the process so important in communication: he is re-coding the protocol so as to pass on a simpler message that still contains the important information but without redundancy; for the redundant part is passed on once and for all by a compact statement about this permanent "property".

Perhaps the most important property that is testable on the protocol is that of whether the system is suitable for study at all! That is, whether it is information-tight. What this means, in essence, is that the protocol should be invariant in time, in regard to the constraints it shows. If this is so, the Investigator may legitimately claim that the system, at least so far as this property is concerned, is information-tight, that is, not subject to unpredictable vagaries. In fact, at the level of fundamental concepts, such invariance may be regarded as the operational definition of what is meant by "information-tight".

The next fundamental property that can be deduced from the protocol is whether the system is or is not behaving in a "machine-like" way. By that I mean whether knowledge of its present state (as shown at the output) and of the conditions within which it is working (that is, the state of its input) is *sufficient* to determine what it will do next. Whether the Box is behaving in a machine-like way does not require study of its internal details; by a straightforward, operationally-defined process the question can be answered from the protocol. Space does not allow me to describe the method at the moment. I will merely remark that the method also allows the Investigator to establish whether the system, though not strictly determinate, is determinate in the statistical sense of behaving with an unvarying probability.

An important question that often arises in the investigation of any particular system is: what functional connections exist between the parts? (Here I refer exclusively to the functional aspects, that is, of what affects what.) When we know the connections between the parts we often draw a diagram of the immediate, or direct effects. So we get the endocrinologist's diagram showing how the various glands, tissues, and nervous centers act on each other; and the business administrator's diagram showing how the various departments are connected. Can such a diagram be obtained from a Black Box by deduction, from the protocol?

It can, up to an isomorphism.

Let me ignore the qualification for a moment. The fact then is that *functional connections within a Black Box can be deduced from observations made from without.* Information in this respect is to be found in the protocol. To find something of the connections does not demand the opening of the Box.

But not everything of the internal connections can be so deduced. The protocol contains information about the connections that will enable us to specify the connections only up to an isomorphism. No re-coding of the protocol can go past this limit: the information necessary just does not exist in the protocol.

To see what this limitation means, let me make clear what is meant by two Black Boxes being "isomorphic."

Suppose we have before us two Black Boxes. We are privileged to look inside.

The first contains a heavy wheel, which can rotate, and a dial outside showing its position. Attached to the wheel is a spring, and the input is a lever attached to the spring's other end. So moving the lever distorts the spring and

applies a force to the wheel, making it turn, or perhaps to stop turning. If the lever is given a complicated sequence of positions, the wheel will respond with some complicated sequence of turning, which will show on the dial.

The second Box is electrical, and contains an inductance and a capacitance in series. The input is a lever that controls a variable potential; it is under the experimenter's control and can be varied in any arbitrary way. Recorded as output on a dial is the total amount of current that passes round the circuit. If now the experimenter moves the lever in some arbitrary way, the potential will affect the components, causing a varying current, which will show as a complex sequence of changes at the output.

Let us further suppose that the various constants – stiffness of spring, mass of wheel, inductance, and capacitance – have been set once and for all so as to impose a certain relationship between the two systems. (As this example is for illustration only, I need not specify further.)

It will then be found that when the first Box is taken, and a particular input applied to the spring, and the consequent movement of the wheel observed, application to the second Box of the *same* values to *its* input will evoke the *same* pattern of change at the output. In other words, equality of the two inputs is always followed by equality of the two outputs. In fact, through an infinity of possibilities, whenever the two Boxes are given the same trajectory of input, no matter how long or complex, the outputs, however long or complex, will also be equal. Thus if the actual mechanisms are covered up, leaving only inputs and outputs visible, the two Boxes become indistinguishable, for they will respond similarly to all possible tests that can be applied. The two machines are then said to be "isomorphic".

Suppose now that a differential analyzer has been programmed to predict the behavior of one of the Boxes for all possible inputs. Since the analyzer, when given an input, gives the same output as the Box, the analyzer is by definition isomorphic with the Box. Thus an analogue computer might correctly be defined as a machine that can easily be made isomorphic with any of a wide class of dynamic systems.

It has been shown, as I said, that the internal connections can be deduced from the protocol *up to an isomorphism*. It is also readily provable that no deduction, on a given protocol, can go further. We thus encounter here one of the fundamental limitations that I spoke of earlier.

It must not be thought that this limitation is merely technical, to be swept away by the invention of some new gimmick. What it means is that any finite protocol can give only a certain amount of information about any particular question of connection; and that for further questions the information in the protocol *does not exist*.

Knowing that this limitation exists may sometimes be of value in saving us from attempting the impossible. Thus suppose we have before us not only the two Black Boxes just mentioned, *and* the differential analyzer suitably programmed to copy them, but also an engineer, so well trained that, if told the input he can predict what the output will be, perhaps by drawing a graph of the actual changes. If we regard *him* as a neuronic mechanism, then we have said that this mechanism is itself isomorphic with the other three; for, given the same input to all four, all four will produce identical outputs. If now we become neurophysiologists, and start to think about the engineer's brain and the neuronic connections within it, we are warned before we start that the pattern of connections is not uniquely defined by the engineer's behavior. Anything we say can only be about a class of mechanisms. And this fact should be reflected in what we try to say about the mechanism. Some of our difficulties in treating the theory of these neuronic mechanisms may be due to our tending to forget this fact.

Degrees of freedom

Yet another characteristic of the Black Box that can be deduced rigorously from the protocol is the number of its degrees of freedom. By this is meant the number of variables that must be observed or specified if its behavior is to become determinate, that is, unique, single-valued, not subject to random variations. As example, take the simple pendulum of fixed length. It has a determinate trajectory only if both its position and its velocity are specified; so it has two degrees of freedom. And a desk calculator that multiplies eight figures by eight has 16 degrees of freedom, for only when 16 figures have been specified does the number that will appear as product become determined.

To return to the Black Box. We assume that its output is shown on a row of dials. Now this row may show what is occurring internally either completely or only partially. Thus, if the Box really contained a simple pendulum, the output might tell us only its *position* at each moment. Study of this Box would soon show that knowledge of its output was not sufficient to make the output's behavior predictable.

Now the number of degrees of freedom is an intrinsic property of a system and can be deduced by finding how many observations have to be made if the behavior is to become predictable. Thus suppose that some new Box has actually 20 degrees of freedom internally, and that five dials are reporting on the events within. With 15 degrees of freedom unobserved, the Investigator will find that the behavior of the Box, as shown on the dials, is not determinate. (The apparent indeterminacy comes from the fact that 15 variables are not

being taken account of.) What is important is that the Investigator can restore determinacy by taking account of earlier values of the variables he *can* see on the dials. And as 15 degrees of freedom are not directly observable, the necessary information can be obtained by making an extra 15 observations on the same five dials (three on each, say).

Thus Black Box theory leads us naturally into the theory – most important for those who study the brain – of the mechanism that, for whatever reason, is not wholly accessible to observation.

Thus we are led to a statement that can be proved rigorously (though for simplicity I shall omit here the qualifications that are strictly necessary): – When a system is really determinate, but cannot be observed at every significant point, determinacy can be restored by the use of supplementary observations, at the same points, that is, on the same dials, about what happened earlier. And the total number of observations to be made must always equal the number of the system's degrees of freedom. In other words, we can find how many degrees of freedom the Black Box has by finding how many observations on the dials are necessary to make correct prediction possible.

Memory

I have just said that when the Box is not completely observable, the Investigator may restore predictability *by taking account of what happened earlier.* Now this process of appealing to earlier events is also well known under another name. Suppose, for instance, that I am at a friend's house and, as a car goes past outside, his dog runs to a corner of the room and cringes. To me the behavior is causeless and inexplicable. Then my friend says "He was run over by a car a month ago." The behavior is now accounted for *by my taking account of what happened earlier.*

The psychologist would say I was appealing to the concept of "memory", as shown by the dog. What we can now see is that the concept of "memory" arises most naturally in the Investigator's mind when not all of the system is accessible to observation, so that he must use information of what happened earlier to take the place of what he cannot observe now. "Memory", from this point of view, is not an objective and intrinsic property of a system, but a reflection of the Investigator's limited powers of observation. Recognition of this fact may help us to remove some of the paradoxes that have tended to collect around the subject.

I referred earlier to the fact that when the scientist decides to include only "natural" or "reasonable" systems in the set he studies, he selects somewhat intensively, and may well impose peculiarities that are later discovered empirically, just as the ichthyologist discovered that all sea creatures exceed two inches in length.

We are not surprised, then, when study of the Black Box shows that certain properties, long known to be common in the real world, are necessary consequences of *our* act in only accepting for study such systems as are information- and noise-tight.

Space is running out, so I must pass over these properties somewhat briefly.

The first depends on the fact that every noise-tight system, if subjected to no disturbance at its input, that is if "isolated", cannot *gain* information. Any change in the quantity of the information can then only be a decrease. Every isolated system shows this decrease when it goes to a state of equilibrium; for when many trajectories, from many distinct initial states, converge to one state of equilibrium, the system, when at the equilibrium, has lost the information about which initial state it came from. This is a first example of the general principle that information about what happened earlier in the system tends always to decay.

A more elaborate instance of what is essentially the same principle occurs when a noise-tight system is subjected to a long sequence of events as inputs. Let us regard the system's state now as showing various traces of what has happened to it in the past. If, as is usually the case, the system's capacity for information is finite, information about what has happened to it in the remoter past tends to be swamped and destroyed by information about what has happened recently. More formally: if a noise-tight system is subjected to a long sequence of events as input, then the longer the sequence, the more will its final state depend on which sequence was applied rather than upon which state the system happened to start from.

In psychology the phenomenon has long been known in various forms. We know it from every day experience when we notice that a group of boys of varied characteristics, if put through a uniform experience, such as being sent to sea, becomes later more characterized by the fact that they are sailors than by their previous idiosyncracies. The same phenomenon has also been encountered in the laboratory as retroactive inhibition, which names the fact that later learning tends to destroy earlier learning.

Various more or less complex mechanisms have been invented to explain these well known phenomena. The possibility however exists that they may in some cases be due to the fact that the Scientist will only investigate such systems as are information-tight. He thereby unwittingly selects such systems as must show the phenomenon to some degree.

To conclude, let me offer some justification for the title of this paper, which suggests that general system theory should be regarded as a new discipline.

You will have noticed that a good deal of what I have had to say has not been concerned directly with the Black Box but rather with what the Investigator can or cannot achieve when faced with one. *We*, the system theorists, have in fact been studying, not a Black Box, but a larger system composed of two parts, the Black Box and the Investigator, each acting on the other. We have used communication theory in its non-ergodic form to deduce something of the laws of their interacting. Thus if the Investigator is a scientist studying the Box, *we* are meta-scientists, for we are studying both; we are working at an essentially different level.

What I have been able to say cannot do more than to introduce the general idea of a mathematical aspect of general system theory. I hope I have made clear that it is possible and reasonable to work not upwards from the empirical but downward from the abstract and general. I hope I have shown that such a study promises worthwhile results, and that it may help to provide us with what is urgently needed in our studies of such complex systems as the brain and society, namely, a logic of mechanism.

CONSTRAINT ANALYSIS OF MANY-DIMENSIONAL RELATIONS

W. ROSS ASHBY

University of Illinois, College of Engineering, Urbana, Ill. (U.S.A.)

That this article should be offered as a tribute to Norbert Wiener is specially appropriate, for it takes as basis an observation of his that has not yet shown, I feel sure, its full fertility. I refer to his original suggestion (Wiener, 1914) that a 'relation', previously regarded as somewhat metaphysical, be identified, at least for operational purposes, with the set of those n-tuples that satisfy the relation. At one stroke the 'relation' becomes an ordinary mathematical object that can be subjected to the ordinary mathematical operations, even when the relation is wholly arbitrary.

To the biologist, the freedom that allows it to be wholly arbitrary is most welcome, for in his science, though relations are of the greatest importance, they seldom have the tidiness common to more formal mathematics.

The attempt to apply Relation Theory to the biological sciences soon runs into great difficulties however. As soon as the biologist attempts to deal with realistically large numbers (*e.g.* with 10^{10} nerve cells) or with realistically intricate patterns of interaction (*e.g.* those between species in the Amazon jungle) the combinatorial possibilities soon generate fantastically large numbers; equally fantastic is the quantity of information-processing demanded of any system (cerebral or electronic) that would handle the questions involved. Exponentials, factorials, or even more explosively increasing functions appear. Bremermann (1962) has shown that no computer made of matter, and therefore subject to the mass-energy relation and to Heisenbergian uncertainty, can possibly process more than 1.4×10^{47} bits per g per sec; so 10^{70}, say, is certainly an absolute upper bound to what is practical. Yet even the simplest problems with more than a few variables and more than infinitesimal interaction generate numbers vastly greater than 10^{70}. An example is given below. Any science, such as cybernetics, that would treat large systems with strong interactions, urgently needs

methods by which the excessively complex can be reduced to complexities that are within our resources. In this paper is described one such method.

The method is based on the common observation that when the number of variables is large — a thousand and over, say — many of the significant relations are not really intricate to the full degree, but are really built out of simpler relations. In dynamics, for instance, the linear system, both common and important, has the peculiarity that the complicated output evoked by a complicated input can be found by simply adding a number of simple outputs, each evoked by a simple input. Thus in this case the whole output–input relation is really composed of many simple relations that combine only by adding. The parts of the system interact, but not the sub-relations.

Again, a camera lens with ten elements and fifteen surfaces at first seems optically very complex; yet in fact the total effect, from incident ray to emergent ray, can be obtained by merely repeating one *ternary* relation (incident ray/surface/refracted ray) fifteen times in succession. Thus the lens designer is able to avoid the fantastic combinatorial possibilities initially presented by the 15-variable relation.

Not only the physical sciences but everyday life shows the same feature. 'The Law', as it affects John Citizen, has hundreds, even thousands, of variables. Yet it can, in fact, be dealt with piecemeal; for it is built by the intersection of such sub-relations as: Drivers of age x may drive only automobiles of class y: Stores selling goods p must be closed on days q.

The set of all events that are 'legal' is then obtained by simple compounding of all the sub-relations, each of which uses only a tiny fraction of the totality of variables.

An indication of what threatens may emphasize the point here. If there are n variables, and each variable can take k values, the number of relations possible (*e.g.* of Laws that J. C. may face), as subsets of the product set, is

$$2^{(k^n)}.$$

As a function of n, it is an exponential of an exponential, a rate of increase vastly more 'explosive' than those commonly encountered in other branches of science. If k is merely 10, for instance, by the time n has risen to five (well below the 15 of the 'optical' example) the number of relations has risen to about $10^{30\,000}$, a number that shows how intensely restrictive Bremermann's limit really is. Let us then consider how one complex n-ary relation may be reduced to a set of simpler relations.

When the dimensions are 2, the relation, as a subset of a product-set $E \times F$, has only the simplifying possibility that it is itself a product set,

$A \times B$ say, with $A \subset E$ and $B \subset F$. This simplification is too extreme to interest us here, for it reduces the relation to the mere conjunction of two properties, for $(x,y) \in R$ is here everywhere equivalent to $x \in \mathrm{pr}_1 R$ and $y \in \mathrm{pr}_2 R$. (I shall use throughout the notation of Bourbaki.) Here there is not really a *relation* between E and F: only two properties ($\mathrm{pr}_1 R$ in E, and $\mathrm{pr}_2 R$ in F) that happen to be mentioned in the same sentence.

The case of 3 dimensions is more interesting and more suitable as a starting point. To dispose of the two extremes first: the subset R may be arbitrarily irregular, or it may be a product set. There is, however, an intermediate case. This occurs when

$$(x,y,z) \in R \text{ is equivalent to } \begin{cases} (x,y) \in \mathrm{pr}_{12} R \\ \text{and } (y,z) \in \mathrm{pr}_{23} R \\ \text{and } (x,z) \in \mathrm{pr}_{13} R\ . \end{cases}$$

The possibility is illustrated by the model in Fig. 1. It was easy to make, for just as a product set may be modelled in 3 dimensions by cuts with a hand-saw, parallel to the axes, so this set may be modelled by cuts with a band-saw, moving in any direction provided its edge stays parallel to the axes. The model may be helpful as illustration, but the reader should notice that the theory given in this article nowhere assumes any metric over the sets.

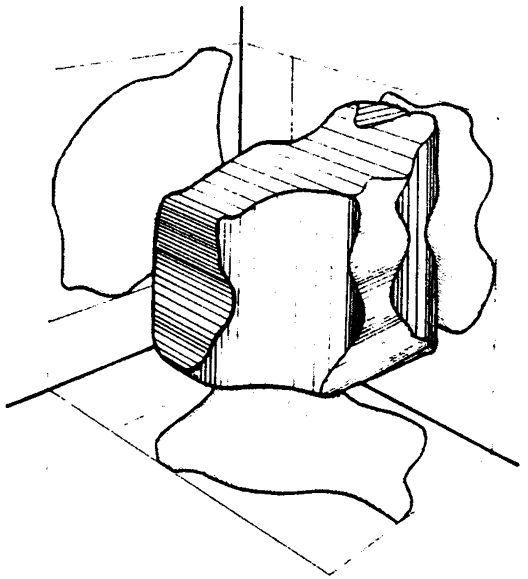


Fig. 1.

That this set has some constraint is intuitively evident. We will now proceed to treat the subject rigorously.

We start with a total set E of elements, the product of the set I of n sets E_i: $E = \prod_{i \in I} E_i$. The typical element is the n-tuple $(a_i)_{i \in I}$ composed of one element a_i from each set E_i. No order is assumed in the set I, nor any metric over the sets E_i. By $X - Y$ will be meant the set of elements that is in X but not in Y.

If J is a subset of I, by pr_J will be meant the mapping of $\prod_{i \in I} E_i$ into $\prod_{i \in J} E_i$ such that

$$\mathrm{pr}_J (a_i)_{i \in I} = (a_i)_{i \in J} \tag{1}$$

The following propositions will be required later; they are easily proved:

$$\text{If } J \subset K \subset I,\ (a_i)_{i \in K} \in \mathrm{pr}_K A \Rightarrow (a_i)_{i \in J} \in \mathrm{pr}_J A \tag{2}$$

$$A \subset B \Rightarrow \mathrm{pr}_J A \subset \mathrm{pr}_J B \tag{3}$$

$$\mathrm{pr}_J (A \cap B) \subset (\mathrm{pr}_J A) \cap (\mathrm{pr}_J B) \tag{4}$$

By the definition of an inverse, pr_J^{-1} acting on $(a_i)_{i \in J}$ will give all those elements in E that have $(a_i)_{i \in J}$ as their components in J; thus

$$\mathrm{pr}_J^{-1} (a_i)_{i \in J} = \{(a_i)_{i \in J}\} \times \prod_{i \in I - J} E_i \tag{5}$$

Similarly, if $Z \subset \prod_{i \in J} E_i$,

$$\mathrm{pr}_J^{-1} Z = Z \times \prod_{i \in I - J} E_i \tag{6}$$

If B is a subset of $\prod_{i \in J} E_i$, the set $B \times \prod_{i \in I - J} E_i$ is the 'cylindrical' set on B as 'base'. Thus, pr_J^{-1}, acting on Z, has the effect of forming the cylindrical set on Z as base.

The following propositions will be required later; they are easily proved:

$$A \subset B \Rightarrow \mathrm{pr}_J^{-1} A \subset \mathrm{pr}_J^{-1} B \tag{7}$$

$$\text{If } J \subset K \subset I, \quad \mathrm{pr}_K^{-1} \mathrm{pr}_K R \subset \mathrm{pr}_J^{-1} \mathrm{pr}_J R \tag{8}$$

We now reach the main working tool of the method. Given a set R, (a subset of E), its 'cylindrical closure of order p' is also a subset of E, formed by taking R's projections on all the subspaces of p dimensions, forming all cylinders in E on these projections as bases, and then taking the intersection of all the cylinders. Thus, if $p = 3$, R's cylindrical closure of order 3 is the set, in E:

References p. 18

$$\bigcap_{(ijk) \in I} \mathrm{pr}_{ijk}^{-1} \mathrm{pr}_{ijk} R,$$

where i, j, k runs through all triples in I. We will write it as $C_p R$. By definition, we will make $C_0 R = E$, and $C_n R = R$.

The most important property of these closures is that they form a nested set, only losing elements as they shrink down from E to R. A formal proof is advisable here.

Proposition: For any R in E, if $p \geq q$, then $C_p R \subset C_q R$.

Proof: The proposition is equivalent, in the same conditions, to

$$\cap\, \mathrm{pr}_P^{-1}\, \mathrm{pr}_P\, R \subset \cap\, \mathrm{pr}_Q^{-1}\, \mathrm{pr}_Q\, R,$$

where P runs through all subsets (of I) with p elements; and similarly for Q. Write the intersecting left-hand sets in a column on the left, and those of the right-hand sets on the right, arranging the items in each column so that each subset P is opposite a subset Q that is contained in it. (Residual blank spaces may be filled with selected repetitions of those already in.) Between any horizontal pair we now have the relation, by (8) above,

$$\text{left set} \subset \text{right set}.$$

So the intersection of the left column must be contained in the intersection of the right.

(*Corollary*: Every set is contained in all its cylindrical closures.)

Thus, given a total set E and some subset R of it, a sequence of sets — the 'cylindrical closures' — can be formed, the first of which equals E, the last of which equals R, and such that each is contained in the set preceding it in order. Each is associated with the number of dimensions necessary for its formation. Somewhere in the sequence the closure must first become equal to R; the set of smallest order equal to R gives the number that measures what we shall call R's 'cylindrance'. Thus the set in Fig. 1 has its constraint concisely specified by the fact that its cylindrance is 2.

We can now approach the matter from another point of view. Earlier we referred to the 'Law' as built up of many simple relations, and it seemed intuitively likely that the whole Law thereby was in some way restricted in complexity. We can in fact establish the matter rigorously with the following theorem.

Theorem: If a set R is the intersection of cylinders whose bases are all of m dimensions or fewer, then R's cylindrance cannot exceed m.

Proof: The difficulty is mostly notational; one way is as follows. From I, form all subsets having m elements. Call this new set M, and let μ be any

element in it. Let B_μ represent any base in $\prod_{i\in\mu} E_i$, where μ now represents the corresponding subset of I. By hypothesis, $R = \bigcap_{\mu\in M} \mathrm{pr}_\mu^{-1} B_\mu$. The theorem is proved, and R's cylindrance cannot exceed m, if and only if

$$\bigcap_{\mu\in M} \mathrm{pr}_\mu^{-1} \mathrm{pr}_\mu R = R.$$

Lemma: If $B_\mu \subset \prod_{i\in\mu} E_i$, and if $\nu\in M$,

$$\mathrm{pr}_\mu \left(\bigcap_{\nu\in M} \mathrm{pr}_\nu^{-1} B_\nu\right) \subset B_\mu .$$

By (4) above, the left side is contained in

$$\bigcap_{\nu\in M} \mathrm{pr}_\mu \mathrm{pr}_\nu^{-1} B_\nu .$$

Now

$$\mathrm{pr}_\mu \mathrm{pr}_\nu^{-1} B_\nu = \mathrm{pr}_\mu (B_\nu \times \prod_{j\in I-\nu} E_j).$$

The set in the parentheses is just a subset of E, and pr_μ, acting on it, will act according to whether each component in it is in μ or not. Thus,

(a) components not in μ are ignored completely;
(b) components in μ but not in ν will provide factors E_j;
(c) components in μ and in ν will act to form that projection.

More precisely, the μ-projection will be

$$(\mathrm{pr}_{\mu\cap\nu} B_\nu) \times \left(\prod_{j\in\mu-\nu} E_j\right).$$

The intersection of these sets, with ν taking all values in M, may be found as follows:

$$(a_i)_{i\in\mu} \in \bigcap_{\nu\in M} \mathrm{pr}_\mu \mathrm{pr}_\nu^{-1} B_\nu$$

$$\Leftrightarrow \quad \forall \nu \in M: \quad (a_i)_{i\in\mu} \in (\mathrm{pr}_{\mu\cap\nu} B_\nu) \times \left(\prod_{j\in\mu-\nu} E_j\right)$$

$$\Leftrightarrow \quad \forall \nu \in M: \quad (a_i)_{i\in\mu} \in \mathrm{pr}_{\mu\cap\nu} B_\nu \text{ and } (a_i)_{i\in\mu} \in \prod_{j\in\mu-\nu} E_j$$

$$\Leftrightarrow \quad \forall \nu \in M: \quad (a_i)_{i\in\mu} \in \mathrm{pr}_{\mu\cap\nu} B_\nu, \text{ (for the second part adds or excludes nothing).}$$

The equivalence is possible only if $\mu \cap \nu = \mu$, *i.e.* if $\nu \subset \mu$. As they have the same number of elements, $\nu = \mu$.

$$\Rightarrow (a_i)_{i\in\mu} \in \mathrm{pr}_\mu B_\mu \ (= B_\mu).$$

Thus the intersection must be contained in B_μ; which proves the Lemma.

If now both sides (of the expression in the Lemma) are operated on by pr_μ^{-1}

References p. 18

and then the intersection taken for all values of μ, we obtain

$$\bigcap_{\mu\in M} \mathrm{pr}_\mu^{-1} \mathrm{pr}_\mu \left(\bigcap_{\nu\in M} \mathrm{pr}_\nu^{-1} B_\nu\right) \subset \bigcap_{\nu\in M} \mathrm{pr}_\mu^{-1} B_\mu, \textit{ i.e. } \bigcap_{\mu\in M} \mathrm{pr}_\mu^{-1} \mathrm{pr}_\mu R \subset R$$

But R is always contained in its cylindrical closure, so the two sets must be equal. Thus, R's cylindrance is at most m, and the theorem is proved.

(The cylindrance may be less than m: the intersection of the arbitrarily given cylinders may be void. Equally one can see that as only the intersecting regions of the cylinders matter, the other regions may be of unlimited cylindrance without raising that of the intersection above m).

CONSTRAINT ANALYSIS

Given a set E, the 'constraint' introduced by a relation R is most naturally identified with the set $E - R$. When $R = E$ the constraint is zero; as R shrinks, so does the constraint become more intense. With the cylindrance as scale, we can compare relations otherwise incomparable. In particular, by forming the sequence of the cylindrical closures of a relation, and by seeing how fast, and at what stage, it shrinks, we can locate the relation's intrinsic complexities.

The amount by which it shrinks at each stage is shown by the size of the set $C_{m-1}R - C_mR$. By an analogy that will be developed later, it may be called the 'interaction of order m'. Call it D_m. Its properties reflect certain essential characteristics of R.

D_1, for instance, is $C_0R - C_1R$. Since C_1R is easily shown to be the smallest product set that contains R, D_1 is E minus this product set. It thus shows how much of R's constraint is due simply to the fact that R's variables have domains that do not use all that is offered by the sets E_i. This constraint D_1 is thus that due to the properties that R imposes on the variables individually.

D_2 shows the extent of R's constraint by binary relations. D_1's effect has been removed, so D_2 shows how much of R's constraint is due to the variables as unique pairs. D_3 shows how much of R's constraint is due to variables acting in triples, over and above the effects due to their actions in pairs, and so on.

The total constraint can thus be partitioned into sets of different degrees. As example, consider the case in which $n = 5$, $E_i = \{0, 1\}$ for all i, and R is the set of quintuples corresponding, in binary notation, to the seventeen denaries: 0, 1, 4, 5, 6, 8, 10, 11, 17, 19, 20, 22, 24, 26, 29, 30, 31. It is easily found that the cylindrical closures are:

$C_0R = E$ and has 32 elements
$C_1R = E$ and has 32 elements
$C_2R = E$ and has 32 elements
C_3R ..... has 28 elements
$C_4R = R$ and has 17 elements
$C_5R = R$ and has 17 elements

The relation is thus shown to be not of full complexity, for its cylindrance is 4; but it is of only slightly less than full complexity.

As contrast, consider the example, otherwise similar, in which R is the 16-element set represented by 2, 4, 5, 6, 10, 12, 13, 14, 18, 20, 21, 22, 26, 28, 29, 30. Its constraint analysis goes

$$C_0R = C_1R = E \text{ and has 32 elements}$$
$$C_2R = C_3R = C_4R = C_5R = R \text{ and has 16 elements.}$$

Thus R's cylindrance is 2, and R, for all its five dimensions, is shown to be really a collection of sub-relations, no one of which is more complex than a binary.

DISCUSSION

When the sets are small, as in the examples just given, the gain may seem to be hardly worth the trouble. The method may prove useful, however, when the number n of variables becomes so large that the combinatorial possibilities get quite out of hand. The method may then show that the actual relation, as it exists in some practical or experimental case, is in fact within the bounds of what is practical.

We have found that we can make rigorous our intuitive idea that 'The Law', being composed of Acts or Paragraphs no one of which relates more than k items, has an essential simplicity that is somehow related to k, no matter how many Acts may be added to The Law. In this way a Constraint Analysis is provided that may be usable in sociology.

In dynamics, the fact that the state-determined system, of m variables, can be described by m equations relating the m time-derivatives to the m variables:

$$\frac{dx_i}{dt} = f_i(x_1, \ldots, x_m) \qquad (i = 1, \ldots, m)$$

shows that the whole $2m$-ary relation of the total constraint has cylindrance not exceeding $m + 1$, for it is built from sub-relations, each one line of the equation, each of which relates m variables and 1 derivative.

Should the parts (or variables) of the system not be fully connected, then each f_i will have fewer than m arguments. If each f_i has only h arguments,

References p. 18

then the behavior of the whole must have the constraint corresponding to a cylindrance of $h + 1$. Since the 10^{10} nerve cells of the human brain are by no means fully connected each to each, the behavior of the living man must be subject to this quantitative and objectively demonstrable constraint. In this example the demonstration might be very difficult, but in simpler systems the demonstration might be found easier, especially as the nature of the analysis is well suited to processes of automatic computation.

The investigation described in this article was originally undertaken to provide a clear framework for studies of similar type, leading to a similar 'constraint analysis', with quantitative methods based on information theory, the idea being that the total entropy $H(x_1, \ldots, x_n)$ corresponds to the relation R, the maximal entropy $H(x_1) + \ldots + H(x_n)$ to the set E, and the total transmission (their difference) to the constraint $E - R$. The extension of these ideas in this direction cannot, however, be undertaken here.

ACKNOWLEDGEMENT

The work on which this article is based was supported by the Air Force Office of Scientific Research, Grant 7–63.

SUMMARY

Relations between large numbers of variables are often not as complex as they seem, for they are often constructed from simpler sub-relations, and retain something of their simplicity.

The idea is here treated rigorously, and a method is developed for detecting and measuring the degree of essential simplicity. The individual relation is made to generate a sequence of progressively simpler relations; where it comes in the sequence determines and measures its degree of essential simplicity.

The method may be useful when one wishes to consider relations (or systems) that, while involving very large numbers of variables, retain some simplicity derived from the sub-relations that formed them.

REFERENCES

BREMERMANN, H. J., (1962); Optimization through evolution and re-combination. *Self-Organizing Systems*. M.C. Yovits, G. T. Jacobi and G. D. Goldstein, Editors. Washington Spartan Books (pp. 93–106).

WIENER, N., (1914); A simplification of the logic of relations. *Proc. Cambridge phil. Soc.*, 17, 387–390.

INT. J. SYSTEMS SCI., 1972, VOL. 3, NO. 4, 343-356

THE IDENTIFICATION OF MANY-DIMENSIONAL RELATIONS*

R.F. MADDEN
and
W. ROSS ASHBY

[Received 11 October 1971]

Any system, no matter how general or complex, can be described by a relationship amongst nominal variables. In this paper we investigate the problem of identifying such a relationship, and the properties of the system corresponding to it, when the data descriptive of the system have been gathered from diverse and uncorrelated sources. We represent this problem theoretically as one of reconstructing and identifying an n-dimensional relation from its projections, and discuss the practical implications of this representation. We show that such reconstruction and identification is possible only if certain basic features of the corresponding system are constrained in a very definite manner. For example, if the relation is to be identifiable, i.e., if the set of relevant variables is to be "perfectly constrained", then the system must contain no more than a certain number of functionally determined variables, and no more than a certain number of independent variables. Furthermore, as the number of variables becomes large then, in theory at least, only a vanishingly small fraction of relations are identifiable. The treatment throughout is confined to nominal variables, and is intended to contribute to the constraint analysis of large and complex systems.

1. Introduction

Vast and diverse collections of data are often available to the investigator of large and complex systems such as the social, biological, or economic systems. These data may not be explicitly related in that they may have been collected neither to test specific hypotheses nor with any presumption of relevance to the objectives of the investigator. For example a sociologist (Lang and Lang 1961) may want to "reconstruct underlying patterns out of the partial and divergent perspectives of a multiplicity of observers" as recorded from newspaper

* This research was sponsored by the Air Force Office of Scientific Research under AFOSR Grant 70-1865.

and radio reports, television and film tapes, and in answers to questionnaires; or an economist (Morgenstern 1954) may want to design new experiments without specific guidance from theory, using data which are a mere by-product of administrative acts and government or business operations.

Such an investigator may not be immediately concerned with detailed statistical techniques, for example multiple regression or factor analysis, but may instead seek to distinguish in the data such basic features as consistencies and inconsistencies, or functional dependencies of some variables on others. For this purpose the data must be expressed in a form suitable for computer processing, a program must be composed to search for the above features, and, finally, the search must be carried out, probably in stages so that the investigator can intervene and interpret results after each stage. Even a partial execution of these tasks may prove to be immensely costly and difficult for large and complex systems. In such investigations there is a pressing need for what Tukey (1962) has called "procedures to extract indications rather than conclusions."

We shall be concerned in this paper with some general system properties and the indications which they can provide concerning the above mentioned basic features of a system.

We consider the case in which the system under investigation is defined by a set of n variables, representing those selected by the investigator, and the overall relationship holding between these variables. The overall relationship between the variables is yet unknown to the investigator to whom only the collections of data are available. Each collection of data is assumed to define a relationship among a subset of at most say, p, variables. By combining these relationships the investigator can attempt to reconstruct the overall n-variable relationship and thus determine all the properties and characteristics of the system.

If the data are not contradictory, such reconstruction is possible in principle. It may result in the identification of the overall relationship, and thus in exhaustive knowledge of the system, or, alternatively, it may result in an approximation to the overall relationship. The relationship describing the system can then be either identified or approximated to by combining p-variable relationships. The possibility of such identification or approximation, for a given p and n, corresponds to certain constraints on the basic features of a system.

To determine these constraints we proceed as follows: A mathematical representation of the problem is obtained by making one fundamental assumption concerning the available data. We then consider the reconstruction and identification of the original relationship. Finally, we examine the system properties implied by the possibility of such reconstruction and identification.

2. The data

In investigations such as those indicated above many a variable may have no natural mathematical structure, and different mathematical structures may be appropriate to different viarables. For example, a ratio scale is appropriate to the variable "density" and to the variable "resistance", an interval scale is appropriate to "temperature" and "energy" (Stevens 1951), and such variables as "cloud formation", "weather condition", or "type of disease" have no natural mathematical structure. A variable without mathematical structure we call a "nominal" variable. This paper concerns nominal variables, and consequently all results are applicable to a system which has variables of different types.

The variables under investigation are denoted by $X_1, X_2, \ldots, X_n$ and the variable X_i takes its value from the set E_i, $i = 1, 2, \ldots, n$; where any set E_i is called a "dimension". The total space under investigation will thus be the cartesian product E of the sets $E_1, E_2, \ldots, E_n$. If a collection, c, of data defines a relationship among the variables $X_i, X_j, \ldots, X_k$, say, then the relationship defined by c can be represented (Ashby 1964) as a subset R_c of the product set $E_i \times E_j \times \ldots \times E_k$. All of the data available to the investigator will therefore be represented by a finite collection, D, of such subsets. The overall relationship defining the system under investigation will be represented by the subset S of E.

Our fundamental assumption concerning the collections of data and the relations in D is as follows: each collection of data is assumed to define completely the values adopted by the set of variables to which it refers. Consequently, each relation in D can be assumed to be a projection from the subset S; thus R_c for example would be the projection of S into $X_i \times X_j \times \ldots \times X_k$. This amounts to representing the "partial perspective" of each of a multiplicity of observers in terms of a product set of variables, and the values adopted by those variables. With this assumption the relations in D need only be combined by intersection to obtain a reconstruction of S.

When dealing with actual data some preliminary organization will now be required in an effort to make the corresponding relations qualify as projections from a single set S. For example, the collections of data must be so chosen that if a variable X_i is relevant to two different collections A and B, then every value of X_i which occurs in A must also occur in B. If this cannot be arranged then, in our scheme, the variable X_i must be redefined or temporarily dropped from consideration. This assumption also means that a variable having relative frequencies as its values cannot easily be taken into account; consequently, we will ignore the shape of any frequency distributions associated with S.

As an illustration of some of the above notions let us suppose that an

investigation concerns the effects of an industrial environment on leguminous plant life. This may involve many hundreds of such disparate variables as "indigenous fauna" or "sulphur production". Suppose there are available some items of data resulting from a previous study of, say, the migration or non-migration (X_3) of an avian species (X_m) from rural or urban areas (X_2) under various weather conditions (X_1). The dimensions E_1, E_2, E_3, E_m; corresponding to the variables X_1, X_2, X_3, X_m, respectively, might be as follows:

'weather': $E_1 = \{x_{11}, x_{12}, x_{13}\}$; x_{11} = clear, x_{12} = precipitation, x_{13} = cold,

'location': $E_2 = \{x_{21}, x_{22}\}$; x_{21} = urban, x_{22} = rural,

'behaviour': $E_3 = \{x_{31}, x_{32}\}$; x_{31} = migration, x_{32} = no migration,

'species': $E_m = \{x_{m1}, x_{m2}\}$.

Two items of data concerning species x_{m1} and x_{m2} might be as represented in tabular form in Fig. 1. Both of these items must be assigned to one collection. If this collection contained no other items it would define a relation such as that in Fig. 2

Fig. 1

ITEM 1 $X_m = x_{m1}$				ITEM 2 $X_m = x_{m2}$			
X_1	X_2	X_3	f	X_1	X_2	X_3	f
x_{11}	x_{21}	x_{31}	0	x_{11}	x_{21}	x_{31}	0
x_{11}	x_{21}	x_{32}	0·2	x_{11}	x_{21}	x_{32}	0·3
x_{11}	x_{22}	x_{31}	0·1	x_{11}	x_{22}	x_{31}	0·1
x_{11}	x_{22}	x_{32}	0	x_{11}	x_{22}	x_{32}	0
x_{12}	x_{21}	x_{31}	0·1	x_{12}	x_{21}	x_{31}	0
x_{12}	x_{21}	x_{32}	0·2	x_{12}	x_{21}	x_{32}	0·1
x_{12}	x_{22}	x_{31}	0·1	x_{12}	x_{22}	x_{31}	0·1
x_{12}	x_{22}	x_{32}	0·1	x_{12}	x_{22}	x_{32}	0
x_{13}	x_{21}	x_{31}	0	x_{13}	x_{21}	x_{31}	0·1
x_{13}	x_{21}	x_{32}	0	x_{13}	x_{21}	x_{32}	0
x_{13}	x_{22}	x_{31}	0·2	x_{13}	x_{22}	x_{31}	0
x_{13}	x_{22}	x_{32}	0	x_{13}	x_{22}	x_{32}	0·3
			1·0				1·0

Example of two items of data.

Fig. 2

X_m	X_1	X_2	X_3
x_{m1}	x_{11}	x_{21}	x_{32}
x_{m1}	x_{11}	x_{22}	x_{31}
x_{m1}	x_{12}	x_{21}	x_{31}
x_{m1}	x_{12}	x_{21}	x_{32}
x_{m1}	x_{12}	x_{22}	x_{31}
x_{m1}	x_{12}	x_{22}	x_{32}
x_{m1}	x_{13}	x_{22}	x_{31}
x_{m2}	x_{11}	x_{21}	x_{32}
x_{m2}	x_{11}	x_{22}	x_{31}
x_{m2}	x_{12}	x_{21}	x_{32}
x_{m2}	x_{12}	x_{22}	x_{31}
x_{m2}	x_{13}	x_{21}	x_{31}
x_{m2}	x_{13}	x_{22}	x_{32}

Relation corresponding to data items in Fig. 1.

Mathematical expressions must be represented as a subset of some appropriate product set. For instance, if we are interested in variables X_1 and X_2, and no boundary condition is specified for a previously known relationship expressed as $dX_1/dX_2 = k$, then both this expression, and the expression $X_1 = kX_2 + X_3$ represent the same four-dimensional subset of the product set $E_1 \times E_2 \times E_3 \times (k)$ where E_3 is a set of boundary conditions.

Following the above assumption we are now concerned with the problem of identifying, or approximating to, an n-dimensional relation S from a set D of relations obtained by projection from S.

3. Previous work

In 1964 Ashby (1964) considered how the sequence of its cylindrical closures approximates to an n-dimensional relation, where a relation was defined, following Wiener (1914), to be the set of n-tuples that satisfy it. The ρth cylindrical closure of a relation will be defined theoretically in section 5. Its relevance to our concern is as follows. If a physical process, or a set of activities or behavior patterns is observable as the set of values adopted by a set of n variables, then an investigator can reconstruct from such ρ-dimensional data only the ρth cylindrical closure of the original n-dimensional relation. The approximation referred to in the preceding paragraph is therefore an approximation by cylindrical closures.

If a relation can be only approximated to by its ρth cylindrical closures then the set of n variables is under-constrained by ρ-dimensional relations. We

are, therefore, concerned with problems of under-constraint. A series of papers by Friedman and Leondes in 1969 (1969 a, b) treated in a fundamental way of the well-posed or constraint problem. The emphasis in these papers was on the problem of variables being over-constrained by sets of relations. What follows is, for the most part, complementary to, and is in some cases a development of, the approach of Friedman and Leondes. We do, however, confine our attention to nominal variables and will not be concerned with different representations of relations.

4. Notations and conventions

Variables and dimensions will usually be referred to by their indices. Thus, instead of 'variable X_i' or 'dimension E_j' we will write 'variable i' or dimension j'. The set of all variables or dimensions under investigation will be denoted by I, $I = \{1, 2, 3, \ldots, n\}$ The cardinality of any set A will be denoted by $\#(A)$. Thus $\#(\phi) = 0$ and $\#(\{1, 2, 3\}) = 3$. We assume that for every $i \in I$, $\#(E_i) \geqslant 2$.

If all the elements in a set A are contained in a set B we write $A \subset B$; and if also some elements of B are not contained in A we denote this by $A \subsetneq B$. We denote the complement of a subset by a superscript c when the set containing it is understood; thus the set $E - S$ is denoted by S^c and $I - K$ by K^c. If $K \subset I$ the cartesian product of the variables K and the dimensions K will be denoted by X_K and E_K, respectively, so that $E_K = \prod_{i \in K} E_i$. If $A \subset E_J$ and $B \subset E_K$, by the 'intersection' of A and B, $A \cap B$, we mean the subset $(A \times E_{K \cap J^c}) \cap (B \times E_{J \cap K^c})$.

If $\#(K) = p$ and $B \subset E_{Kc}$ then the subset $B \times E_K$ will be called a 'p-dimensional cylinder', or a 'cylinder having p dimensions', on the '$(n-p)$-dimensional base' B.

We use the following two basic set operators:

(1) The projection operator: pr_K.

If $A = pr_K B$ then A consists of all those elements of E_K which appear in B.

(2) The spreading operator: V_J.

If $A = V_J B$, and if $B \subset E_K$ then $A = (pr_{K-J} B) \times E_J$, where $J \subset K$, and $A \subset E_K$. Thus V_J "spreads" each element of B along the dimensions J.

5. Reconstructing and identifying S

The cylindrical closure of order p, of the set S, denoted by $C_p S$, is formed by projecting S onto all p-dimensional subspaces of E, and intersecting all cylinders into E having these projections as bases (Ashby 1964).

Definition 1

$$\text{For } p = 1, 2, 3, \ldots, n-1: \quad \text{and any } S \subset E:$$
$$C_p S = (\bigcap_J V_J S : J \subset I \quad \text{and} \quad \#(J) = n - p).$$

We define $C_0 S = E$ and $C_n S = S$.

The set $(C_p S: p = 1, 2, 3, \ldots, n)$ forms a nested set of "approximations" to S as follows:

Proposition 1 (Ashby 1964)

$$E = C_0 S \supset C_1 S \supset C_2 S \supset \ldots \supset C_{n-1} S \supset C_n S = S.$$

The set $C_p S$, therefore, contracts from E to S as p increases. One quantity which measures the sharpness of this contraction is the minimum value of p for which $C_p S = S$. This quantity has been termed the cylindrance of S (Ashby 1964), and is denoted in the following by "cyl S".

Definition 2

$$\text{cyl } S = \min (p : C_p S = S).$$

If S describes a physical process, or a set of activities, and the values adopted by the variables $X_1, X_2, X_3, \ldots, X_n$ are recorded by a sufficiently numerous set of observers, each of whom records the values adopted by p of the variables, then S can be reconstructed from these records only on the condition that cyl $S \leqslant p$.

If cyl $S > p$ the relation reconstructed from the data, S_{rec} will, according to Proposition 1, contain some n-tuples which are not contained in S. The variables I are then under-constrained by p-dimensional relations.

If S is such that $C_p S = E$, then, since each subset of p variables will adopt all possible values, the data provided by our hypothetical group of observers give no information whatever concerning the structure of S. The maximum value of p for which this is the case will be called the 'dimensional scope' of S, denoted by sep S.

Definition 3

$$\text{sep } S = \max_p (p : C_p S = E).$$

Any relation whose dimensional scope exceeds zero is a 'universal relation' in the terminology of Friedman and Leondes (1969 a, b).

A 'relevant variable' (Friedman and Leondes 1969 a, Definition 5) can be defined as follows:

Definition 4

Variable X_i is "relevant" to relation R if $V_i R \neq R$.

If scp $S \geqslant p$ then any record of the values adopted by p, or fewer, of the variables $X_1, X_2, X_3, \ldots, X_n$ will define a relation having no relevant variables and will therefore be totally uninformative. In practice, if no significant statistical linkages exist between sets of less than p variables, the existence of a constraint such as S could not even be detected except from records of the values adopted by sets of at least $(p + 1)$ variables. This is illustrated by Fig. 3

Fig. 3

S: X_1	X_2	X_3	X_4
x_{11}	x_{21}	x_{31}	x_{41}
x_{11}	x_{21}	x_{32}	x_{42}
x_{11}	x_{22}	x_{31}	x_{42}
x_{12}	x_{21}	x_{31}	x_{42}
x_{11}	x_{22}	x_{32}	x_{41}
x_{12}	x_{21}	x_{32}	x_{41}
x_{12}	x_{22}	x_{31}	x_{41}
x_{12}	x_{22}	x_{32}	x_{42}

$E_1 = \{x_{11}, x_{12}\}$ $E_2 = \{x_{21}, x_{22}\}$
$E_3 = \{x_{31}, x_{32}\}$ $E_4 = \{x_{41}, x_{42}\}$
scp $S = 3$
cyl $S = 4$

Each variable is a function of the remaining three variables.

where any triple of variables appears totally unconstrained. In experimental situations events can usually be clocked or timed. The dimensions of interest are then, say, $E_1, E_2, E_3, \ldots, E_n$ and t. If the time divisions are small enough to enable a unique time value to be associated with every n-tuple in S then this has the effect of producing a composite relation S'; where $S = pr_1 S'$ and $S' \subset E \times t$. For this composite relation cyl $S' = 2$, and scp $S' = 0$, irrespective of the values of cyl S and scp S. The relation S' can therefore be correctly reconstructed from two-dimensional relations, at least if one of the two dimensions in each relation is t. In reconstructing past events using data from divergent and uncorrelated sources such convenient timing information will seldom be available.

We have introduced so far the quantities cyl S and sep S. In dealing with the three sets S, D, and S_{rec}, we have seen that if the relations in D have at least p relevant variables then $C_p S \supset S_{rec} \supset S$; and if cyl $S \leqslant p$ then $S = S_{rec}$. But cyl S is unknown to the investigator, so to identify S it is necessary to construct S_{rec}. If S_{rec} is the only relation which projects precisely onto the elements of

D then $S = S_{rec}$ and S are thus identified. If S is not such a relation, then even if S equals S_{rec} the investigator has no way of knowing this to be the case. If S can be identified from p-dimensional relations we say that S is p-identifiable, and write ident $S \leqslant p$. Ident S is defined formally as follows:

Definition 5

Ident $S \leqslant p$ iff $R \neq S$ implies that for some K $\#(K) = p$, $pr_K R \neq pr_K S$.

If ident $S \leqslant p$ and ident $S > p - 1$ we say that ident $S = p$. If ident $S > n - 1$ then exhaustive knowledge of S cannot be obtained from partial or incomplete perspectives and we say that S is not identifiable.

The following two propositions link ident S with the quantities of cyl S and sep S:

Proposition 2

For any $S \subset E$ the following conditions are equivalent: (i) ident $S \leqslant p$; (ii) no subset of S has cylindrance greater than p.

Proof

(ii) (i). Let $R \neq S$ and $pr_K S$ for any K such that $\#(K) = p$. If cyl $S \leqslant p$ then $R \subset S$ and cyl $R > p$. Therefore some subset of S has cylindrance greater than p.

(i) (ii). Let $A \subset S$ and cyl $A > p$. Let $R = (S \cap C_p A) \cup A$. Then $R \neq S$, and if $\#(K) = p$ we have $pr_K R = pr_K S$. Therefore ident $S > p$.

Proposition 3

For any $S \subset E$, if ident $S \leqslant p$, then sep $S^c \geqslant n - p$.

Proof

If sep $S^c < n - p$ it can be shown that for some K, $\#(K) = p + 1$, there exists an element e of E such that $V_K e \subset S$. Let $R = (V_K e) - e$. Then $R \subset S$ and cyl $R > p$.

These propositions show that if ident S is small then S is very different from its complement.

We have now obtained conditions under which an n-dimensional relation can be identified from its p-dimensional projections. The remainder of this paper will be devoted to exploring the implications of p-identifiability.

We will examine first of all how ident S is related to functional dependencies amongst the variables $X_1, X_2, X_3, \ldots, X_n$.

6. Functional dependencies

By considering a "system" to be a subset S of E_I, where I is any set of variables, we have made the "system under investigation" totally dependent on the set of variables chosen for investigation. However, in order for S to be identifiable it is necessary that some of the variables in I "interact" with others. The following proposition shows that a mere juxtaposition of independent variables can never lead to an identifiable system.

Proposition 4

The number of irrelevant variables in I cannot exceed (ident S) and cannot exceed (n − cyl S).

Proof

If ident $S = p$, then it can be shown that for any e S, if $\#(K) > p$, then $(V_K e)$ $S^c \neq \phi$. Hence $V_K S \neq S$, so I contains at most p irrelevant variables. If cyl $S = q$, then for $\#(K) > (n - q)$, we can show that for some a S^c, $(V_K a)$ $S \neq \phi$. Hence $V_K S \neq S$, so I contains at most $(n - q)$ irrelevant variables.

If S_1 and S_2 describe independent systems, $I_1 \cap I_2 = \phi$; $S_1 \subset E_{I1}$; $S_2 \subset E_{I2}$; then the values adopted by $I_1 \cup I_2$, consistituting the relation $S_1 \times S_2$, cannot be identified except from relations having (ident S_1) + (ident S_2) dimensions. But if variables I_2, say, are dropped temporarily from consideration an identifiable reconstruction of S_1 may be more readily obtainable.

In general, dropping variables from consideration can reduce the changes of obtaining an identifiable reconstruction. For example, if S is in fact p-identifiable, $S \subset E_I$, and variables K are ignored, $K \subset I$, we are then dealing with $pr_{I-K} S$ which may not be identifiable.

Although some of the variables in I must interact with others if S is to be identifiable, the extent to which such interactions can be functional in character is limited by ident S. We say that a variable X_i is "functionally determined" by the relation S if there is some $K, K \subset I - i$, such that for any value of the variables $I - (K \cup i)$ which occurs in S, variable X_i is some function of variables X_K; i.e. $X_i = f_i(X_K)$. In the following proposition we allow the function f_i to be different for different values of the variables $I - (K \cup i)$.

Proposition 5

The number of variables functionally determined by a relation S cannot exceed ident S.

Proof

Let $X_i = f_i(X_{K_i})$ where $i \in P$, $\#(P) = p$, $i \notin K_i$, $K_i \subset I$. Let $K = P^c \cap (\bigcup_{i \in P} K_i)^c$

If there is some $a \in E - S$ such that $pr_K = pr_K b$ for some $b \in S$, then for $i \in P$, by definition of f_i, we have $(V_i a) \cap S \neq \phi$. Therefore, $(V_L a) \subset E - S$ implies that $\#(L) \leqslant n - p$; and from this it can be shown that cyl $S \geqslant p$. If there is no such $a \in E - S$, then sep $S^c < \#(K) \leqslant n - p$ so that $p <$ ident S. Hence $p \leqslant$ ident S.

Corollary

If $X_i = f_i(X_{I-i})$ for $i = 1, 2, 3, \ldots, n$; then S is not identifiable.

If the function f_i is not different for different values of the variables $I - (K \cup i)$, but is instead independent of these variables, then it is easily shown that the number of such independent variables is again restricted by ident S.

Proposition 6

If $X_i = f_i(X_K)$ and f_i is independent of the variables $(X_j : j \notin K \cup I)$ then ident $S > n - 1 - \#(K)$.

It is possible that some variables, in investigations such as we are considering, belong to state determined systems, i.e. systems in which the value of the variables at the end of some time interval is functionally determined by their value at the beginning of that interval. Any change in cylindrance over a time interval can tell us something about the number of such variables.

Suppose each n-tuple in S_1 represents the value adopted by the variables I at some time t_j during the set of time instants $(t_j : j \quad \alpha)$, and S_2 represents the values adopted by I during the set of time instants $(t_j + \Delta t : j \quad \alpha)$; i.e. $S_1 = \bigcup_{j \in X} X_I(t = t_j)$ and $S_2 = \bigcup_{j \in X} X_I(t = t_j + \Delta t$.

If the set K of variables belongs to a state-determined system, then for some function, g, $X_K(t_j + \Delta t) = g(X_K(t_j))$. If we suppose that the variables K^c change value only during the intervals $(t_j + \Delta t, t_{j+1})$, and assuming for convenience that g is one-to-one, we get the following result:

Proposition 7

$$(\text{cyl } S_1) - \#(K) < \text{cyl } S_2 < (\text{cyl } S_1) + \#(K).$$

Proof

Noting that $S_2 = \bigcup \{a \times g(S_1(a)) : a \in pr_{E-K} S_1\}$ where $S_1(a)$ is the "section" of S_1 at a (Ashby 1964), a proof follows straightforwardly from the

fact that there is some $e \in S_1{}^c$, such that for any $J \subset I$, if $\#(J) = (\text{cyl } S_1) - 1$, then $S_1 \cap (V_{E-J}e) \neq \phi$.

Thus, if cyl S changes, the number of variables involved in state determined behavior must exceed the change in cyl S.

7. Consistency and inconsistency

When dealing with several identified relations $S_1, S_2, S_3, \ldots$; among the variables $I_1, I_2, I_3, \ldots$; the question of their consistency or inconsistency arises. Following Friedman and Leondes (1969 a, p. 53) we say that a pair S_1, S_2 ; of relations is consistent if the intersection $S_1 \cap S_2$ is non-empty, otherwise S_1 and S_2 are inconsistent.

Definition 6

S_1 is consistent with S_2 iff $S_1 \cap S_2 \neq \phi$. In investigating systems such as those indicated in section 1 it is often important to know whether or not some relations are consistent with the complements of others. For example, if S_1 represents values of ecological variables essential to the survival of a species, and S_2 represents a set of ecological conditions, then if S_1 is consistent with $S_2{}^c$ the species cannot survive under conditions S_2. The following proposition shows how any inequality in dimensional scope indicates a consistent pair of relations.

Proposition 8

For any pair S_1, S_2 ; of relations, if either scp $S_1 >$ scp $S_2{}^c$, or scp $S_2 >$ scp $S_1{}^c$, then S_1 is consistent with S_2.

Proof

If S_1 and S_2 are inconsistent, then, if $S_1 \subset E_J$ and $S_2 \subset E_K$; $S_1 \times E_J{}^c \subset S_2{}^c \times E_K{}^c$, so that scp $S_1 \leqslant$ scp $S_2{}^c$. Similarly scp $S_2 \leqslant$ scp $S_1{}^c$.

Corollary

If ident $S_1 \leqslant n/2$ and ident $S_2 \leqslant n/2$ then $S_1{}^c$ is consistent with $S_2{}^c$.

The result expressed in the table is obtained by complementing in all possible ways the relations S_1 and S_2 in the preceding proposition. In this table if a condition in the left-hand column is satisfied, then the pair of relations in the box horizontally opposite is consistent.

Taking the variables I_1 and I_2 into account we get the following two propositions, the first of which is a generalization of the Friedman and Leondes Theorem 5:

Condition	Consistent pair
scp $R_i <$ scp R_j	$R_i{}^c$, R_j
scp $R_i <$ scp $R_j{}^c$	$R_i{}^c$, $R_j{}^c$
scp $R_i{}^c <$ scp R_j	R_i, R_j
scp $R_i{}^c <$ scp $R_j{}^c$	R_i, $R_j{}^c$
scp $R_j <$ scp R_i	R_i, $R_j{}^c$
scp $R_j <$ scp $R_i{}^c$	$R_i{}^c$, $R_j{}^c$
scp $R_j{}^c <$ scp R_i	R_i, R_j
scp $R_j{}^c <$ scp $R_i{}^c$	$R_i{}^c$, R_j

Proposition 9

If S_1 has no more than k variables in common with S_2, and scp $S_2 \geqslant k$, then S_1 is consistent with S_2, and $\#(S_1 \cap S_2) \geqslant \#S_1)$.

Corollary

If S_1 and S_2 have no more than $\#(I_1 \cup I_2) -$ (ident S_1 + ident S_2) variables in common, then $S_1{}^c$ is consistent with $S_2{}^c$.

Proposition 10

If scp $S_1 \geqslant k$, and scp $S_2 \geqslant k$, and $\#(I_1 \cap I_2) \leqslant k/2$, then scp $(S_1 \cap S_2) \geqslant k$.

The consistency or inconsistency of small numbers of relations can often be ascertained on inspection using the preceding propositions. No such immediate results can be expected for large numbers of relations unless other properties of the relations are taken into account. For example, from Proposition 3, and Theorem 7 (Friedman and Leondes 1969 a), or from Proposition 9 above, we get the following:

Proposition 11

If the model graph of a set of identifiable relations is a tree, then the complements of all of these relations are consistent.

The "local dimension", $I(S)$, of a set S has been used by Friedman and

Leondes (1969 b) in examining the consistency of large numbers of "regular" relations. By writing ident $S = \max_{A \subset S}$ (cyl A) we see that ident S is a sort of local dimension suitable for nominal variables. If S is a discrete relation, for instance, then ident S, unlike $I(S)$, need not necessarily be zero. A fundamental distinction between ident S and $I(S)$ is that ident S is always given with respect to the sets $E_1, E_2, E_3, \ldots, E_n$ so that ident $E_i = 1$, ident $E_i \times E_j = 2$, etc.; while $I(E_i)$ depends on the structure of E_i. If the definition of local dimension were expressed as a subset of a product set then ident S and $I(S)$ could be examined from the same viewpoint. Assuming, however, that $I(E_i) \leqslant 1$, $i = 1, 2, \ldots, n$; it is easy to show, from $I(A \times B) \leqslant I(A) + I(B)$, that $I(S) \leqslant$ ident S. If ident $S = p$, we can think of S as a "surface", of at most $I(S)$ dimensions, twisted into at most p dimensions throughout n-space.

If we re-define "regularity" (Friedman and Leondes 1969 b, Definition 31) in terms of ident S rather than $I(S)$, and call it N-regularity, then those of the Friedman and Leondes model-graph techniques which depend on the "regularity hypothesis" can be modified for use with nominal variables. These modified techniques, however, can useful be applied only to small numbers of relations. One such modification demonstrates why this is so, and also indicates how some "irregular" sets of R_2-type relations might be detected before applying the "regularity hypothesis" (Friedman and Leondes 1969 b): If S_1 and S_2 have no relevant variables in common, then ident $(S_1 \cap S_2)$ = (ident S_1) + (ident S_2) so that no set containing both S_1 and S_2 can be N-regular. For example, no nodal tree containing more than two nodes can be N-regular. Similarly for R_2-type relations in spaces for which $I(A \times B) = I(A) + I(B)$, any set containing two relations which have no relevant variables in common must be "irregular".

We have seen that non-identifiable relations are those which cannot be exactly determined from partial or incomplete perspectices. The following proposition demonstrates that for large n, and at least for $\#(E_i) \geqslant 16$, almost all n-variable relations are non-identifiable.

Let $W_n(r)$ be the fraction of relations, on n variables, having cylindrance less than n, where each variable takes r values, i.e. $\#(E_i) = r$, $i = 1, 2, 3, \ldots, n$.

Proposition 12

For some integer k, $2 < k \leqslant 16$, and for all $r \geqslant k$, $\lim_{n \to \infty} W_n(r) = 0$.

Proof

This can be proved by showing that if $\#(E_i) = 2^m$, m $\geqslant 4$, the the fraction of relations in E_I having cylindrance less than n tends to zero with increasing n. Furthermore $\lim_{n \to \infty} W_n(r+1) \leqslant \lim_{n \to \infty} W_n(r)$, which can be shown as follows:

If $G = G_I$; and $\#(G_i) = r + 1$, with each subset S of E can be associated a unique collection $Q(S)$ of $2^{(r+1)^n - r^n}$ subsets of G such that the cylindrance of each element of $Q(S)$ is not less than cyl S; thus $W_n(r+1) \leqslant W_n(r)$.

Consequently, since ident $S \geqslant$ cyl S, only a vanishingly small fraction of relations are identifiable if $r \geqslant 16$. The exact value of the integer k referred to in this proposition is not known. For the case in which $r = 2$ we have the following:

Proposition 13

$$\frac{29}{48} \leqslant \lim_{n \to \infty} W_n(2) \leqslant \frac{30}{48}.$$

Proof

For any $e \in E$, let $Q_e = \bigcup_{i \in I} V_i e)$. We call the set $\{Q_{e1}, Q_{e2}, \ldots, Q_{ek}\}$ a k-configuration, and if for some set $\{e_1, e_2, \ldots, e_{ek}\}$ we have $\{e_1 \cup e_2 \cup \ldots \cup e_k\} \subset S^c$ and $\{Q_{e1}, Q_{e2} \cup \ldots \cup Q_{ek}\} \subset S$, we say that S "contains" the k-configuration $\{Q_{e1}, Q_{e2}, \ldots, Q_{ek}\}$. $W_n(2)$ is the fraction of relations which contain no k-configuration for any k.

Let $N_n(\{Q_{e1}, Q_{e2}, \ldots, Q_{ek}\})$ denote the number of n-variable relations "containing" the k configuration $\{Q_{e1}, Q_{e2}, \ldots, Q_{ek}\}$; and let $N_n(k)$ denote the sum of $N_n(\{Q_{e1}, Q_{e2}, \ldots, Q_{ek}\})$ over all k-configurations. Let $N_n(0) = 2^{(2n)}$.

Then

$$W_n(2) = \sum_{k=0}^{r} (-1)^k \frac{N_n(k)}{N_n(0)}$$

where $r < 2^n$.

Thus

$$\sum_{k=0}^{\ddot{r}} (-1)^k \frac{N_n(k)}{N_n(0)} < W_n(2) < \sum_{k=0}^{\ddot{2}} (-1)^k \frac{N_n(k)}{N_n(k)}$$

where $N_n(k) \neq 0$ for $k > 3$.

Finding $N_n^L(i)$ and $N_n^U(i)$ such that $N_n^L(i) \leqslant N_n(i) \leqslant N_n^U(i)$, since

$$1 - \frac{N_n(1)}{N_n(0)} + \frac{N_n^L(2)}{N_n(0)} - \frac{N_n^U(3)}{N_n(0)} < W_n(2) < 1 - \frac{N_n(1)}{N_n(0)} + \frac{N_n^U(2)}{N_n(0)},$$

gives

$$1-\tfrac{1}{2}+\frac{1}{2\cdot 2!}-\frac{1}{3\cdot 2!}\leqslant \lim_{n\to\infty} W_n(2)\leqslant 1-\tfrac{1}{2}+\frac{1}{2\cdot 2!};$$

which suggests that $\lim_{n\to\infty} W_n(2)$ may equal $\frac{1}{\sqrt{e}}$.

A relation, therefore, which is chosen at random from the set of all possible relations on a large number of variables will almost certainly not be identifiable.

References

ASHBY, W.R., 1964, *General Systems, Yearbook of the Society for General Systems Research*, vol. 9 (Ann Arbor, Michigan), pp. 83, 99.

FRIEDMAN, G.J., and LEONDES, C.T., 1969 a, *I.E.E.E. Trans. Systems Sci. Cybernetics*, 5, 58; 1969 b, *Ibid*, 5, 132, 191.

LANG, K., and LANG, G.E., 1961. *Collective Dynamics* (New York: Thomas Crowell), p. 553

MORGENSTERN, O., 1954, *Economic Activity Analysis* (New York: John Wiley & Sons), p. 539.

STEVENS, S.S., 1951, *Handbook of Experimental Psychology* (New York: John Wiley & Sons), p. 21.

TUKEY, J.W., 1962, *Ann. math. Statist.*, 33, 1.

WIENER, N., 1914, *Proc. Camb. phil. Soc. math. phys. Sci.*, 17, 387.

VI.

BRAINS, INTELLIGENCE, CREATIVITY, AND GENIUS

BRAINS, INTELLIGENCE, CREATIVITY, AND GENIUS

INTRODUCTION

In our increasingly complex and interconnected world one sometimes hears wistful expressions to the effect that what we need is some superintelligent leader who can cope with it all on our behalf, since our own intelligence seems so inadequate. As Ashby points out in "Design for an Intelligence Amplifier," this is a forlorn hope, since human intelligence is severely limited. In a fruitful analogy he points out that mankind's big advance in mechanical strength did not come from breeding burlier slaves but rather when Watt discovered a way to tap natural sources of power. The steam engine is a power-amplifier in the sense that a small amount of power (used to manipulate the controls, fuel the fire, etc.) is used to control a much larger amount (liberated from the fuel, ultimately). For intelligence-amplification, then, we need some way to use human intelligence, which Ashby argues is basically selection, to control much larger sources of selection.

In the field of artificial intelligence, which was in its infancy when the article was written, the basic process outlined by Ashby is used very commonly. Always the goal of intelligent action, by human or machine, can be viewed abstractly as an appropriate selection from a set: a chess move from all moves legally possible; the answer to a test question from the set of all available answers; a sequence of actions which will accomplish an objective, taken from the set of possible actions, and so on. To generate a set of moves, answers, or actions is rather easy and no sign of great intelligence; to select the right one(s) from the set is the hard part. When we write (or abstractly, select) a computer program to do the selection for us we are obtaining selection-amplification. To the extent that the essence of intelligence is selection, we are obtaining intelligence-amplification, and there is no problem in devising a machine which exhibits more intelligence than its designer.

In this farsighted and penetrating article Ashby also discusses solutions to some of the problematic combinatorial aspects of "selection by equilibrium." All of these are reflected in modern efforts in artificial intelligence.

"Can a Mechanical Chess-Player Outplay its Designer?" From the arguments in the first article in the chapter one can conclude that the answer is 'yes,' and it is also an historical fact that computers do outplay their programmers. (Ashby would have been delighted to hear that as this book was being edited, a computer program defeated the world's champion of backgammon, a substantially difficult intellectual feat.) The question addressed by the paper, however, is broader: can a machine exhibit more "design" than is put into it by its designer? The famous "Argument from Design" for the existence of God is based on the allegedly "self-evident truth" that it could not, so the question is of philosophical interest. Ashby casts the philosophical question in modern cybernetic terms and shows that if the designer contrives the machine in such a way that it can use information from its own experience, then it can eventually show "more design" than that put in by its designer. The true Designer of a machine is thus its explicit designer, together with the environment which by sending it information allows the machine to enhance the original design.

"What is an Intelligent Machine?" In this and other articles Ashby takes to task the notion that the brain is a magical device with supernatural skills, arguing that the comparison between "smart" brains and "stupid" computers is stacked against the machines when the intelligence demanded is matched to the everyday environment for which we have been highly preprogrammed (designed, in the language of the prior article) both genetically and through decades of learning. For intelligence is basically goal-seeking activity and appropriate selection, and for this both brains and machines are subject to the same limits, such as the Law of Requisite Variety. A high intelligence is shown by a device which utilizes information efficiently to achieve a high degree of appropriate selection. Even the so-called "genius" cannot escape the laws of information which govern the process. In this excellent article Ashby gives us an exceptionally clear and crisp discussion of the subject of intelligence, as applied to brains and synthetic machines.

DESIGN FOR AN INTELLIGENCE-AMPLIFIER

SECTION I

1. Introduction

For over a century man has been able to use, for his own advantage, physical power that far transcend those produced by his own muscles. Is it impossible that he should develop machines with "synthetic" intellectual powers that will equally surpass those of his own brain? I hope to show that recent developments have made such machines possible – possible in the sense that their building can start today. Let us then consider the question of building a mechanistic system for the solution of problems that are beyond the human intellect. I hope to show that such a construction is by no means impossible, even though the constructors are themselves quite averagely human.

There is certainly no lack of difficult problems awaiting solution. Mathematics provides plenty, and so does almost every branch of science. It is perhaps in the socail and economic world that such problems occur most noticeably, both in regard to their complexity and to the great issues that depend on them. Success in solving these problems is a matter of some urgency. We have built a civilization beyond our understanding and we are finding that it is getting out of hand. Faced with such problems, what are we do do?

Our first instinctive action is to look for someone with corresponding intellectual powers: we think of a Napoleon or an Archimedes. But detailed study of the distribution of man's intelligence shows that this method can give little. Figure 1, for instance, shows the distribution of the Intelligence Quotient in the

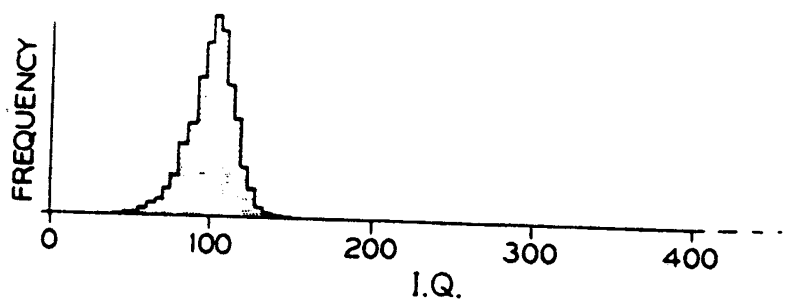


FIGURE 1. Distribution of the adult human Intelligence Quotient (after Wechsler, 1).

normal adult population, as found by Wechsler[1]. What is important for us now

is not the shape on the left but the absolute emptiness on the right. A variety of tests by other workers have always yielded about the same result: a scarcity of people with I.Q.s over 150, and a total absence of I.Q.s over 200. Let us admit frankly that man's intellectual powers are as bounded as are those of his muscles. What then are we to do?

We can see something of how to proceed by comparing our position today in respect to intellectual problems with the position of the Romans in respect to physical problems. The Romans were doubtless often confronted by engineering and mechanical problems that demanded extreme physical strength. Doubtless the exceptionally strong slave was most useful, and doubtless the Romans sometimes considered the possibility of breeding slaves of even greater strength. Nevertheless, such plans were misdirected: only when men turned from their own powers to the powers latent in nature was the revolution inaugurated by Watt possible. Today, a workman comes to his task with a thousand horsepower available, though his own muscles will provide only about one-tenth. He gets this extra power by using a "power-amplifier". Had the present day brain-worker an "intelligence-amplifier" of the same ratio, he would be able to bring to his problems an I.Q. of a million.

If intellectual power is to be so developed, we must, somehow, construct amplifiers for intelligence – devices that, supplied with a little intelligence, will emit a lot. To see how this is to be done, let us look more closely at what is implied.

2. The Criterion of Intelligence

Let us first be clear about what we want. There is no intention here to inquire into the "real" nature of intelligence (whatever that may mean). The position is simple: we have problems and we want answers. We proceed then to ask, where are the answers to be found?

It has often been remarked that any random sequence, if long enough, will contain *all* the answers. Nothing prevents a child from doodling

$$\cos^2 x + \sin^2 x = 1,$$

or a dancing mote in the sunlight from emitting the same message in Morse or a similar code. Let us be more definite. If each of the above 13 symbols might have been any one of 50 letters and elementary signs, then as 50^{13} is approximately 2^{73}, the equation can be given in coded form by 73 binary symbols. Now consider a cubic centimeter of air as a turmoil of colliding molecules. A particular molecule's turnings after collision, sometimes to the left and sometimes to the right, will provide a series of binary symbols, each 73 of which, on some given code, either will or will not represent the equation. A simple calculation from the known facts shows that the molecules in every cubic centimeter of air are emitting this sequence *correctly* over a hundred thousand times a second. The objection that "such things don't happen" cannot stand.

Doodling, then, or any other random activity, is capable of producing all that is required. What spoils the child's claim to be a mathematician is that he will doodle, with equal readiness, such forms as

$$\cos^2 x + \sin^2 x = 2 \qquad \text{or} \qquad \text{ci)xsi} = \text{nx1}$$

or any other variation. After the child has had some mathematical experience he will stop producing these other variations. He becomes not more but less productive: he becomes selective.

The close, indeed essential, relation between intelligence and selection is shown clearly if we examine the tests specially devised for its objective measurement. Take, for instance, those of the Terman and Merrill[2] series for Year IV. In the first test the child is shown a picture of a common object and is asked to give its name. Out of all the words he knows he is asked to select one. In the second test, three model objects – motor-car, dog, show – are placed in a row and seen by the child; then all are hidden from him and a cover is placed over the dog; he is then shown motor-car, cover, shoe, and asked what is under the cover. Again his response is correct if, out of all possible words, he can select the appropriate one. Similarly the other tests, for all ages, evoke a response that is judged "correct" or "incorrect" simply by the subject's power of appropriate selection.

The same fact, that getting a solution implies selection, is shown with special clarity in the biological world. There the problems are all ultimately of how to achieve survival, and survival implies that the essential variables – the supply of food, water, etc. – are to be kept within physiological limits. The solutions to these problems are thus all selections from the totality of possibilities.

The same is true of the most important social and economic problems. What is wanted is often simple enough in aim – a way of ensuring food for all with an increasing population, or a way of keeping international frictions small in spite of provocations. In most of these problems the aim is the keeping of certain variables within assigned limits; and the problem is to find, amid the possibilities, some set of dynamic linkages that will keep the system both stable, and stable within those limits. Thus, finding the answer is again equivalent to achieving an appropriate selection.

The fact is that in admiring the *productivity* of genius our admiration has been misplaced. Nothing is easier than the generation of new ideas: with some suitable interpretation, a kaleidoscope, the entrails of a sheep, or a noisy

vacuum tube will generate them in profusion. What is remarkable in the genius is the discrimination with which the possibilities are winnowed.

A possible method, then, is to use some random source for the generation of all the possibilities and to pass its output through some device that will select the answer. But before we proceed to make the device we must dispose of the critic who puts forward this well known argument: as the device will be made by some designer, it can select only what he has made it to select, so it can do no more than he can. Since this argument is clearly plausible, we must examine it with some care.

To see it in perspective, let us remember that the engineers of the middle ages, familiar with the principles of the lever and cog and pulley, must often have said that as no machine, worked by a man, could pull out more work than he put in, therefore no machine could ever amplify a man's power. Yet today we see one man keeping all the wheels in a factory turning by shovelling coal into a furnace. It is instructive to notice just how it is that today's stoker defeats the medieval engineer's dictum, while being still subject to the law of the conservation of energy. A little thought shows that the process occurs in two stages. In Stage One the stoker lifts the coal into the furnace; and over this stage energy is conserved strictly. The arrival of the coal in the furnace is then the beginning of Stage Two, in which again energy is conserved, as the burning of the coal leads to the generation of steam and ultimately to the turning of the factory's wheels. By making the whole process, from stoker's muscles to factory wheel, take place in two stages, involving two lots of energy whose sizes can vary with some independence, the modern engineer can obtain an overall amplification. Can we copy this method in principle so as to get an amplification in selection?

3. The Selection-amplifier

The essence of the stoker's method is that he uses his (small) power to bring into action that which will provide the main power. The designer, therefore, should use his (small) selectivity to bring into action that which is going to do the main selecting. Examples of this happening are common-place once one knows what to look for. Thus a garden sieve selects stones from soil; so if a gardener has sieves of different mesh, his act of selecting a sieve means that he is selecting, not the stones from the soil, but that which will do the selecting. The end result is that the stones are selected from the soil, and this has occurred as a consequence of his primary act; but he has achieved the selection mediately, in two stages. Again, when the directors of a large firm appoint a Manager of Personnel, who will attend to the selection of the staff generally, they are

selecting that which will do the main selecting. When the whole process of selection is thus broken into two stages the details need only a little care for there to occur an amplification in the degree of selection exerted.

In this connection it must be appreciated that the *degree* of selection exerted is not defined by what is selected: it depends also on what the object is selected from. Thus, suppose I want to telephone for a plumber, and hesitate for a moment between calling Brown or Green, who are the only two I know. If I decide to ring up Green's number I have made a one-bit selection. My secretary, who will get the number for me, is waiting with directory in hand; she also will select Green's number, but she will select it from 50,000 other numbers, a 15.6-bit selection. (Since a 1-bit selection has directly determined a 15.6-bit selection, some aplification has occurred.) Thus two different selectors can select the same thing and yet exert quite different degrees of selection.

The same distinction will occur in the machine we are going to build. Thus, suppose we are tackling a difficult social and economic problem; we first select what we want, which might be:

An organization that will be stable at the conditions:		
Unemployed	<	100,000 persons
Crimes of violence	<	10 per week
Minimal income per family	>	£500 per annum

This is *our* selection, and its degree depends of what other conditions we might have named but did not. The solving-machine now has to make *its* selection, finding this organization among the multitudinous other possibilities in the society. We and the solving-machine are selecting the same entity, but we are selecting it from quite different sets, or contexts, and the degrees of selection exerted can vary with some independence. (The similarity of this relation with those occurring in information theory is unmistakable; for in the latter the information-content of a message depends not only on what is in the message but on what population of message it came from[3,4].)

The building of a true selection-amplifier – one that selects over a greater range than that covered when it was designed – is thus possible. We can now proceed to build the system whose selectivity, and therefore whose intelligence, exceeds that of its designer.

(From now on we shall have to distinguish carefully between *two* problems: our problem, which is to design and build the solving-machine, and the solving-machine's problem – the one we want it to solve.)

4. Basic Design

Let us suppose for definiteness that the social and economic problem of the

previous article is to be the solver's problem. How can we design a solver for it? The construction would in practice be a formidable task, but here we are concerned only with the principles. First, how is the selection to be achieved automatically?

SELECTION BY EQUILIBRIUM. We can take advantage of the fact that if any two determinate dynamic systems (X and S in Figure 2) are coupled through channels G and U so that each affects the other, then any resting state of the whole (that is, any state at which it can stay permanently), must be a resting state in each of the two parts individually, each being in the conditions provided by the other. To put it more picturesquely, each part has a power of veto over resting states proposed by the other. (The formulation can be made perfectly precise in the terms used in Article 6.)

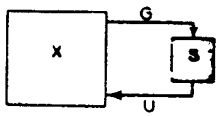


FIGURE 2

It is only a change of words to say that each part acts selectively towards the resting states of the other. So if S has been specially built to have resting states only on the occurrence of some condition ξ in S, then S's power of veto ensures that a resting state of the whole will always imply ξ in S. Suppose next that the linkage G is such that G will allow ξ to occur in S if and only if the condition η occurs in X. S's power of veto now ensures that any resting state of the whole must have condition η in X. So the selection of S and G to have these properties ensures that the only states in X that can be permanent are those that have the condition η.

It must be noticed that the selection of η, in the sense of its retention in X, has been done in two stages. The first occurred when the designer specified S and G and ξ. The second occurred when S, acting without further reference to the designer, rejected state after state of X, accepting finally one that gave the condition η in X. The designer has, in a sense, selected η, as an ultimate consequence of his actions, but his actions have worked through two stages, so the selectivity achieved in the second stage may be larger, perhaps much larger, than that used in the first.

The application of this method to the solving of the economic problem is, in principle, simple. We identify the real economic world with X and the conditions that we want to achieve in it with η. The selection of η in X is beyond our power, so we build, and couple to it, a system S, so built that it has a resting state if and only if its information through G is that η has occurred in a resting state in X. As time progresses, the limit of the whole system, X and S, is the permanent retention of η in X. The designer has to design and build S and G, and to couple it to X; after that the process occurs, so far as he is concerned, automatically.

5. The Homeostat

To see the process actually at work, we can turn briefly to the Hemeostat. Though it has been described fully elsewhere[5], a description of how its action appears in the terms used here may be helpful in illustration. (Figure 3 is intended to show its principle, not its actual appearance.)

It consists of four boxes (F) of components, freely supplied with energy, that act on one another in a complex pattern of feedbacks, providing initially a somewhat chaotic system, showing properties not unlike those sometimes seen in our own society. In this machine, S has been built, and G arranged, so that S has a resting state when and only when four needles N are stable at the central positions. These are the conditions η. N and the F's correspond to X in Figure 2. S affects the F's through the channel U, whose activity causes changes in the conditions within the boxes.

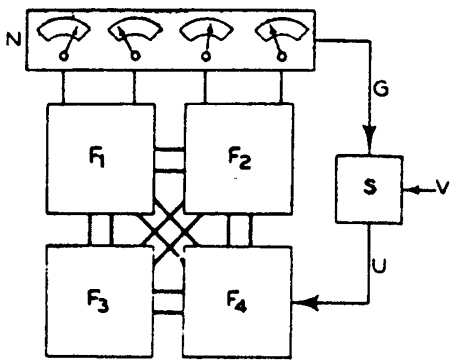


FIGURE 3

Suppose now that the conditions within the boxes, or in the connections between them, are set in some random way, say as a by-stander pleases; then if the conditions so set do not satisfy η, S goes into activity and enforces changes until η is restored. Since η may refer to certain properties of the stability within F, the system has been called "ultrastable", for it can regulate the conditions of its own stability within F. What is important in principle is that the combinations in F that restore η were not foreseen by the designer and programmer in detail; he provided only a random collection of about 300,000 combinations (from a table of random numbers), leaving it to S to make the detailed selection.

One possible objection on a matter of principle is that all the variation going to the trials in X seems, in Figure 2, to be coming from S and therefore from the designer, who has provided S. The objection can easily be met, however, and the alteration introduces an important technical improvement. To make the objection invalid, all the designer has to do is to couple S, as shown in Figure 3, to some convenient source of random variation V – a noisy vacuum tube say – so that the SV combination

(i) sends disturbance of inexhaustible variety along U if ξ is not occurring in S, and

(ii) keeps U constant, i.e., blocks the way from V to U, if ξ *is* occurring.

In the Hemeostat, V is represented by the table of random numbers which determined what entered F along U. In this way the whole system, X and S, has available the inexhaustible random variation that was suggested in Article 2 as a suitable source for the solutions.

6. Abstract Formulation

It is now instructive to view the whole process from another point of view, so as to bring out more clearly the deep analogy that exists between the amplification of power and that of intelligence.

Consider the engineer who has, say, some ore at the foot of a mine-shaft and who wants it brought to the surface. The power required is more than he can supply personally. What he does is to take some system that is going to change, by the laws of nature, from low entropy to high, and he couples this system to his ore, perhaps through pistons and ropes, so that "low entropy" is couple to "ore down" and "high entropy" to "ore up". He then lets the whole system go, confident that as the entropy goes from low to high so will it change the ore's position from down to up.

Abstractly (Figure 4) he has a process that is going, by the laws of nature, to pass from state H_1 to state H_2. He wants C_1 to change to C_2. So he couples H_1 to C_1 and H_2 to C_2. Then the system, in changing from H_1 to H_2, will change from C_1 to C_2, which is what he wants. The arrangmement is clearly both necessary and sufficient.

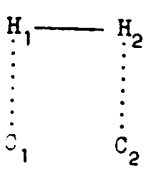


FIGURE 4

The method of getting the problem-solver to solve the set problem can now be seen to be of essentially the same form. The job to be done is the bring of X, in Figure 2, to a certain condition or "sollution" η. What the intelligence engineer does first to build a system, X and S, that has the tendency, by the laws of nature, to go to a state of equilibrium. He arranges the coupling between them so that "not at equilibrium" is coupled to not-η, and "at equilibrium" to η. He then lets the system go, confident that as the passage of time takes the whole to an equilibrium, so will the conditions in X have to change from not η to η. He does not make the conditions in X change by his own efforts, but allows the basic drive of nature to do the work.

This is the fundamental principle of our intelligence-amplifier. Its driving

power is the tendency for entropy to increase, where "entropy" is used, not as understood in heat engines but as understood in stochastic processes.

AXIOMATIC STATEMENT. Since we are considering systems of extreme generality, the best representation of them is given in terms of the theory of sets. I use the concepts and terminology of Bourbaki[6].

From this point of view a machine, or any system that behaves in a determinate way, can be at any one of a set of states at a given moment. Let M be the set of states and μ some one of them. Time is assumed to be discrete, changing by unit intervals. The internal nature of the machine, whose details are irrelevant in this context, causes a transformation to occur in each interval of time, the state μ passing over determinately to some state μ' (not necessarily different from μ), thereby defining a mapping t of M in M:

$$t: \mu \to \mu' = t(\mu).$$

If the machine has an input, there will be a set I of input states ι, to each of which will correspond a mapping t_ι. The states ι may, of course, be those of some other machine, or may be determined by it; in this way machine may be coupled to machine. Thus if machine N with states ν has a set K of inputs κ and transformations u_κ, then machines M and N can be coupled by defining a mapping ζ of M in K, $\kappa = \zeta(\mu)$, and a mapping of m of N in I, $\iota = m(\nu)$, giving a system whose states are the couples (μ, ν) and whose changes with time are defined by the mapping, of M x N in M x N:

$$(\mu, \nu) \to ({}^{t}m(\nu)(\mu),\ u_{\zeta(\mu)}(\nu)).$$

The abstract specification of the principle of ultrastability is as follows, Figure 5 corresponding to Figure 2:

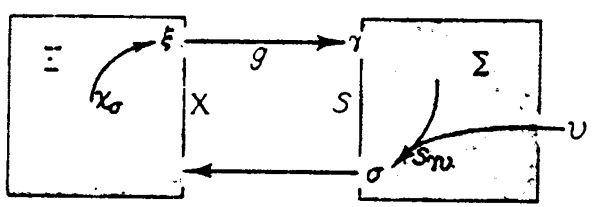


FIGURE 5

GIVEN:

(1) A set Γ consisting of two elements γ_1 and γ_2;

(2) A set Ξ of elements ξ;

(3) A mapping g of Ξ in Γ;

(4) A set Σ of elements σ;

(5) A family of mappings χ_σ of Ξ in Ξ;

(6) A random variable ν, with inexhaustible variety;

(7) A double family of mappings $s_{\gamma\nu}$ of Σ in Σ, with the property that, for all $\sigma \in \Sigma$ and all values of ν,

$$s_{\gamma_1}(\sigma) \neq \sigma \qquad \text{and} \qquad s_{\gamma_2}(\sigma) \neq \sigma;$$

(8) Time, advancing by discrete intervals, induces the operations χ_σ and $s_{\gamma\nu}$ and the successive values of ν simultaneously, once in each interval.

THEOREM: If the series of states of Ξ, induced by time, has a limit $\xi^* \in g^{-1}(\gamma_2)$.

PROOF. The state of the whole system, apart from ν, is given by the couple (ξ, σ), an element in $\Xi \times \Sigma$. The passage of one interval of time induces the mapping $(\xi, \sigma) \to \left(\chi_\sigma(\xi), s_{g(\xi), \nu}(\sigma) \right)$. If the series is at a limit-state (ξ^*, σ^*), then $s_{g(\xi^*), \nu}(\sigma^*) = \sigma^*$ for all values of ν. Therefore $g(\xi^*) = \gamma_2$, and $\xi^* \in g^{-1}(\gamma_2)$.

SECTION II

With this theorem our problem is solved, at least in principle. Systems so built can find solutions, and are not bounded by the powers of their designers. Nevertheless, there is still a long way to go before a man-made intelligence-amplifier will be actually in operation. Space prohibits any discussion of the many subsidiary problems that arise, but there is one objection that will probably be raised, suggesting that the method can never be of use, that can be dealt with here. It is of the greatest importance in the subject, so the remainder of the paper will be devoted to its consideration.

7. Duration of Trials

What we have discussed so far has related essentially to a process that, it is claimed, has "solution" as its limit. The question we are now coming to is, how fast is the convergence? how much time will elapse before the limit is reached?

A first estimate is readily made. Many of the problems require that the answer is an n-tuple, the solver being required to specify the value of each of n components. Thus the answer to an economic problem might require answeres to each of the questions:

(1) What is the optimal production-ratio of coal to oil?

(2) What amount should be invested annually in heavy industry?

(3) What should be the differential between the wages of skilled and unskilled workers?

except that in a real system n would be far larger than three.

A first estimate of the number of states to be searched can be made by finding the number of states in each component, assuming independence, and then finding the product so as to give the number of combinations. The time, with a state by state search, will then be proportional to this product. Thus, suppose that there are on a chessboard ten White and ten Black men; each can move to one of six squares; how many possibilities will have to be searched if I am to find the best next two moves, taking White's two moves and Black's two into account? With each man having about six moves (when captures are allowed for), the possible moves at each step are approximately 6^{10}; and the total possibilities are about 6^{40}. To find the best of them, even if some machine ran through them at a million a second, would take nearly a billion billion years – a prohibitively long time. The reason is that the time taken increases, in this estimate, exponentially with the number of components; and the exponential rate of increase is extremely fast. The calculation is not encouraging, but it is very crude; may it be seriously in error?

It is certainly in error to some extent, for it is not strictly an estimate but an upper bound. It will therefore always err by over-estimation. This however is not the chief reason for thinking that our method may yet prove practical. The reason for thinking this will be given in the next three articles.

8. The Method of Models

The first factor that can be used to reduce the time of search is of interest because it is almost synonymous with the method of science itself. It is, to conduct the search, not in the real physical thing itself but in a model of it, the model being chosen so that the search can proceed in it very much more rapidly. Thus Leverrier and Adams, searcing for a planet to explain the aberrations of Uranus, used pencil, paper and mathematics rather than the more obvious telescope; in that way they found Neptune in a few months where a telescopic search might have taken a lifetime.

The essence of the method is worth noticing explicitly. There is a set R, containing the solutions r, a subset of R; the task is to find a number of r. The method of models can be used if we can find some other set R′ whose elements can be put into correspondence with those R in such a way that the elements (a set r′) in R′ that correspond to those in r can be recognized. The search is then conducted in R′ for one of r′; when successful, the correspondence, used inversely, identifies a solution in R. For the method to be worth using, the

search in R′ must be so much faster than that in R that the time taken in the three operations

(i) change from R to R′,
(ii) search in R′,
(iii) change back from r′ to r

is less than that taken in the single operation of searching in R.

Such models are common and are used widely. Pilot plants are used for trials rather than a complete workshop. Trials are conducted in the drawing-office rather than at the bench. The analogue computer is, of course, a model in this sense, and so, in a more subtle way, is the digital computer. Mathematics itself provides a vast range of models which can be handled on paper, or made to "behave", far faster than the systems to which they refer.

The use of models with the word extended to mean any structure isomorphic with that of the primary system, can thus often reduce the time taken to a fraction of what might at first seem necessary.

9. Constraints

A second reason why the time tends to fall below that of the exponential upper bound is that often the components are *not* independent, and the effect of this is always to reduce the range of possibilities. It will, therefore, other things being equal, reduce the time of search.

CONSTRAINT BY RELATION. Suppose we are looking for a solution in the n-tuple $(a_1, \ldots, a_n)$, where a_1 is an element in a set A_1, etc. The solution is then one of a subset of the product set $A_1 \times A_2 \times \ldots \times A_n$. A relation between the a's, $\phi(a_1, \ldots, a_n)$, always defines a subset of the product space [6], so if the relation holds over the a's, the solution will be found in the subset of the product-space defined by ϕ. An obvious example occurs when there exist invariants over the a's. k invariants can be used to eliminate k of the a's, which shows that the original range of variation was over n − k, not over n, dimensions. More generally, every law, whether parliamentary, or natural, or algebraic[7] is a constraint, and acts to narrow the range of variation and, with it, the time of search.

Similarly, every "entity" that can be recognized in the range of variation holds its individuality only if its parts do *not* vary over the full range conceivable. Thus, a "chair" is recognizable as a thing partly because its four legs do not move in space with all the degrees of freedom possible to four independent objects. The fact that their actual degrees of freedom are 6 instead of 24 is a measure of their cohesion. Conversely, their cohesion implies that any reckoning of possibilities must count 6 dimensions in the product or phase space, not 24.

It will be seen therefore that ***every relation that holds between the components of an n-tuple lessens the region of search.***

CONSTRAINT BY CONTINUITY. Another common constraint on the possibilities occurs where there are functional relations within the system such that the function is continuous. Continuity is a restriction, for if y = f(z) and f is continuous and a series of arguments z, z′ and z″, . . . has the limit z*, then the corresponding series y, y′, y″ . . . must have the limit f(z*). Thus f is not free to relate y and z, y′ and z′, y″ and z″, . . . arbitrarily, as it could do if it were unrestricted. This fact can also be expressed by saying that as adjacent values of z make the values of f(z) adjacent, these values of f(z) tend to be highly correlated, so that the values of f(z) can be adequately explored by a mere sampling of the possibilities in z: the values of y do not have to be tested individually.

A rigorous discussion of the subject would lead into the technicalities of topology; here it is sufficient to notice that the continuity of f puts restrictions on the range of possibilities that will have to be searched for a solution.

Continuity helps particularly to make the search easy when the problem is to find what values of a, b, c, . . . will make some function $\lambda(a,b,c, \ldots)$ a maximum. Wherever λ is discontinuous and arbitrary there is no way of finding the maximum, or optimum, except by trying every combination of the arguments individually; but where λ is continuous a maximum can often be proceeded to directly. The thesis is well illustrated in the art of aircraft design, which involves knowing the conditions that will make the strength, lightness, etc., an optimum. In those aspects of design in which the behavior of the aircraft is a continuous function of the variables of design, the finding of an optimum is comparatively direct and rapid; where however the relations are discontinuous, as happens at the critical values of the Reynolds' and Mach numbers, then the finding of an optimum is more laborious.

Chess, in fact, is a difficult game largely because of the degree to which it is discontinuous, in the sense that the "value" of a position, to White say, is by no means a continuous function of the positions of the component pieces. If a rook, for instance, is moved square by square up a column, the successive values of the positions vary, sometimes more or less continuously, but often with marked discontinuity. The high degree of discontinuity occurring throughout a game of chess makes it very unlike the more "natural" systems, in which continuity is common. With this goes the corollary that the test so often proposed for a mechanical brain – that it should play chess – may be misleading, in that it is by no means representative of the class of problem that a real mechanical brain will one day have to deal with.

The commonness of continuity in the world around us has undoubtedly

played an important part in Darwinian evolution in that progress in evolution would have been far slower had not continuity been common. In this paper I have tended to stress the analogy of the solving process with that of the amplification of physical power I could equally have stressed its deep analogy with the processes of evolution, for there is the closest formal similarity between the process by which adaptation is produced automatically by Darwinian selection and the process by which a solution is produced automatically by mechanical selection of the type considered in Article 4. Be that as it may, every stock-breeder knows that selection for good strains can procede much more rapidly when the relation between genotype and phenotype is continuous[8]. The time taken for a given degree of improvement to be achieved is, of course, correspondingly reduced.

To sum up: Continuity being common in the natural world, the time taken in the solution of problems coming from it may be substantially below that given by the exponential bound.

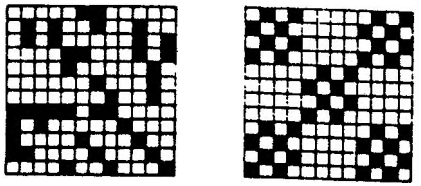
CONSTRAINT BY PRIOR KNOWLEDGE. It is also only realistic to consider, in this connection, the effect on solving of knowledge accumulated in the past. Few problems are wholly new, and it is merely common sense that we, and the problem-solver, should make use of whatever knowledge has already been won.

The effect of such knowledge is again to put a constraint on the possibilities, lessening the regions that have to be searched, for past experience will act essentially by warning us that a solution is unlikely to lie in certain regions. The constraint is most marked when the problem is one of a class, of which several have already been solved. Having such knowledge about other members of the same class is equivalent to starting the solving process at some point that is already partly advanced towards the goal. Such knowledge can naturally be used to shorten the search.

"Solving a problem" can in fact be given a perfectly general representation. The phase- or sample-space of possibilities contains many points, most of them corresponding to "no solution" but a few of them corresponding to an acceptable "solution". Finding a solution then becomes a matter of starting somewhere in this unknown distribution of states and trying to find one of the acceptable states.

If the acceptable states are distributed wholly at random, i.e., with no recognizable pattern, then we are considering the case of the problem about which nothing, absolutely nothing, is known. After 200 years of scientific activity, such problems today are rare; usually some knowledge is available from past experience, and this knowledge can be used to constrain the region of search, making it smaller and the search more rapid. Thus suppose, as a simple example,

that there are 114 states to be examined and that 40 of them are acceptable, i.e., correspond to solutions. If they are really scattered at random, as are the black squares in I of Figure 6, then the searcher has no better resource than to start somewhere and to wander at random. With this method, in this particular example he will usually require about 3.6 trials to find a black square. If, however, past experience has shown that the black squares are distributed as in II, advantage can be taken to shorten this search; for wherever the search may start, the occurrence of two white squares in succession shows the searcher that he is in an all-white quadrant. A move of four squares will take him to a checkered quadrant where he will find a dark square either at once or on the next move. In this case his average number of trials need not exceed about 2.4 (if we ignore the complications at the boundaries).



I FIGURE 6 II

Information theory and the strategy of games are both clearly involved here, and an advanced knowledge of such theories will undoubtedly be part of the technical equipment of those who will have to handle the problem-solvers of the future. Meanwhile we can sum up this Article by saying that it has shown several factors that tend to make the time of search less than that given by the exponential bound.

10. Selection by Components

We now come to what is perhaps the most potent factor of all for reducing the time taken. Let us assume, as before, that the solver is searching for an element in a set and that the elements are n-tuples. The solver searches, then, for the n values of the n components.

In the most "awkward" problems there is no means of selection available other than the testing of every element. Thus, if each of the n components can take any one of k values, the number of states to be searched is k^n, and the time taken will be proportional, as we saw in Article 7, to the exponential bound. This case, however, is the worst possible.

Next let us consider the other extreme. In this case the selection can be conducted component by component, each component being identified independently of the others. When this occurs, a very great reduction occurs in the time taken, for now the number of states to be searched is kn.

This second number may be very much smaller than the first. To get some idea of the degree of change implied, let us take a simple example. Suppose there are a thousand components, each of which has a hundred possibilities. If

the search has to be over every *combination* separately, the number of states to be searched is the exponential bound, 100^{1000}. If the states could be examined at one a second it would take about 10^{1993} years, a duration almost beyond thinking. If we tried to get through the same selection by the method of models, using some device that could test, say, a million in each microsecond, the time would drop to 10^{1981} years, which is practically no change at all. If however the components could be selected individually and independently, then the number of selections would drop to 100,000; and at one per second the problem could be solved in a little over a day.

This illustration will suffice to illustrate the general rule that selection by components is so much faster than selection by elements that a search that is utterly impractical by the one may be quite easy by the other.

The next questions is to what extent problems are susceptible of this faster method. About *all* problems one can say nothing, for the class is not defined, but about the problems of social and economic type, to which this paper is specially directed, it is possible to form some estimate of what is to be expected. The social and economic system is highly dynamic, active by its own energies, and controllable only through a number of parameters. The solver thus looks for suitable values in the n-tuple of parameters. (Its number of components will not usually be the same as those of the variables mentioned in the next paragraph.)

Now a dynamic system may be "reducible"; this happens when the system that seems to be single really consists of two or more adjacent, but functionally independent, parts. If the parts, P and Q say, are quite independent, the whole is "completely" reducible; if part P affects part Q, but Q has no effect on P, then the whole is simple "reducible". In the equations of such a system, which can be given in the canonical form[5]

$$\frac{dx_1}{dt} = f_1(x_1, \ldots, x_n)$$

$$\frac{dx_n}{dt} = f_n(x_1, \ldots, x_n)$$

reducibility is shown by the Jacobian of the n functions f_1 with respect to the n variables x_j having, when partitioned into the variables that belong to P and those that belong to Q, one or more all-zero quadrants, being like I when reducible and like II when completely reducible:

$$\begin{bmatrix} J_1 & O \\ K & J_2 \end{bmatrix} \qquad \begin{bmatrix} J_1 & O \\ O & J_2 \end{bmatrix}$$

I II

Such reducibility corresponds, in a real dynamic system, with the possibility of searching to some extent by components, the components being the projections[6] of the abstract space on those variables, or "coordinates", that occur together in one part (P, or Q) of the reducible system. The more a system is reducible, the more does it offer the possibility of the search being made by the quick method of finding the components separately.

Though the question has not yet been adequately explored, there is reason to believe that part-functions, i.e., variables whose f_1 (above) are zero for many values of their arguments, are common in social and economic systems. They are ubiquitous in the physical world, as I have described elsewhere[5]. Whenever they occur, they introduce zeros into the Jacobian of the system (for if x_j is constant, over some interval of time, f_1 must be zero and therefore so will be $\delta f_1 / \delta x_j$); they therefore tend to introduce temporary reducibilities. This means that the problem may be broken up into a series of conditional and transient sub-problems, each of which has a solution that can be found more or less easily. The whole search thus has something of the method of search by components.

In this connection it is instructive to consider an observation of Shannon's[9] on the practicability of using relays as devices for switching, for it involves a property closely related to that of reducibility. One of the problems he solved was to find how many elements each relay would have to operate if the network was to be capable of realizing all possible functions on n variables. The calculation gave a number that was, apparently, ridiculously high for, did one not know, it would suggest that relays were unsuitable for practical use in switching. The apparent discrepancy proved to be due to the fact that the functions commonly required in switching are not as complicated as the class that can be considered in theory. The more they *look* complicated the more they tend to have hidden simplicities. These simplicities are of the form in which the function is "separable", , that is to say, the variables go in sets, much as the variables in a reducible system go in sets. Separability thus makes what seems to be an impractical number of elements become practical. It is not unlikely that reducibility will act similarly towards the time of search.

11. The Lower Bound

It can now be seen that problem-solving, as a process of selection, is related to the process of message-receiving as treated in information theory[3, 10]. The connection can be seen most readily by imagining that the selection of one object from N is to be made by an agent A who acts according to instructions received from B. As the successive elements appear, B will issue signals: ". . . ,

reject, reject, . . . , accept" thereby giving information in calculable quantity to A. Usually the number of binary signals so given is likely to be greater than the number necessary, for the probabilities of the two signals, "reject" and "accept", are by no means equal. The most efficient method of selection is that which makes the two signals equally likely. This will happen in the whole set of N can be dichotomized at each set of selection. In this way we can find a *lower bound* to the time taken by the process, which will be proportional to log N. This time is none other than the least time in which, using binary notation, the solution can be written down, that is, identified from its alternatives.

We see therefore that, though the upper (exponential) bould is forbiddingly high, the lower bound is reassuringly low. What time will actually be taken in some particular problem can only be estimated after a direct study of the particular problem and of the resources available. I hope, however, that I have said enough to show that the mere mention of the exponential bound is not enough to discredit the method proposed here. The possibility that the method will work is still open.

Summary

The question is considered whether it is possible for human constructors to build a machine that can solve problems of more than human difficulty. If physical power can be amplified, why not intellectual?

Considerations show that:

Getting an answer to a problem is essentially a matter of selection.

Selection can be amplified.

A system with a selection-amplifier can be more selective than the man who built it.

Such a system is, in principle, capable of solving problems, perhaps in the social and economic world, beyond the intellectual powers of its designer.

A first estimate of the time it will take to solve a difficult problem suggests that the time will be excessively long: closer examination shows that the estimate is biased, being an upper bound.

The lower bound, achievable in some cases, is the time necessary for the answer to be written down in binary notation.

It is not impossible that the method may be successful in those social and economic problems to which the paper is specially addressed.

Bibliography

1 WECHSLER, D., *Management of Adult Intelligence.* Baltimore, Williams & Wilkins, 3rd Ed., 1944.

2 TERMAN, L.M., and MERRILL, M.A., *Measuring Intelligence.* London, Harrap & Co., 1937.

3 SHANNON, C.E. and WEAVER, W., *The Mathematical Theory of Communication.* Urbana, University of Illinois Press, 1949.

4 ASHBY, W. ROSS, *Can A Mechanical Chess-player Outplay its Designer?* Brit. J. Phil. Sci., 3, 44: 1952.

5. ASHBY, W. ROSS, *Design for a Brain.* London, Chapman & Hall; New York, John Wiley & Sons, 1952.

6 BOURBAKI, N., "Theorie des Ensembles." *A.S.E.I.* 1141. Paris, Herman et Cie, 2nd ed., 1951.

BOURBAKI, N. "Structures Algebriques." *A.S.E.I.* 1144. Paris, Hermann et Cie, 2nd ed., 1951.

8 LERNER, I.M. *Population Genetics and Animal Improvement.* Cambridge University Press, 1950.

9 SHANNON, C.E., "Synthesis of Two-terminal Switching Circuits." *Bell System tech. J.,* 28, 59-98, 1949.

10 WIENER, N., *Cybernetics.* New York, John Wiley & Sons, 1948.

CAN A MECHANICAL CHESS-PLAYER OUTPLAY ITS DESIGNER?*

W. Ross Ashby

1 *Introduction*

Mr Adkins' letter[1] is welcome, for it insists on the discussion of a question often heard in conversation but not yet, so far as I am aware, adequately treated in print: to what extent is the machine restricted by the limitations of its designer? The best form of the question is uncertain, but we can usually recognise it in its different forms. Thus, Descartes declared that there must be at least as much reality and perfection in the cause as in the effect. Kant[2] asked, 'How can work full of design build itself up without a design and without a builder?' Another form has been used as title for this paper.

The question is not only of philosophic interest but is fast approaching practical importance. Some of us are studying the physiology of the brain by building models that will mimic its characteristic properties. Our working hypothesis is that we can eventually build a 'real' brain. But for it to be a 'real' brain it must produce cleverness of its own and not just give us back the ingenuity we have put into it. Here the truth or falsity of Descartes' dictum is crucial. If it is true, we are wasting our time; only if it is false can we succeed. So we must try to disprove it. If we can prove it false, not only may the way be shown open to a 'real' brain but, by scrutinising the proof, we may be able to get a hint about how to proceed, for the proof may show what the real difficulty is, and how it can be overcome.

2 *Measuring 'Design'*

Examining Descartes' dictum, we notice first that it is essentially quantitative. It says, in effect: if the quantity of design supplied by the designer is D, and if the quantity of design shown by the machine is M, then necessarily

$$D > M \quad . \quad . \quad . \quad . \quad . \quad (1)$$

* Received 11. xii. 51

[1] This *Journal*, 1951, 2, 248

[2] *General History of Nature*, 1755

Descartes, however, gave no hint as to how these quantities were to be measured, so we must develop a method. Fortunately we need not enquire exhaustively into all that is implied by 'design,' for the measurement of a quantity can properly precede the understanding of what it is that is being measured—engineers were measuring the 'electric fluid' and were lighting towns with it several decades before its real nature was understood.

How are we to obtain an objective and consistent measure of the 'amount of design' put into, or shown by, a machine? Abstractly, 'designing' a machine means giving selected numerical values to the available parameters. How long shall the lever be? where shall its fulcrum be placed? how many teeth shall the cog have? what value shall be given to the electrical resistance? what composition shall the alloy have? and so on. Clearly, the amount of design must be related in some way to the number of decisions made and also to the fineness of the discrimination made in the selection. I suggest that the measure appropriate to our purpose is one already developed in Shannon's theory of information.[1] Though this theory was developed by communication engineers for technical purposes, it will, I believe, be found to have applications, and to be of philosophic importance, over a much wider range. Its use enables us to treat the question quantitatively, and it provides the discussion with a secure basis of experimental and practical experience.

Shannon's measure is defined fundamentally as follows. Suppose certain events $E_1, E_2, \ldots, E_n$ have probabilities $p_1, p_2, \ldots, p_n$, respectively, of occurring, with

$$p_1 + p_2, + \ldots + p_n = 1,$$

then the occurrence of the actual event is associated with the quantity

$$-\Sigma_j p_j \log p_j \qquad . \quad . \quad . \quad . \quad (2)$$

(If the probabilities are all equal, and each equal to $1/n$, then the quantity becomes simply $\log n$.) As an example, suppose that we are going to draw and inspect one card from a shuffled pack and that we are interested only in the distinction between the three events

E_1 : the drawing of the King of Clubs,
E_2 : the drawing of any Spade,
E_3 : the drawing of any other card;

[1] C. E. Shannon, 'A mathematical theory of communication,' *Bell System tech. J.*, 1948, 27, 379-423, 623-656; Norbert Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, New York, 1949

then $p_1 = \frac{1}{52}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{39}{52}$; and the quantity associated with the drawing of the card, and the discovery of which event has actually occurred, is

$$-\tfrac{1}{52}\log\tfrac{1}{52} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{39}{52}\log\tfrac{39}{52},$$

which, if the logarithms are to the base 2 (see below), has the value 0·94. This number measures the amount of information, relative to these E's, that is to be obtained by drawing a card.

The number 0·94, it should be noticed, does not belong to any particular card, for it can be calculated before the drawing takes place. It does not answer the question 'I have drawn the Three of Hearts: how much information have I received?'; rather it answers the question 'when I draw and look at the card, how much uncertainty will be removed?' It measures, primarily, the uncertainty existing before the drawing is made. This uncertainty will be dispelled by the actual drawing and inspection of the card, and may be used as a measure of the information to be expected in the drawn card if we regard 'information' as 'that which dispels uncertainty.' At first impression this way of measuring information may seem unnatural, almost the reverse of what one would expect; but its usefulness in communication engineering has put its practical value and the soundness of its method beyond doubt.

The amount can be measured in various units, which depend on the base used for the logarithms. The most convenient here is the 'bit'[1]—a contraction of BInary digiT—which is the amount of information given, or uncertainty removed, when a decision is made between two equally probable alternatives: a decision made by the spin of a coin, for instance, or that made after asking 'to be or not to be?' To work in this unit, the logarithms are taken to the base 2. So the drawing of a card in the example above gives a little less information than is obtained by seeing whether a penny has fallen heads or tails.

It is an important property of Shannon's function (2) that if two or more events become indistinguishable, so that some of the E's and p's have to be combined, then the numerical value of the function always diminishes. Thus suppose, in the example above, that we no longer distinguish between E_1 and E_2. p_1 and p_2 then become the single probability $\frac{7}{26}$, i.e. $\frac{1}{52} + \frac{1}{4}$, and the quantity becomes

$$-\tfrac{7}{26}\log\tfrac{7}{26} - \tfrac{39}{52}\log\tfrac{39}{52},$$

[1] The word is inelegant, but it is already firmly established in use.

escartes, however, gave no hint as to how these quantities were to e measured, so we must develop a method. Fortunately we need ot enquire exhaustively into all that is implied by 'design,' for the easurement of a quantity can properly precede the understanding f what it is that is being measured—engineers were measuring the electric fluid' and were lighting towns with it several decades efore its real nature was understood.

How are we to obtain an objective and consistent measure of the amount of design' put into, or shown by, a machine? Abstractly, designing' a machine means giving selected numerical values to the vailable parameters. How long shall the lever be? where shall its ulcrum be placed? how many teeth shall the cog have? what value hall be given to the electrical resistance? what composition shall the lloy have? and so on. Clearly, the amount of design must be elated in some way to the number of decisions made and also to the neness of the discrimination made in the selection. I suggest that he measure appropriate to our purpose is one already developed in hannon's theory of information.[1] Though this theory was deeloped by communication engineers for technical purposes, it will, believe, be found to have applications, and to be of philosophic mportance, over a much wider range. Its use enables us to treat he question quantitatively, and it provides the discussion with a ecure basis of experimental and practical experience.

Shannon's measure is defined fundamentally as follows. Suppose ertain events $E_1, E_2, \ldots, E_n$ have probabilities $p_1, p_2, \ldots, p_n$, espectively, of occurring, with

$$p_1 + p_2, + \ldots + p_n = 1,$$

hen the occurrence of the actual event is associated with the quantity

$$-\Sigma_j p_j \log p_j \qquad . \quad . \quad . \quad . \quad (2)$$

If the probabilities are all equal, and each equal to $1/n$, then the quanity becomes simply $\log n$.) .As an example, suppose that we are going to draw and inspect one card from a shuffled pack and that we re interested only in the distinction between the three events

E_1 : the drawing of the King of Clubs,
E_2 : the drawing of any Spade,
E_3 : the drawing of any other card;

[1] C. E. Shannon, 'A mathematical theory of communication,' *Bell System tech. .*, 1948, 27, 379-423, 623-656; Norbert Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, New York, 1949

then $p_1 = \frac{1}{52}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{39}{52}$; and the quantity associated with the drawing of the card, and the discovery of which event has actually occurred, is

$$-\tfrac{1}{52}\log\tfrac{1}{52} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{39}{52}\log\tfrac{39}{52},$$

which, if the logarithms are to the base 2 (see below), has the value 0·94. This number measures the amount of information, relative to these E's, that is to be obtained by drawing a card.

The number 0·94, it should be noticed, does not belong to any particular card, for it can be calculated before the drawing takes place. It does not answer the question 'I have drawn the Three of Hearts: how much information have I received?'; rather it answers the question 'when I draw and look at the card, how much uncertainty will be removed?' It measures, primarily, the uncertainty existing before the drawing is made. This uncertainty will be dispelled by the actual drawing and inspection of the card, and may be used as a measure of the information to be expected in the drawn card if we regard 'information' as 'that which dispels uncertainty.' At first impression this way of measuring information may seem unnatural, almost the reverse of what one would expect; but its usefulness in communication engineering has put its practical value and the soundness of its method beyond doubt.

The amount can be measured in various units, which depend on the base used for the logarithms. The most convenient here is the 'bit'[1]—a contraction of BInary digiT—which is the amount of information given, or uncertainty removed, when a decision is made between two equally probable alternatives: a decision made by the spin of a coin, for instance, or that made after asking 'to be or not to be?' To work in this unit, the logarithms are taken to the base 2. So the drawing of a card in the example above gives a little less information than is obtained by seeing whether a penny has fallen heads or tails.

It is an important property of Shannon's function (2) that if two or more events become indistinguishable, so that some of the E's and p's have to be combined, then the numerical value of the function always diminishes. Thus suppose, in the example above, that we no longer distinguish between E_1 and E_2. p_1 and p_2 then become the single probability $\frac{7}{26}$, i.e. $\frac{1}{52} + \frac{1}{4}$, and the quantity becomes

$$-\tfrac{7}{26}\log\tfrac{7}{26} - \tfrac{19}{26}\log\tfrac{19}{26}.$$

which is 0·84, i.e. 0·10 less than the previous 0·94. Conversely, if we can make a distinction where previously none was made, then the quantity will be increased.

Any quantity calculated by the formula (2) will be referred to throughout this paper as the 'amount of information,' but I use the phrase purely for convenience, and wish to imply only what is either contained in the definition or deducible from it. I shall attempt to show that this quantity, in this context, is equivalent, in its properties, to the amount of design. This is not to show that information and design are intrinsically the same, but to show that as the object of our study is changed in ways that unquestionably alter the amount of design in it, so does the quantity given by Shannon's function always change similarly. The latter can then conveniently be used as an indicator for the former, just as the height of the mercury in a thermometer-tube can be used as an indicator of the hotness of its surroundings, though 'height' and 'hotness' are by no means identical.

To apply the measure to a designed machine, we regard the machine as something specified by a designer and produced, as output, from a workshop. We must therefore consider not only the particular machine but the *ensemble* of machines from which the final model has been selected. Let me give some examples. They will illustrate the method in detail and will show that the measure agrees satisfactorily with our intuitive sense of 'amount of design.'

First suppose an electrical engineer is going to design a network of resistances. Suppose he can take for granted that there will be three resistances radiating from a point, that their values will lie somewhere between 10 and 100 ohms, and that he need not specify closer than 20 per cent. All other variations have zero probability. What he has to do is to allot to each resistance one of the values 10, 15, 22, 33, 47, 67, or 100 ohms. If the *a priori* probabilities are all equal to $\frac{1}{7}$, the amount of information is, for each resistance, $\log_2 7$ bits, and, for the whole network, $3 \log_2 7$, i.e. 8·42 bits. This number gives an exact value to the symbol D in (1).

Had a finer discrimination been necessary, the amount of information would have been larger. Thus, had each resistance to be specified to the nearest ohm, each resistance would have to be selected in value from 10, 11, 12, . . . , 99, 100 ohms, i.e. from 91 possibilities. So the amount of information would be $3 \log_2 91$, i.e. 19·52 bits. The

extra 11·10 bits (the increase from 8·42 to 19·52) represents, and measures, the extra discrimination that has been used; for it is an important property of Shannon's function that if selections are made in stages, the components combine linearly.

We can use the same function to measure the amount of information (or design) shown by the network (M of (1)). Here again we consider the possible networks, as they differ in their particular values for the resistances; we associate a probability with each distinct network, and we calculate the amount of information in the usual way. It proves to be 8·42 bits; so in this example M equals D.

This equality does not occur necessarily, and in other networks the two quantities may be unequal. Suppose, for instance, that another network, instead of being arranged as three radii, had been arranged as three resistances in parallel. Two variations that differed only by a permutation of their resistances would not be distinguishable electrically; so the designer could make variations that did not make any effective difference to the network. In such a case, M would be less than D.

On the other hand, no possible rearrangement can make M greater than D, so this first example exemplifies Descartes' dictum.

As a second example, consider the case of the designer who, to provide himself with materials, has bought a toy engineering set that contains a definite number of parts capable of being joined in a certain variety of ways. If he constructs his machine from these parts we could, with some labour, calculate the exact amount of information in his final model. Again, the variety of machines constructible cannot exceed the variety available to the designer; and it may be less if some of the forms are indistinguishable. So this second example also exemplifies the dictum. In fact, if this way of measuring 'amount of information' (or design) be accepted, the whole of information theory and all the practical experience of communication engineering becomes available to support the thesis that, in cases such as these, no possible rearrangement of parts can make the amount of information in the machine greater than that used by the designer.

3 *Dynamic Systems*

Exactly the same conclusion is reached if we ignore the form or construction of the machine and consider only its behaviour.

Every determinate machine can be specified in its behaviour by a set of ordinary simultaneous differential equations of the first order

$$\left.\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, \ldots, x_n) \\ &\cdots\cdots \\ \frac{dx_n}{dt} &= f_n(x_1, \ldots, x_n) \end{aligned}\right\} \quad \ldots \quad (3)$$

where the f's are single valued, and where the right-hand side contains no function of t (the time) other than those whose fluxions occur on the left.[1] (For a machine is 'determinate' in its behaviour if the occurrence of a particular state determines what it will do. The equations (3) say this in mathematical form, for if the f's are given, by being appropriate to some particular machine, the occurrence of a particular set of values of the x's is sufficient to determine the changes, the dx's, that will occur during the next instant dt.) If the machine, with variables $x_1, \ldots, x_n$, is brought, at time $t = t_0$, to an arbitrary state $x_1^\circ, \ldots, x_n^\circ$ and then released, the equations, through their integrals, determine how each x will change in value with time.

When a machine is 'designed,' the designer starts with more possibilities available:

$$\left.\begin{aligned} \frac{dx_1}{dt} &= \phi_1(x_1, \ldots, x_n; \alpha_1, \alpha_2, \ldots) \\ &\cdots\cdots\cdots \\ \frac{dx_n}{dt} &= \phi_n(x_1, \ldots, x_n; \alpha_1, \alpha_2, \ldots) \end{aligned}\right\} \quad \ldots \quad (4)$$

where the α's are parameters under his arbitrary control. He then selects the numerical values that they are to have, fixes them permanently, and so restricts the forms ϕ, regarded as functions of the x's, to some particular form, such as (3).

It is well known that the equations (3) define, from each point $(x_1, \ldots, x_n)$, a single trajectory. It follows that if the system (4) has k possible combinations of the α's, the machine cannot show more than k trajectories from any point, though it may show fewer. Measured in this way, the amount of information in its trajectories,

[1] W. Ross Ashby, 'The physical basis of adaptation by trial and error,' *J. gen. Psychol.*, 1945, 32, 13-25; 'The nervous system as physical machine,' *Mind*, 1947, 56, 1-16; L. von Bertalanffy, 'An outline of general system theory,' this *Journal*, 1950, 1, 134-165

or behaviours, cannot exceed the amount primarily introduced by the designer when he decides the values to be given to the α's.

The examples given so far are thus wholly in accord with Descartes' dictum.

4 *Evolution and Design*

The question might seem settled, were it not for the fact, known to every biologist, that Descartes' dictum was proved false over ninety years ago by Darwin. He showed that quite a simple rule, acting over a great length of time, could produce design and adaptation far more complex than the rule that had generated it. The status of his proof was uncertain for some time, but the work of the last thirty years, especially that of the geneticists, has shown beyond all reasonable doubt the sufficiency of natural selection. We face therefore something of a paradox.

There can be no escape by denying the great complexity of living organisms. Neither Descartes nor Kant would have attempted this, for they appealed to just this richness of design as evidence for their arguments. Information theory, too, confirms the richness. Thus, suppose we try to measure the amount of design involved in the construction of a bird that can fly a hundred miles without resting. As a machine, it must have a very large number of parameters adjusted. How many cannot be stated accurately, but it is of the same order as the number of all the facts of avian anatomy, histology, and biochemistry. Unquestionably, therefore, evolution by natural selection produces great richness of design.

Whence comes the richness? Can it be that the rule of natural selection—'the dead shall not breed'—together with its ancillary rules, is really more complex than it seems, and is at least as complex as the design it produces? This seems most unlikely, but even if it were true the paradox would not really be resolved, for the complexity that can be generated by evolution is *independent* of the complexity in these rules. To make the argument clear and quantitative, let me give an example. The essence of the evolutionary process is selection, a single selective operator acting over and over again. Consider, then, a very simple selective machine: a magnet that selects any iron ball from a stream of balls that rolls beneath it. Let us compare the quantity of design that went to its construction with the quantity of selection that it can make, using Shannon's measure for both. When it was designed, the designer had to select its components from those

that were available to him: cogs, magnets, wires, batteries, etc.—what was available in his workshop and what could be bought with his resources. If there were A components, with equal *a priori* probabilities of being selected, the selection of a magnet involved $\log_2 A$ bits. (Theoretically, the accuracy of this measurement is limited only by our ability to give accurate definition to 'workshop,' etc.) The decision to allow balls to roll underneath involves, say, $\log_2 B$ bits; we need not enquire its exact value here. So the total amount of information put into the machine's design is $\log_2 A + \log_2 B$ bits. Now consider what the machine can achieve as a selector: if there is one iron ball in C the machine will achieve a selection of $\log_2 C$ bits. It follows that $\log_2 C$ may well be greater than $\log_2 A + \log_2 B$, for A, B and C have no necessary relationship. Similarly, however complex the specification of 'natural selection,' evolution can in time produce a greater complexity.

Information theory, however, makes clear whence comes the extra information. The law that information cannot be created is not violated by evolution, for the evolving system receives an endless stream of information in the form of mutations. Whatever their origin, whether in cosmic rays or thermal noise, the fact that each gene may, during each second, change unpredictably to some other form makes each gene a typical information source. The information received each second by the whole gene-pattern, or by the species, is then simply the sum of the separate contributions. The evolving system has thus *two* sources of information, that implied in the specification of the rules of natural selection and that implied by the inpouring stream of mutations.

It is now clear that the paradox arose simply because the words 'cause' or 'designer,' in relation to a system, can be used in two senses. If they are used comprehensively, to mean 'everything that contributes to the determination of the system,' then Shannon and Descartes can agree that 'a noiseless transducer or determinate machine can emit only such information as is supplied to it.' This formulation will include the process of evolution if the 'cause' is understood to include not only the rules of natural selection but also the mutations, *specified in every detail*. If, on the other hand, by 'cause' or 'designer' we mean something more restricted—a human designer, say—so that the designer is only a part of the total determination, then the dictum is no longer true.

This gives us what we need. The evolutionary process transcends,

in a sense, the bounds set by the dictum. What we have to do now is to develop a machine that shall, in some way, use an evolution-like process in its working.

5 *Darwinian Machinery* [1]

To develop the machine, first observe that we now have two ways of giving the instructions necessary for the production of, say, a bird. One way is to specify the details in all their complexity and individuality, filling perhaps several volumes in the process. The other way is just to write down the instructions: 'Take a planet with some carbon and oxygen; irradiate it with sunshine and cosmic rays; and leave it alone for a few hundred million years.' We may feel that this second way is somewhat irregular, but we cannot deny that it is effective.

To see what is involved, let us consider an even simpler example. Suppose we offer a prize to any inventor who, by using only 1000 bits [2] in the specification of his machine, can build one that will itself produce more than 1000 bits. Suppose that one inventor proceeds as follows. He makes a little machine, using 100 bits; he makes a Geiger counter, using 500 bits; and he then uses 1 bit to decide that the counter shall, rather than shall not, feed into the machine. He then comes to us and claims that he has built a machine that (*a*) was designed, so far as *he* was concerned, with 601 bits, and that (*b*) can emit far more than this quantity. Can we deny his claim?

He has shown us how a human being, limited by his finite intelligence so that he cannot contribute more than 1000 bits to the design, can none the less produce a machine that will emit more than this quantity. The point is that the decision to use, or not to use, a source of information—or to select one source from several—requires only a bit or two, while the source itself then contributes a quantity which may be much greater.

He has, in fact, devised an 'information-amplifier.' For what is an 'amplifier?' A power-amplifier is a device that, if given a small amount of power, will emit a larger amount. A sound-amplifier, if subjected to a small sound, will emit a larger sound. A 'money-amplifier' would obviously be a device that, if given a small amount

[1] The author is indebted to Professor Wiener for this apt adjective.

[2] Units of information, not parts.

of money, would emit a larger amount of money. In all cases the output is of the same quality as the input, but the input is not used to provide part of the output: it is consumed in controlling the flow of 'material' from a copious source, or large reservoir, of the same 'material'; and it is the reservoir that provides the output. Exactly the same method was used by the inventor: his 601 bits were used, not to provide part of the output but to control the flow of fresh information from a much larger source.

At this point the critic may well object that such information, being unorganised and chaotic, is useless. To this I would reply that chaotic information is by no means useless, but is, in fact, perfectly usable (as evolution has shown by its use of mutations) *provided that the machine has been designed to make the necessary selection.*

The main lines of the machine that is to go past the limit set by Descartes' dictum is now clear. The designer should make it selective, like the 'magnet-machine' considered earlier, and he should couple it to an abundant source of information such as a Geiger counter or a device using thermal noise. The amount of design needed to specify the machine's construction, and the amount of information that it can emit as output, will then be dissociated, and the amount that can be emitted is no longer limited to being equal to, or less than, the amount used in the design. If P bits went to its construction, and if it selects something as rare as 1 in Q, then if Q is greater than 2^P the designer can claim that the machine is showing more design than he has put into it.

6 *The Homeostat*

The argument may be made clearer by now turning to the homeostat, which was built precisely to admit and to use unorganised information.[1]

First, what is the homeostat? Here we need its abstract functional, rather than its practical, details.

In a sense it is a machine within a machine, so its description is best taken in two stages. The primary machine consists of four Units, all of which act on each other, and each of which bears on its top a needle whose deviation from the central position is the focus of

[1] W. Ross Ashby, in *Electronic Engineering*, 1948, **20**, 379–383. (A full account of the machine and of its principles in relation to the living brain, especially with regard to the origin of adaptive behaviour, is given in my *Design for a Brain*, to be published shortly by Chapman and Hall, London.)

interest. The four deviations (x_1, x_2, x_3, x_4) depend on sixteen parameters a_{ij} in accordance with the equations (of the same form as (3) above)

$$\left.\begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 \\ &\cdots\cdots\cdots \\ \frac{dx_4}{dt} &= a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 \end{aligned}\right\} \quad \ldots \quad (5)$$

Its only resting state is at the origin, where it can be in either stable or unstable equilibrium.

If the latent roots of the matrix $[a_{ij}]$ have real parts that are all negative, the system will be stable. Its needles will then move to the central position and, if displaced, the Units will so interact that a co-ordinated movement brings the needles back to the centre. If one or more of the real parts is positive, the system will be unstable, 'vicious circles' of action will develop, and the needles will move away from the centre. To this extent the machine is simply an analogue computor, for if the values of the parameters a_{ij} are given, its behaviour provides integrals of the equations (5).

It has, however, further components which influence this first activity. If any one of the x's goes outside certain bounds (a deviation of more than about $\pm 45°$) the extra components act and change the a's in the corresponding row to a new set of values that are taken at random from a distribution that lies evenly between -9 and $+9$. The actual values came originally from a table of random numbers.[1] So long as the needle stays outside the bounds, so long do the a's keep changing, at about one-second intervals. As soon as the needle returns to within the bounds, the a's stop changing and retain their latest set of values.

When set working it behaves in the following way. If stable, the needles come to the centre; it will then demonstrate, if disturbed, the co-ordinated response typical of a stable system. If, however, it is unstable, the needles diverge, and the a's undergo substitution after substitution. The first stable combination of a's to arrive is retained. Whether a particular combination will be stable depends partly on what other conditions (the 'problem') have been imposed on the mechanical and other parts of the system (for commutators can reverse

[1] R. A. Fisher and F. Yates, *Statistical Tables*, Edinburgh, 1938

the normal polarities of connection, a needle or two can be locked, needles can be tied together, and so on, making the equations (5) more complicated than is shown). Its behaviour can therefore be described by saying that it hunts for, and retains, a matrix that is stable in the imposed conditions. If the conditions are changed, so that the matrix is no longer stable in these new conditions, it will at once hunt for, and eventually retain, a new matrix that gives stability in the new conditions.

Its process is clearly similar to that occurring in evolution. There the rules are: test the organism against the environment; if the organism is unfit remove it; replace it by a new organism that differs from it in some random way. In the homeostat the rules are: test the matrix for stability in the imposed conditions; if it is unstable remove it; replace it by a new matrix with random elements. In both, the new material varies merely randomly from the old.

The homeostat was built before Shannon's theory of information had been published. It is instructive now to apply the theory to the homeostat's design. Let us consider the original model. There are four uniselectors (stepping-switches) and each can take, independently, one of twenty-five positions. The amount of information in its output is thus $4 \log_2 25$, i.e. 18·6 bits. The amount that went to its design can be very roughly estimated at 200–800 bits. It seems, therefore, that though the homeostat, in principle, can claim to produce more design than was put into it by the designer, the original model cannot so claim—its reservoir of new information is too quickly exhausted. Its amplification-factor is almost ludicrously below unity.

This deduction is somewhat sobering, but from it come two reflections. We can see how powerful is Shannon's theory in its quantitative grasp of these questions, and we can also notice, what the evolutionary process itself suggests, that a very great deal of selection may be necessary if Descartes' limit is actually to be passed.

7 *Conclusion*

It seems now to be clear that while Descartes' dictum is true if by 'designer' we mean 'every detail that contributes to the machine's performance,' it is not true if by 'designer' we mean 'the man who specifies its construction.' To defeat the limit implied by the dictum, the designer must include, as one of his specifications, 'admit other

information.' Stated thus the method may seem to be mere trickery, on a par with that of the boy who, given three wishes, made one of them a request for three more wishes. Trickery or not, however, it offers the designer a practical method for overcoming the limitations of his own powers of design.

Once the method has been identified, we can see that it is not unreasonable. We can see, for instance, that the more minutely we design a machine to play chess just as we want it to—admitting no other information—the more certain it is to play just *our* sort of chess, with all our faults and wrongly-conceived strategies. We are, in fact, in exactly the same position as the father, a keen but mediocre chess-player, who wants his son to become world champion. It is true that the father should teach the child, but he must not teach his son every reply in detail lest he limit the son's play to being the merest replica of the father's. If the father really wants the son to beat him he must sooner or later stop telling the son what to do and must send him out into the world to be subjected to all sorts of unselected experiences. The understanding father will not try to teach his son all chess but will try to teach him how to profit by future experiences.

Designing a machine has much in common with teaching a child, for in each case the almost infinite possibilities have to be reduced to a selection. Were Descartes' dictum to be stated in the form 'no child can know more than he has been taught' we would at once see the equivocation, for the teacher can not only teach the child facts but can also teach him how to use the 'free' information in the world, and thus how to surpass his teacher. If we wish to build a machine that can beat us at chess, or to build a 'real' brain, we must follow the same method; we must aim in design, not a machine that will play chess, but at a machine that can make trials, and select. The homeostat was intended to be a first step in this direction.

8 *Summary*

The quantitative aspects of the question 'how much of a machine's behaviour is attributable to the designer?' can be treated accurately and meaningfully by the methods of information theory. Its application leads to the deductions:

It is quite possible for a mechanical chess-player to outplay the man who designed it.

For this to be possible, the designer must construct the machine so that it can receive and use information not provided by him in detail.

The extra information need not be organised or intelligible—that given by thermal noise may be sufficient.

(An examination of the homeostat illustrates the method in detail.)

Department of Research
Barnwood House
Gloucester

WHAT IS AN INTELLIGENT MACHINE?

Summary

From the "intelligent" processes we must first split off those that are peculiar to the living brain, but only because they are not commonly met with elsewhere. These processes are of interest but are neither intelligent nor stupid, neither good nor bad.

The "intelligent" processes *par excellence* are the goal-seeking – those tha show high power of appropriate selection. Man and computer show their power alike, by appropriate selection. Both are bounded by the fact that appropriate selection (to a degree better than chance) can be achieved only as a consequence of information received and processed.

Machines can be made as intelligent as we please, but both they and man are bounded by the fact that their intelligence cannot exceed their powers of receiving and processing information.

1. Introduction

I am very pleased to have the privilege of addressing this conference today because I believe the time has come when we should notice a turning point in our views on the nature of brain and of brain-like mechanisms. The 1950's were largely a decade of ferment and progress. The 1960's, I believe, will be a decade of consolidation, of the establishment of a firm framework of ideas within which the whole science of brain-like machines will move for quite a number of years to come.

The point is that until recently we all tended to assume that the capacities of the brain, especially of the human brain, were unlimited. We felt that if a man were clever enough he could do anything – the genius could solve any problem. I say that that belief must go. It is obstructing progress. In the 60's it will be recognized as being as ignorant and superstitious as the belief of the small boy who thinks that his big brother can lift anything. Today, we know what the word "brain-like" means, and we know what are a brain's limitations. We know, too, that these limitations are *exactly* the same for the human brain and for the machine because they are the limitations inherent in any system that behaves in an orderly and law-abiding way. The system that passes these limitations gets its results by pure magic. Before I go on, however, to treat these matters in more detail, I would like to discuss some minor matters so that we can get them out of the way.

Brain-like processes can be clearly divided into two classes, according to whether the process is goal-seeking or not. It is the goal-seeking processes that

are *par excellence* the intelligent ones, whether they occur in a machine or a brain. But there are also a number of processes that occur in the brain that are not goal-seeking. Let me deal with them first.

The living brain has, of course, a great number of interesting properties that are of interest simply because they do not occur commonly elsewhere. The brain, for instance, has some unique biochemical processes; and it has interesting electromechanical processes. Of special interest to the computing engineer are the special network properties that it has developed, and the stochastic properties that it has developed for special purposes. The chief point about these non-goal-seeking processes is that they are neither good nor bad in themselves – they are simply processes that the laws of nature provide – like oxidation – and the brain, under the guidance of natural selection and evolution, develops or suppresses them in accordance with whether they are useful or harmful. They are brain-like only in the sense that they are seldom seen outside the brain. They can all be simulated on computers because they are straightforward natural processes.

In considering these brain-like processes, we should remember that the computer is actually superior to the brain, because the computer can be made to behave as if it were totally devoid of any operational structure. As a result, the computer can, in principle, carry out any well defined process. The living brain, however, has been so molded by five billion years of evolution that it is now very highly specialized to match the needs of this terrestrial environment. This environment, we are beginning to realize, is nothing like so general as one is apt to think it. Its distribution in space, with a three-dimensional Euclidian metric, the extraordinary commonness of continuity in it, its tendency for effects to be much localized, and the tendency for the same properties to turn up again and again in different places, all these features are highly characteristic of the terrestrial environment. For them to occur in the computer, they would have to be programmed in with great labor. Because the living brain has faced this very special environment for so long, the brain has become equally specialized in its operational methods. As a result, the brain, far from being a remarkably flexible mechanism, is now appreciated as a system of remarkable inflexibility.

Of these brain-like processes, other than the goal-seeking, I wish to say little here, but before I leave the topic I would like to say that I think the most promising line of research at the present time is the study of systems with large numbers of equilibria. Our knowledge of such systems is extraordinarily small. A great deal is known about the statistical mechanics of large physical (i.e. Newtonian) systems, but these usually have very few states of equilibrium. The simplest example of a system with a really large number of states of equilibrium is a dish of sand, in which the particles will rest in a great number of different configurations. But the only activity of this system is the tiny movement of the grain of sand as it moves from a nonequilibrium to an equilibrium – a movement too small to be interesting. What we need to know more about is the system that has a very great number of states of equilibrium and, sufficient dynamism so that its trajectories, before it reaches a state of equilibrium, are sufficiently long and complex to be interesting. I have shown elsewhere[2] how such systems tend to show some of the elementary properties of living organisms, and I have little doubt that much more remains to be discovered. Here, of course, we recognize that a system with thresholds, such as is the nervous system is just such a system with a great number of states of equilibrium. It seems to me to be most incredible that having known for 50 years that the nervous system works largely on threshold, we should, in the 1960's, know practically nothing about how such a system tends to behave when the system is large enough for it to show something of its statistical mechanics.

So much then, for those properties that the brain possesses, but that are essentially ordinary as natural processes. I now come to the other brain processes – those that are universally recognized as somewhat extraordinary. They are – the goal-seeking.

2. What Is Intelligent"?

Until a few years ago there could have been a considerable dispute about what was meant by an "intelligent machine", but that time is past. The position was clarified some years ago, and has been known long enough for any refutation to have come forward. No refutation has been offered. Not a single clear counter-example has been given in the last 10 years to show that an intelligent system is anything other than one which achieves appropriate selection. This is the touchstone of intelligence. According to this view, intelligent is as intelligent does.

Let me give some examples to make clearer what I mean. If a man plays chess, we need not judge his powers by listening to his boasting – we simply observe whether the moves he makes are very highly selected out of the totality of legal moves, being selected from just those few moves that bring him rapidly nearer the win.

Again, the good workshop manager is one who, in spite of all the confusions and difficulties of the day, issues such carefully selected instructions as will steer all the work through by the end of the day. Again, signal men show their intelligence by selecting just those patterns of operations in their box that gives, over long intervals of busy traffic, an accident number of zero. And in the

so-called intelligence tests, which do test something of what we mean by intelligence, the operational criterion is simply "did the candidate select the right answers?"

Thus an intelligent machine can be defined as a system that utilizes information, and processes it with high efficiency, so as to achieve a high intensity of appropriate selection. If it is to show *really* high intelligence, it must process a really large quantity of information, and the efficiency should be really high.

In biological processes, appropriate selection and intelligence is shown essentially by regulation; the living organism, when it acts "intelligently," acts so as to keep itself alive. It acts, in other words, so as to keep the essential variables on which its existence depends within physiological limits. This is a straightforward act of appropriate selection, and the animals, as they ascend the scale of intelligence, show their ascent precisely by their power of regulating their environment in spite of greater ranges of stresses coming to them. In man, the primary goals are what evolution and natural selection have built into him. The other goals are all secondary, developed either as species characteristics or by learning.

This approach to the nature of intelligence gives us an angle on the subject quite different from the older philosophers', and one which at a stroke ties it firmly to the modern theory of information. For "regulation" simply means that in spite of many threatened deviations from the optimum, the organism so behaves that the deviation does not occur; that is to say, the correct form is maintained. This achieving of a correct final form, repeatedly in spite of a stream of disturbances, is clearly homologous with the correction of noise by a correction channel. The noise threatens to drive the form or message from its desired shape and the correction channel so acts as to bring it back to the true form. A natural measure of the degree of intelligence can thus be given in the terms of Shannon's theory of communication, and with it not merely a measure but a complete grasp of the logic of the situation and of what is implied.

3. The Limit to Intelligence

As soon as we recognize that an intelligent system, whether living or mechanical, is simply one that behaves in an intelligent way, we appreciate that the test of intelligence is the power of appropriate selection. All intelligent actions are actions of appropriate selection. As a result every intelligent system is subject to the following postulate:

> Any system that achieves appropriate selection (to a degree better than chance) does so as a consequence of information received.

One would imagine this postulate to be completely obvious were it not for the fact that many discussions about the powers of living brains subtly and tacitly deny it. Yet what would happen if the postulate did not hold? We would have the case of the examination candidate who starts to give the appropriate answer before the particular question has been given! We would have the case of the man who submits an accurate insurance claim for damage by fire before the fire has broken out! We would have the case of the machine put on the market for which the claim is made that it is now so intelligent that it will start to give the answer before the program tape is run in!

Science knows nothing of these things. What will happen in the future we can't say; but it is quite clear that in the middle of the 20th century we must reject such possibilities and proceed on the assumption that they do not occur.

The moment we say such events do not occur we are implicitly saying that all systems whether human or mechanical are subject to this postulate – they can achieve appropriate selection only if they receive and process the appropriate amount of information.

This point of view at once brings them under a quantitative limitation. For appropriate selection is fundamentally homologous, as I said, with the correction of noise, and therefore the amount of correction that can be applied is subject to Shannon's tenth theorem[3]. Though the theorem has a somewhat different aim, it says that if a certain quantity of error is to be removed from the final form (that is to say, a certain degree of appropriate selection is to be made), then at least that quantity of information must be carried along the correction channel. When a human being undertakes such activities of correction, or of regulation, or of appropriate selection, he is acting as the correction channel, and he cannot achieve this appropriate selection unless he receives and transmits the necessary *quantity* of information.

The same point can be made in a simpler and more primitive form, as I have done in the law of requisite variety[1]; which shows that in the most obvious and common sense way the processing of the necessary quantity of information must be done if appropriate selection is to be achieved by law, and not by mere magic.

Today then, we are in the position of being able to say of the human brain that it must work in one of two ways. Either it works subject to this postulate, in which case it achieves appropriate selection because it has received and processed the necessary amount of information, or it is behaving in an entirely magical way, producing correct effects without corresponding causes.

I do not say that it is impossible that the human brain should sometimes do wonderful things – I believe that the universe is still full of surprises;

but what I do say is that those who maintain that the human brain is not subject to my postulate must accept the consequences of the alternative and must declare that the human brain sometimes achieves appropriate selection without receiving the necessary information. And it is obviously desirable that they should produce evidence to show that this remarkable event does actually occur. Until such evidence is produced, the postulate must stand.

It may perhaps be of interest to turn aside for a moment to glance at the reasons that may have led us to misunderstand the nature of human intelligence and cleverness. The point seems to be, as we can now see with the clearer quantitative grasp that we have today, that we tended grossly to misestimate the quantities of information that were used by computers and by people. When we program a computer, we have to write down every detail of the supplied information, and we are acutely aware of the quantity of information that must be made available to it. As a result, we tend to think that the quantity of information is extremely large in fact, on any comparable scale of measurement it is quite small. The human mathematician, however, who solves a problem in three-dimensional geometry for instance, may do it very quickly and easily, and he may think that the amount of information that he has used is quite small. In fact, it is very large; and the measure of its largeness is precisely the amount of programming that would have to go into the computer in order to enable the computer to carry through the same process and to arrive at the same answer. The point is, of course, that when it comes to things like three-dimensional geometry, the human being has within himself an enormous quantity of information obtained by a form of preprogramming. Before he picked up his pencil, he already had behind him many years of childhood, in which he moved his arms and legs in three-dimensional space until he had learned a great deal about the intricacies of its metric. Then he spent years at school, learning formal Euclidian methods. He has done carpentry, and has learned how to make simple boxes and three-dimensional furniture. And behind him is five billion years of evolutionary molding all occurring in three-dimensional space; because it induced the survival of those organisms with an organization suited to three-dimensional space rather than to any other of the metrics that the cerebral cortex could hold, evolution has provided him with a cerebral organization that must be peculiarly suited to the manipulation of three-dimensional entities. So when a mathematician solves a problem in three-dimensional geometry, he tends grossly to underestimate the amount of information involved in the process. When he does it in a computer, he tends grossly to overestimate it. What I am saying is that if the measure is applied to both on a similar basis it will be found that each, computer and living brain, can achieve appropriate selection pricesely so far as it is allowed to by the quantity of information that it has received and processed.

Because of this hidden preprogramming of every human being, nothing is easier than for him to achieve results with extreme quickness, provided the question falls within his specialized range. But this is no more miraculous than the power of any other machine that is heavily preprogrammed to be quick. Most of the examples commonly given purporting to show some peculiar facility possessed by human beings are of problems in which human beings are peculiarly experienced, either personally or by the hereditary equipment that has come to them. Take for instance the playing of chess. The first thing that has to be explained to the boy of 10 is that the rows, columns, and diagonals are significant. Because he is a human boy aged 10, long experienced in two-dimensional Euclidean geometry, we can indicate rows, columns, and diagonals to him by merely flicking a finger at the board. The computer, however, being stripped down to an absolute zero of metrical properties, has to have the whole metric of the chessboard explained to it in detail, because it would just as readily play on a board with a metric that would seem crazy and quite impossibly difficult to a human being. Thus the fact that a human being is especially good at human problems is no more remarkable than that the digital computer is especially good at problems involving powers of two, or that the analog computer is especially good at handling continuous functions. I say that whenever a human being is found to be pecularly good at a particular class of problems he will always be found to have had substantial preprogramming in those problems. The alternative is that he is getting the answers by magic.

4. What is a "Genius"?

We can now consider briefly the question of the so-called "genius", and the question of his nature. There are two gross fallacies that infest our thinking about the genius.

The first is that after many scientists have tried to solve a problem, we imagine that the one who solves it must have some peculiar power. This is about as reasonable as letting 1024 people predict how a coin will fall 10 times in succession, and then, when one person gets all 10 right, trying to find the explanation of his phenomenal powers of prediction!

Isaac Newton, for instance, recorded when he was quite young that he always thought of everything as flowing into everything else; this was just his natural way of thinking and very congenial to him. He used this way of thinking on almost everything. Is it surprising that this was the man, who, at a time when the calculus was on the verge of being discovered, was actually the man who got it first? Compare him with, say, Planck, at the beginning of this

century, when science was crying out for a man who could think of everything as going in small, discrete jumps. Had Newton been unlucky enough to have been born in 1900 he would have found himself peculiarly handicapped at the time when the quantum theory was just being formulated. Clearly, the concept of a genius is apt to arise because after a number of workers have tried various ways of solving a problem, none of them knowing beforehand which is the right way, and one of them succeeds, *we* come along, pick this person out, and say he is remarkable. Now, part of the selection involved here was not made by that person; the selection is made by us, who pick out this person because of his performance. This very common mistake in statistical logic must be responsible for a substantial amount of our allocation of the title of "genius".

The second fallacy is the idea that the genius can go, as it were, straight to the answer without doing the work. In actual fact much of the work consists of making trials, which is, of course, a powerful way of gaining information. Many of the recognized geniuses are people who, by thinking about the subject day and night, are making trials of new combinations and new ways in great numbers. Take for instance the mathematician Gauss, who is doubtless generally accepted as an excellent example of a genius. Hear his own words about how he achieved a certain result, in a letter of Olbers: "Perhaps you remember my complaints about a theorem which had defied all my attempts. This lack has spoiled for me everything else that I found, and for four years a week has seldom passed when I would not have made one or another vain attempt to solve this problem – recently, very lively again. But all brooding, all searching, has been in vain." Then he adds, "Finally I succeeded a few days ago." And then he adds, "Nobody will have any idea of the long squeeze in which it placed me when I someday lecture on the topic." Undoubtedly, one of the reasons why a person is a genius is that he pays the price for it by sheer hard work. He processes the necessarily large quantity of information.

If the human brain is especially clever and slick at those problems for which it has been preprogrammed, we should find, of course, that it is peculiarly stupid and slow at those problems that are subtly contrary to the preprogramming. As far as I know, very little exploration has been done in this direction. We are not proud of our mistakes and it is only quite recently, almost within my lifetime, that psychologists have seriously paid attention to the defects of the ordinary human being instead of simply trying to exaggerate his abilities. But we do know that there are a certain number of events that show how he can be peculiarly handicapped. It is, of course, obvious that any species that tries to discuss its own sexual habits will always have difficulty, simply because the mixture of the real and the symbolic will always tend to create a confusion. We need not be surprised that we can discuss the sexual behavior of the stickleback with the utmost precision and objectivity, and then fall into a hopeless

muddle when we try to talk about the sexual habits of the young man and woman of today. Other examples, are the well known difficulties that occur when we try to make simultaneous hand and foot movements in a way that do not match the age-old needs of gravitation and locomotion. Again, there are the demonstrations produced by Ames which show how strongly we are impelled to see relationships simply because we are preprogrammed to see them. There is one example, for instance, where one looks through a hole into a box and one sees apparently a toy chair suspended in mid-air. Then one looks through a window in the side and one realizes that there are really a number of pieces scattered throughout the space but so arranged and strung on wires that when seen from one point they present the perspective of a chair. Then the observer, having seen beyond all question that the pieces are widely separated, goes back and looks through the hole again at the appearance. He *cannot* prevent himself from seeing one chair in one place.

5. There Is No "Real" Intelligence

Is there, then, no such thing as "real" intelligence? What I am saying is that if by "real" one means the intelligence that can perform great feats of appropriate selection without prior reception and processing of the equivalent quantity of information; then such "real" intelligence does not exist. It is a myth. It has come into existence in the same way that the idea of "real" magic comes to a child who sees conjuring tricks. At first the child believes in "real" magic. Later, after he has found how tricks are done he no longer believes in transcendental "real" magic; he replaces the myth by genuine knowledge of the processes of actual conjuring.

6. Consequences

What now are the consequences of this point of view? Especially for the computing engineer?

The first fact is that in talking about "intelligence", whether of the living brain or of the machine, we must give up talking about two sorts of intelligence. There is only one sort of intelligence. It is shown in essentially the same way whether the brain is living or mechanical. It shows itself by appropriate selection. It always implies the same underlying activity – that information in the required quantity is taken in (either immediately before the problem is given or at some time earlier as preprogramming) and that this quantity of information is processed with sufficient efficiency so that the total quantity does not fall below the point where it is no longer sufficient to allow the appropriate selection. The living brain has had only one problem throughout evolution: how to get the

necessary information in, and how to process it with reasonable efficiency. The problem of today's computing engineer is exactly the same. From this point of view, the computing engineer should stop asking "How can I make an intelligent machine?" because he is, in fact, doing it at this very moment, and has been doing it for the last 20 years. He should stop being overawed by the so-called geniuses, and he should realize that the so-called genius is simply a rather extreme example of the system towards which he is working steadily. He will doubtless soon develop machines other than the large digital, but these will simply be intelligent in other ways. The contrast here is not between digital and analogue, or germanium vs. protein, but between the true intelligence that processes information in due quantities, and the merely mythical intelligence that human beings have sometimes supposed themselves to possess.

A second application of the basic postulate is that it will provide guidance in a host of different processes. For one must realize that the rule about appropriate selection applies not merely to the final goal, but to all the sub-goals that have to be found on the way to it, and to all the qualifying goals that may be set up. Thus, if the goal is a program to play good chess, the programmer may soon add a subsidiary goal: – the program is to be achieved in the shortest time. Now this "shortest time" is itself a goal, and its achievement demands appropriate selection (among the various ways that consume various times). Thus this demand itself can be met only by processing of the relevant information in sufficient quantity. If the information (about the relative quicknesses) does not exist in 1960, then the appropriate selection ***cannot be made.*** If the goal (of the quickest way) is still desirable, there is no way but that information must be collected. This means that there is no other way than that the program writer should try a tape, see how long it takes, try another tape, see how long it takes, and either by trial and error (another name for "experiment") or in any other way available to him, get the information about which way is the quickest.

There are a host of these subsidiary questions usually coming up during any real appropriate selection, and they can be extremely troublesome. It is a part of what I am saying that the basic postulate will apply to all of these subsidiary questions.

Any attempt to achieve a major goal usually implies the achieving of many minor or qualifying goals. The basic postulate, and the law of requisite variety cover them all. For example, to conduct a search quickly (with speed as the qualifying goal) may require repeated dichotomization. Then "***how*** to dichotomize" becomes the object of a search. Finding how is an act of appropriate selection; it is again subject to the postulate.

There is again the problem of where to bring back corrective feedbacks: Should the correction be fed back to this point or to that? To know which is to make an act of appropriate selection, and this again can be done only insofar as information exists. If it does not exist either a simple random decision must be made or further information must be obtained, either systematically or unsystematically by trial and error; and so on, again and again.

To sum up. What is often referred to with bated breath as "real" intelligence, is a myth. The human being saves himself from being wholly foolish by having a great deal of information, as preprogramming, derived from millions of years of evolution on this earth, and by his personal experience over decades. Give him a problem in this range and he is really slick. This ***is*** his ***real*** intelligence. And any machine equally preprogrammed has an equal amount of ***real*** intelligence.

But this intelligence, whether of man or machine, is absolutely bounded. And what we can build into our machines is similarly bounded. The amount of intelligence we can get into a machine is absolutely bounded by the quantity of information that is put into it. We can get out of a machine as much intelligence as we like, if and only if we insure that at least the corresponding quantity of information gets into it.

The thought of this ultimate limitation is sobering, but the situation here today is not unlike that of power engineering a century ago.

At that time, so many powerful machines were being developed that many engineers took it for granted that the perpetual motion machine would soon be discovered. Then gradually emerged the idea that energy could not be created; and I have little doubt that this idea was seriously disappointing to many of the engineers of the time. They regarded it simply as a limitation.

Nevertheless, we now know that those engineers who accepted the limitation were in fact more realistic than those who went on hoping that perpetual motion would be possible. In the long run, the engineers who accepted it built better engines than those who went on struggling after perpetual motion. I suggest here today the position in computing is similar. If we accept the limitation – that appropriate selection can be achieved only to the degree that information is received and processed – and if we accept that this limitation holds absolutely over all brains, human and mechanical, our work, though less intoxicating, will in fact be more realistic. Those who build intelligent machines on this basis will outdistance those who want to build them on the old and superstitious basis that the human brain can do anything.

References

1 ASHBY, W. ROSS. *An Introduction to Cybernetics.* John Wiley & Sons, New York, 1956.

2 ASHBY, W. Ross. "The Mechanicsm of Habituation." In N.P.L. Symposium on the Mechanization of Thought Processes. Her Majesty's Stationery Office, London, 1959.

3 SHANNON, C.E. and WEAVER, W. *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, 1949.

VII.

OTHER TOPICS AND OVERVIEWS

OTHER TOPICS AND OVERVIEWS

INTRODUCTION

In this chapter are included several articles which do not fit comfortably in preceeding chapters. "The Relativity of Meaning" succinctly points out that meaning is not an intrinsic property of a message. In "Induction, Prediction, and Decision-Making..." Ashby argues against the possibility of absolute truth, claiming that neither machine nor human can obtain infinite certainty (absolute truth) by processing a finite amount of information, and thus all truth is of finite range and reliability. The reader may dispute some of Ashby's stronger claims in this provocative article - yet they are consistent with his view, expressed in several previous chapters, that all intelligent processes such as those named in the title of the article are bounded in the end by the quantities of information involved and thus ultimately are governed by the Law of Requisite Variety.

In "Cybernetics Today..." Ashby assesses some of the past accomplishments of cybernetics research and gives an indication of future directions from the perspective of 1961, when the lecture was given. Here appears his view that past cybernetics research had removed the mystery from brain-like mechanisms and that much future work on them would be devoted to the crucial but unglamorous work of raising the efficiency of information processes. Here also are his only printed comments on what is required in cybernetics education, from his experience at the University of Illinois at Urbana, and also some important research problems whose solutions, incidentally, are yet to be found.

Ashby was interested in model-building and was optimistic about the power of computers to allow construction of and deduction from models far too complex for humans to fully comprehend. In "Analysis of the System to be modeled" he considered the basic function of models as information-losing representations of reality which conserve only crucial features of the system under consideration, and in "Mathematical Models ..." he reviews attempts (through 1965) to model various complex phenomena, from pseudo-neuronal interactions to the intellectual activity in the then-new field of artificial intelligence. Throughout both of these papers runs his incessant concern with the quantities of information involved. Ashby's vision, which is currently being implemented in a minor way (for example in computer systems for medical diagnosis) was that computers would serve as archives for and operators upon collections of knowledge too large and complex to be held in the brain of a single human being, and thus would allow modeling and intelligence beyond the capabilities of individuals.

The review article, "The Contribution of Information Theory..." is of such broad scope that it touches in one way or another on the materials in most earlier chapters. It is not in fact confined to information theory but ranges over such topics as memory, artificial intelligence, originality or creativity in machines, self-organization, automata theory, consciousness, many-dimensional information theory, random networks, Bremerman's Limit, the decomposition of complex tasks, and more. Being a review of the state of the art it is inevitably dated in certain respects, but much of it seems as current now as it was in 1968.

One of the things Ashby enjoyed most, in his classes and papers, was demolishing stereotypes and uprooting commonly held assumptions. In "The Brain of Yesterday and Today" he discusses several "common-sense beliefs" and shows how each is inconsistent with modern knowledge. The handout "How Wrong Can You Get?" in the subsequent chapter is a listing of many more such beliefs, upon which the reader may test his or her wits.

THE RELATIVITY OF 'MEANING'

It is now widely accepted that the information content of a message is not intrinsic to the message; it depends on the set of messages from which the message comes. Not so generally accepted is the postulate that the meaning of a message also depends on the set that the message comes from. The following example, though imaginary and seeming flippant, shows decisively that the presence of another message may grossly affect the meaning of what is transmitted, though the other message is in fact not sent.

The story is told that Mr. and Mrs. A, while on holiday, decided to send a greeting back to the wife's mother. The telegraph company offered a set of standardized messages (actually only two), which were:

Message 1: "How we wish you were here!"

Message 2: "The weather is fine."

In fact they sent message 1 to express a simple and friendly meaning.

Suppose, however, that the company had also offered:

Message 3: "Do come and join us!"

In its presence, the message "How we wish you were here" becomes merely ironic; for to send it is equivalent to a refusal to send the invitation.

Thus, in this example at least, the 'meaning' of message 1 is not intrinsic to the actual words sent: it is a function of the whole set. Passing from the message to the meaning thus resembles such functional operations as taking the average, or the maximal value. According to this view, the 'meaning', on the reception of message i from the set $\{1, 2, \ldots, i, \ldots\}$ is not to be identified with the element i but with forming the i-th function over the set of messages. Applied to the brain (as a dynamic system the behaviour or activity of which forms some function of its input) this view suggests that we should relate the 'meaning' not to the message that comes to it as input but to the particular behaviour or output with which the brain responds.

W. ROSS ASHBY

Burden Neurological Institute
Bristol

III

*Induction, Prediction, and Decision-making in Cybernetic Systems**

By W. Ross Ashby

SUMMARY

The concepts of induction and deduction, as they came to us from the Greeks, were associated with the assumption that they might give a truth that would hold universally without any limit or condition. This assumption, implying that the process ends with an *infinite* quantity of information, is rejected, for no human or mechanical system can process more than a finite quantity.

When the processes of induction and deduction are re-examined in the forms in which they are *actually* used in human thought processes, they are found to follow the same basic methods, and to be subject to the same basic laws, as when they are used in machines.

A NEW START

My first instinct, on approaching the problem of induction, is to go cautiously. The problem is an old one, coming to us from an age that saw the world very differently from the way we see it today. And the way we see it today differs profoundly even from the way it was seen in my adolescence. Before rushing to answer the questions that the Greeks posed, I would prefer today to re-examine the questions, asking whether each question is properly askable.

One aspect at once puts the cyberneticist on the alert. Cybernetics and information theory are alike in being practical, pragmatic, empirical. They think all the time of knowledge as something that is won piece by piece, as something that grows. Statements are not true or false—their probabilities go toward one or zero, but these are the limits, seldom

*This work was supported by the U.S. Department of Health, Education & Welfare Grant No. G.M. 10718-01.

actually achieved. Thus the basic Greek assumption that there are absolute truths, that can be reached in a page or two of logical or mathematical demonstration, leaves the cyberneticist uneasy. He is not used to getting absolute certainty with so little work. I would therefore like to start again from first principles and to see if, perhaps, the Greeks have not unwittingly been misleading us for two millennia.

There comes a time when reconsideration becomes necessary, for however much we may admire the Greeks in some ways, we must admit that they were, in other ways, almost ludicrously ignorant. How long, one wonders, would the average Athenian, alive today, last if he were to start debating with a present-day undergraduate? Sooner or later he would betray that he thought that fire was a substance that came out of wood, or that the winds were blown by big people over the horizon, or that the brain was mucus waiting to come down the nose. We must not delude ourselves into thinking we know everything, but it is a simple fact that a vast amount of knowledge, some of it of profound importance, is available to us today, but was not available to those who first speculated about induction.

The chief reason, of course, for the Greeks' confidence in their philosophy was that they had discovered certain truths which seemed to them to be absolute. Few of us probably have not shared with them the excitement of seeing the demonstration that $\sqrt{2}$ cannot be a fraction, or that the angles inside a triangle add up to two right angles. They felt they had discovered, and demonstrated, that deduction could give absolute truths; so they naturally asked the further question: How may *in*duction lead us to Absolute Truth?

KNOWLEDGE IS FINITE

The cyberneticist, however, is already uneasy. Total certainty, over an *un*bounded universe, is hardly to be obtained except by the reception of an *infinite* quantity of information. The Greek man was no more able to receive an infinitely large quantity of information than modern man is. The communication made in the proof itself—two pages of written characters, say—certainly does not carry an infinite quantity. Are these truths really absolute?

The question is not just philosophic; it has immediate practical consequences, shown at once when we ask: Can a machine find similar

truths? To answer this question, let us take an example (equally typical, but more tractable from the machine's point of view). The Greeks discovered that

$$(a + b)^2 = a^2 + 2ab + b^2$$

In particular they knew that the middle coefficient must be a 2. This was typically the sort of truth that they could *prove*.

The computer can, of course, readily compute the two sides for many pairs of numbers, and could readily find that over all its trials no number other than 2 would give equality. If the computer has no better method than that of trial, never reaching absolute certainty, has Man a better method? I suggest he has not.

We know today, of course, that the Greek certainty in this matter (that the coefficient must be a 2) was simply rooted in ignorance. Knowing nothing of vectors, or groups, or matrices, or quaternions, or non-commutative rings, or algebras in general, he picked on one small portion of knowledge and thought it to be universal. Today we know that, far from the equation

$$(a + b)^2 = a^2 + 2ab + b^2$$

being a universal truth, it simply helps to define which of the many possible algebras we are talking about. Thus the statement is true if and only if we wish it to be so.

We can now see that the same is true of the statement that the three angles of a triangle add to 180 degrees. It is true of some spaces, false in others. To insist that it is true is merely to make clear that one must be thinking of a particular class of spaces.

With this view at a distance, we can now see the Greeks' knowledge in its proper size and place. Not realizing that they were unconsciously restricting themselves to a small portion of all knowledge, they thought when they had mastered their portion that they had mastered all there was.

At this point the modern theory of information becomes applicable, for its measures of information always work within a defined and bounded region. If, for example, the region has n cells, or possibilities, and all are equally probable, then one's knowledge is at its *absolute* zero. If information comes, and the probabilities shrink to a single cell, then the gain in information is $\log_2 n$. Here we have gained knowledge that is perfect, or complete, so far as it goes, in the sense that it cannot be increased without our going outside the original assumptions. Yet, in

spite of the completeness, the quantity $\log_2 n$ is not infinite. If, however, n, the range of possibilities, becomes infinite, complete knowledge could be obtained only by an infinite quantity of information. Cybernetics is not so unwise as to attempt the impossible. What it does generally is to attempt to obtain knowledge that is complete, so far as it goes, and that requires only a finite amount of information. It recognizes, in the same way, that a finite quantity of information can give completeness only over a restricted domain. It is more modest, and more realistic, then the Greeks; they discoursed of the real numbers, and referred to the "universe" of their discourse. We today talk of a far greater realm of numbers, but we do not call it the *universe* of discourse; we refer to the "region" of discourse. And we know today that whatever is true over a certain region may immediately become false if the region is enlarged.

If I may summarize what I have said so far, the Greeks made a fundamental error in thinking they had discovered truth in any absolute or unconditional sense. In fact, they processed a *finite* quantity of information, which is all that our humanity allows, and they arrived at truths of *finite* range and reliability. It is this finiteness that is the point of this paper. Whether the process is one of induction, deduction, prediction, or decision-making, the quantity of information involved in it is always finite. And this finiteness sets bounds to the possibilities.

INDUCTION

Once we reject the essentially superstitious belief that a mathematician's results are of infinite certainty, the whole theoretical structure of induction, deduction, prediction, and decision-making becomes much easier. For what remains, after the mythical part has been removed, is a perfectly straightforward flow of information, first into the subject during the process of induction, within the subject during deduction, and out from the subject at the decision-making. Let me take them in this order.

Induction, once one has removed the mythical element, becomes simply the collecting of information. What can be said, at any moment, is absolutely bounded by the quantity of information that has been taken in. If one seeks the magic theorem or algorithm that would enable one to say much after receiving little information, then such an algorithm does not exist. Many of the so-called "problems" of induction are nothing more than attempts to deduce more than the data permits; it is not

surprising that they are still unsolved. It is to be hoped that as we develop a better intuitive sense of "how much information" we will makes these mistakes less often.

In this connection it must be appreciated that it is only when the facts reach a certain minimal complexity that the concepts of information and induction come into existence. If, for instance, the question were: What can be deduced from the space-traveler's report that he landed on Mars, a white ball appeared on the ground beside him, and then he departed? the answer is: Very little, for the simple reason that such a restricted fact does not permit even the *start* of the concept of information. One is reminded here of how two points on a curve are necessary before the "tangent" becomes definable, three before we can speak of curvature, and four before the concept of torsion becomes meaningful. Attempts to apply the concept of torsion when only three or fewer points exist should not be made. Similarly, the concept of information (and through it of induction) does not exist until a total set of possibilities is properly defined.[1] This criterion will show many of the old so-called problems of induction as merely improper attempts to apply a concept to a set of data inadequate for the concept.

With this point of view, the question of how a person is able to commence the process of induction becomes simply the question of how a system, whether human or mechanical, can receive information. This question has been adequately treated elsewhere and need not be entered into here. The ways that human beings and machines take in information seem to be fundamentally identical. I know of no fact suggesting that there is any fundamental difference.

If it is felt that there must be some sort of special "receptivity" on the part of the system taking in information, I can only say that the evidence is the other way around. It has been shown that if a system is large and has many states of equilibrium (which is equivalent to saying that it is rich in memory), then putting it into an environment that shows law in the transitions it undergoes will result in the system going to a subset of its states such that the subset's transitions resemble (in a certain sense) the pattern of transitions followed by the environment. Thus pattern in the environment inevitably tends to "diffuse" into the system. Only the living brain has been large enough in the past to show the process with

[1] W. Ross Ashby, *An Introduction to Cybernetics*, New York, John Wiley & Sons, 1956.

any clarity but it is of very great generality. An elementary example of it occurs when one stirs a dish of sand with one's fingers; if the finger makes only circular motions one will find afterwards clear marks of circularity in the sand. This example is so well known as to be trivial; what is important is that a similar transfer of pattern can go on in complex mechanisms in which the circularities are complicated and twisted out of all recognition. So in the important case when the world around us shows its properties by its *transitions*, some of its abstract structure will make its way into us whether we like it or not. There is no problem about how to take the information in; it tends to force its way in. This is what "induction" means in cybernetics.

DEDUCTION

The general nature of deduction need not take us long. Once we have disposed of the myth of deduction giving us absolute truths, we can see what is fast becoming recognized by logicians, by mathematicians, and by programmers of computing machines, that "deduction" simply means the carrying through of some well-defined and consistently reproducible process. The emphasis is on the "well-defined." Should the process happen to be isomorphic with some other process, important in itself, then we are pleased. If we find the trajectories of the conic sections isomorphic with those of the planets we are delighted; but the delight is incidental; the deduction requires only that the operations within the conic sections be well defined and not self-contradictory.

From this point of view a digital computer deduces the numerical form of a Bessel function, an analogue computer deduces the flow past an aerofoil, a network of relays deduces the consequences of a set of propositions. Whether a student can deduce the roots of an equation depends simply on whether he can carry through a well-defined process. The *usefulness* of a deduction rests on totally independent criteria, and has nothing whatever to do with the process itself. The work of Newell, Shaw and Simon[2] has made clear that the development of deductive systems of any complexity is only a matter of time and labor. There is no essential mystery about deduction.

[2] A. Newell, J. C. Shaw, and H. A. Simon, *Report on a General Problem-Solving Program*, Internatl. Conf. Inform. Proc., Paris, UNESCO, 1960, pp. 256–64.

PREDICTION

The natural successor to induction and deduction is prediction. Its nature has, I think, been completely clarified by the cybernetic approach. Wiener put his finger on the fundamental axiom when he said that to predict the future is to perform an operation on the past. The essential point is that the agent in the act of prediction depends wholly on the actual past and not in the least on the actual future. When we say of a trained rat that it will not jump through a hole because it *will* receive a shock on the other side, we are guilty of gross confusion of thought and language. We can in fact *demonstrate* that the actual future is quite without effect in this situation, simply by arranging that on this occasion the actual future is to be that the rat is *not* to receive a shock. As everyone knows, the rat's behavior is unchanged. What it is reacting to is its past, which repeatedly contained the sequence

... jump, shock, ... , jump, shock, ... , jump, shock, ...

That living organisms do not react to the actual future, are not affected by it, do not receive information from it, is illustrated vividly when a driver is killed at night by running into an unlit obstacle, or when a soldier steps on a land mine. At ten seconds before the accident he is clearly guided in his actions by what he has seen up to that moment; the fact that he *will* be dead in a few seconds, the *actual* future, has no appreciable effect, though it must be about as powerful a potential stimulus as one can imagine.

That prediction is essentially an operation on the past is also shown very clearly if we reconsider the process of interpolation by Lagrange's method. Here the differences are eliminated, to give a formula in terms of the primary values themselves. Suppose that values, c, b, and a had been observed at times -2 (in the past), -1, and 0 (the present moment). And suppose that we have reason to believe, or are willing to assume, that the quantity is varying on a quadratic law. If we want to interpolate, to estimate the value in the past at

$$t = -\frac{1}{2},$$

we compute the function

$$-\frac{1}{8}c + \frac{3}{4}b + \frac{3}{8}a.$$

If we compute the function

$$\frac{3}{8}c - \frac{5}{4}b + \frac{15}{8}a$$

(as ordinary a function as the previous one), we are in fact *extra*polating it to give a prediction of what the value will be at

$$t = +\frac{1}{2}.$$

And the function

$$55c - 120b + 66a$$

predicts what it will be at $t = +10$.

The cybernetic view of "prediction" is now clear. It sees nothing peculiar in the process: simply an organism (or an anti-aircraft gun) reacting to its immediate and remote past in the way that every system does. The fact that the gun's mechanics make it point ahead of the plane is a consequence of its design (three years ago) and its input (in the last few seconds). It is the *observer* who, knowing a good deal of what happens under various conditions, describes the events in the most misleading terms of the gun "aiming where the plane will be." Though this way of looking at things is vivid and suggestive, it is basically false and therefore dangerous, especially when one attempts to develop a rigorous theory of behaving systems. The concept of "prediction" is meaningful only when the past shows some constraint, some redundancy. When this is so, the organism, or any reacting system, is able in some way to take advantage of the constraint. Always it uses the past to determine its present action, and the constraint then ensures that the present action shall have a better than chance probability of combining with the events in the real world to achieve the desired goal (Sommerhoff's "focal condition"[3]). The act of prediction thus is simply one of the many ways in which adaptation can show itself. In this sense, prediction can be said to occur in almost every movement we make. I step on this platform, predicting that it will support me. I speak into the microphone, predicting that it will carry my words. The predictions of the astronomer are not essentially different. (The question of the *success* of prediction I will take up in a moment.)

DECISION-MAKING

If the processes of induction, deduction, and prediction are all essentially ordinary—that is, physically realizable—the consideration of decision-making will not detain us long. Machines have no trouble decision making—they just act. Even if an automobile engine refuses to start, the refusal is as much its "decision" of what it is to do as would be a starting-up. Consequently what is usually sought is a "good" or "successful" decision rule.

[3] G. Sommerhoff, *Analytical Biology*, Oxford, Clarendon Press, 1950.

This is nothing other than the search for an "adapted" form of behavior, viewed from a slightly changed point of view. The formulations of either myself[4] or Sommerhoff help to show the essentials. There must first be a set of disturbances, dangers, threats—from what Sommerhoff calls the coenetic variable—setting up a real problem; and this set must be defined, for the solution depends upon it. There must also be defined the focal condition—what counts as "good," for in general this cannot be taken for granted. The problem then, in Sommerhoff's terms, is to find the decision-rule ρ such that for each situation, *its* output will so combine (by operation ψ) with the other factors τ of the situation that the end result is achievement of the focal condition G. The relation can be stated quite precisely. If the decision rule is ρ, mapping the coenetic values into ρ's output, and if τ is the mapping of the coenetic values into τ's output, and if ψ is the mapping of the just-mentioned couple of outputs into the set of all possible end results, with G the subset of "good" end results, then a necessary property of ρ if it is to be a good decision rule is

$$\rho \subset [\psi^{-1}(G)] \cdot \tau$$

where I use the notation and operations of the Bourbaki theory of sets. If to this we add the requirement that ρ must be a mapping, i.e., everywhere defined (for otherwise the empty set would satisfy the requirement of never issuing a bad decision by refusing to issue *any* decision), we have given the necessary and sufficient condition that ρ should be a good decision rule in the given conditions of τ, ψ, and G.

This relation specifies the decision rule. Certain of its features are worth notice. First, by being stated in general terms, it includes the special cases of the continuous and the linear, but is by no means restricted to them. So it can be applied without the least difficulty to the more general types of material that occur so frequently in biology and in computer processes. Second, it shows just what is essential, in that it depends on:

[4] W. Ross Ashby, *Design for a Brain*, New York, John Wiley & Sons, 2nd ed., 1960.

τ, the environmental factor
G, the defined goal
ψ, the way the rule and the environment interact, and on the domain of τ, the set of disturbances, eliminated in the composition

And third, it specifies ρ as some member of a class. This is as it should be; for only in the special cases that are continuous and linear will ρ be unique.

SELECTION

When the decision rule acts, it performs an act of appropriate selection, and it at once becomes subject to a rule that has been little used, yet is, I believe, fundamental in the whole theory of brainlike mechanisms. It can be most simply stated as a postulate:

Any system that achieves appropriate selection (to a degree better than chance) does so as a consequence of information received.

The postulate is easily defensible on common-sense grounds, for any system that blatantly violated it would at once be recognized as peculiar. We are suspicious of the examination candidate who gives the correct answer before the question has been put, and of the man who fills in an insurance claim correctly before the fire has broken out! But it can also be supported more technically; for "selection" and "error correction" are abstractly identical, and Shannon's tenth theorem[5] shows, under very general conditions, how the *amount* of appropriate selection is absolutely bounded by the *quantity* of information supplied.

From this point of view, a "good" decision rule is simply one that processes *efficiently* the information that comes to it.

The rules for decision will thus always be of the same form. The possible outcomes are initially many and mixed, good and bad. The set has to be cut down. The decision rule can only process such information as is supplied to it. If it is a good rule, it will use the information efficiently and will thus cut the set down so far as the *quantity* of information permits. *The amount of cutting down is absolutely bounded by the amount of information.* When the information, by its finite quantity, has been used up, no further selection can be justified. The process has reached its "field of ignorance"; any further selection within this field can only be arbitrary.

[5] C. E. Shannon, and W. Weaver, *The Mathematical Theory of Communication*, Urbana, University of Illinois Press, 1949.

Here I think we should chose our words with great care, for it is very easy to express the idea in words that are quite wrong—to our own confusion. Thus, one might say, if some selection *has* to be made within the field of ignorance, that "a random choice is as good as any." But this phrase is grossly misleading. By saying a random choice (by spinning a coin, say) *is* as good as any other, one is making a positive statement that can be checked by experiment. What one really means is that when the information is used up, one has no further *justification* for further selection; any further selection must be made arbitrarily.

When the decision rule has used up the information, and has reached the field of ignorance, no further progress has justification. Commonly one then tries to go on by collecting more information. So far as it is collected and used, the process simply remains again subject to the basic postulate. But there often occurs at this point a matter that has been profoundly misunderstood. I refer to the process of "trial and error." This process is of the greatest practical importance but has often been dismissed by the philosopher as trivial. Here it is essential that we appreciate that it has two aspects. On the one hand, it may be a mere grab at success—if it fails, its worth is nothing. Far more important, however, in the general theory of brainlike functioning, is that the process is capable of giving *new information*, the new information by which alone the field of ignorance may be reduced. This latter aspect is seen clearly when one attempts to find one's way out of a maze. When all other information is exhausted, there remains the fundamental method of using trials. Here the success or failure of each trial is not the point: what is important is that as each trial adds further information, the accumulating information allows further and further appropriate selection until eventually the field of ignorance is reduced to the size at which a solution is identified.

There has been some tendency for discussions of brain and computer to be based on the assumption that the computer can use only the method of trial and error, assumed to be futile, while the brain uses some method that is altogether superior. The fact is that when they are similarly situated with regard to the information available, both are equally subject to the postulate. And both may find, at a certain stage in a problem, such as maze-threading, that progress by efficient trial and error is fundamentally the only process possible.

MAN AND MACHINE

The survey I have given suggests that at no point can we find an instance, in the processes of induction, deduction, prediction, and decision-making, where the powers of Man differ essentially from those of an adequately complex and designed mechanism. Both act essentially as processors of information, both require information if they are to work successfully, and both are bounded by the postulate that they cannot achieve appropriate selection in excess of the information available to them.

Do we then jump directly to the simple deduction that Man is a machine? Here again cybernetics cuts right across the old simple ideas; for cybernetics does not admit the question as a proper one. It denies that one can classify the world's systems into, simply, the mechanical and the non-mechanical. *Every* system has mechanistic aspects, so far as it is law-abiding and orderly in its behavior; and there can be little doubt that the human brain does show a good deal of orderliness in its methods even in induction and the others.

Next arises inevitably the question whether Man is more than a machine. I have identified myself for many years with those who have tried to push to the limit the idea that the brain is essentially a form of mechanism, very complex but essentially mechanistic in the sense of being state-determined. The possible inadequacy of the mechanistic aspect has never worried me, for it has always seemed to me to be clear that a demonstration that the brain is more than a machine can be successful only if one first knows thoroughly what one means by a "machine." Only when the idea of machine is pushed to its limit can we see whether the application to Man breaks down or not. Thus those of us who are pushing to the limit the possibilities of what a mechanical brain can do are *defining* the limit to which the machine can go.

One thing seems to me to be clear. The *demonstration* that Man is more than a machine will be done only by him who really knows what the word "machine" means.

CYBERNETICS TODAY AND ITS FUTURE CONTRIBUTION TO THE ENGINEERING SCIENCES

To be asked to say, this evening, where cybernetics stands, and how it should develop, is a heavy responsibility; for we are considering matters that not only affect the future of a science but also the simple employment of many skilled workers. I am willing to accept the responsibility, for I believe that the position is today clear enough to justify a firm statement; but I trust that any action you may take after hearing my opinions will be based on a reading of the paper when published, rather than on an impression carried away from hearing me speak.

As I said, I am very glad to be able to speak here, for, in my opinion, the science of cybernetics has now crystallized to a point where one can see the field as a whole, and can get some sense of its limits. What these limits are I shall attempt to describe in a minute. They seem to me to be quite clear; and I am certain that cybernetics for many years to come will move within a certain framework of ideas. On the other hand I must not pretend to be speaking as representative of a large professional group with an accepted orthodoxy. At the moment there is little explicit agreement with what I shall say. These views, however, have now been well known for some years and have invited criticism and refutation. None has come. I am therefore of the opinion that these views are likely to endure for quite a time, in spite of the fact that some of them demand a fairly drastic change in our ways of thinking.

Let us then turn to our general topic. I am wholly agreed with those who start from the proposition that Systems Engineering has some form of *control* as its aim. There are the questions of practical control (often included under the title of Operations Research); there is the more abstract theory of control in general; and finally there is the – shall I say "guiding light" or "will-o'-the-wisp"? – of cybernetics: the attempt to develop control processes that are really brain-like.

Of the first two I shall say nothing, for they are essentially straightforward. It is the last – cybernetics – that I shall attempt to assess. What is its value? Is it our brightest hope for the future, or a mere illusion, based on a false analogy?

The Present Position

To get the matter clear, it is essential that we first understand what the human brain really is and what it actually does. This knowledge is essential both to those who must decide what the *aim* of cybernetic research shall be, and to

workers who must carry it out. The fact is, the subject, ever since the days of the Greeks, has been infested with myths and superstitions. It is only, literally, in my life-time that the human brain and its behavior has been studied as it actually is, and not according to what it ought to be. Until the arrival of Sigmund Freud, "psychology" consisted mostly of a polite discussion of the laws of logic, all on the assumption that what should be discussed is the sort of thinking that one can be proud of. To the real processes that actually go on in our heads, with all their confusions, mistakes, hesitations, meanderings, and similar failings, practically no attention was paid. As a result, the Greeks and the psychologists who followed them developed ideas about how the brain *ought* to work. These ideas are our common heritage. They will not easily be corrected. Meanwhile, far too many workers in the field of cybernetics are using ideas about the brain that are really those of the medieval schoolman. They are using powerful and modern techniques to answer questions that should never be asked, because the brain just isn't like that. We would merely laugh if a modern chemist proposed to analyze a spent can of uranium into its component elements of earth, air, fire, and water. Yet too much of today's cybernetic research is open to the criticism that it is of this type. To be more explicit, let me give some examples of the myths about the brain that still persist.

One common idea, for instance, is that the brain can predict the future. The idea certainly has a superficial plausibility, but it is fundamentally wrong. Wiener put the truth succinctly when he said that to predict the future is to carry out an operation on the past. *If* the past shows repetitions, and *if* the future sustains the regularity, then the brain will score a hit when the future arrives. But the process is wholly based on *past* events and on the degree to which the real world has consistent regularities. Let the world produce something really new, and the brain, human or other, is helpless. What happened, for instance, to those who worked with X-rays in 1905? They burned themselves horribly. The brain *cannot* deal effectively with the really new: it must wait until the new has had time to go adequately into the past. All wisdom is wisdom after the event. What we expect of mechanical brains must be based on this fact, if our expectation is to be realistic.

Another myth stands exposed if we ask "How does the brain generalize?" for experiment shows that the answer commonly is: It doesn't. Take for instance the experiments of Papert. He wore prismatic spectacles so that he saw the world, including his own limbs, reversed from side to side, interchanging left and right. With his hands outstretched, he watched someone touch a hand, and he tried to move the hand so as to prevent the touch. After moving the wrong hand repeatedly, he became skilled, and developed a new eye-hand relation. He was then trained to do the same with his feet, until it seemed clear that he had learned the left-right reversal in full generality. He was then tested in the same way at the knees. It was found that, though Dr. Papert was by no means deficient in intelligence, generalization of "reversal" had *not* occurred, and had to be learned, as a special matter, yet again in relation to his knees. The brain has *no* universal faculty of "generalization."

As a last example we may consider the myth that certain people have the special property of being a "genius" – of having some quality that makes them, in some absolute sense, more effective than other people even if all other factors are equalized. The myth is based first on bad statistics. Many workers tackle a problem, using all sorts of preconceived ideas about how to solve it, one is successful, and then *we* pick him out and assume that he must have something that is good for all problems and for all time. In the seventeenth century, for instance, many mathematicians were within a hair's breadth of getting the infinitesimal calculus. Newton came to the problem with a strong personal liking for thinking of everything as flowing, continuously, from one state to another. Is it surprising that he, with Leibniz, was favored with the discovery? Would he have been so successful as a physicist at the beginning of this century, when the need was for someone who could think easily about atoms that changed by discrete, instantaneous jumps? A good deal of the idea of a "genius" arises by the elementary but gross error that *we* are wise after the event, and can name that person who did actually make the discovery first. What remains after this fallacy has been corrected is that the persistently productive geniuses were largely people who were obsessed by their subject, and who devoted to it a fraction of the day's 24 hours far greater than the amount given by the average man. If your Mozart sees a broken twig, the shape at once suggests to him a new musical phrase; a bird call gives him ideas for new harmonies; everything is grist to the musical mill; and in one hour spent apparently doing nothing he may have gotten through more musical investigations than many of us achieve in a lifetime. Is it surprising that he could produce a symphony almost on demand?

These few examples give some indication of the extent to which our everyday thinking about the brain must be revised if we are to think about it realistically; and this we must do if we are successfully to build brains into hardware. But this elimination of the wrong is merely negative. More important is the positive understanding available that sees the brain, in all its activities, as conforming to one fundamental rule. Let me therefore turn from these medieval superstitions to a brief statement of how the brain is seen in 1961.

What is the brain?

The living brain has, of course, many aspects. To the surgeon it is a lump

of jelly that, when concussed, vibrates in a special way. This evening, however, we are obviously concerned with one particular aspect, that by which the living brain shows what may be called its cleverness. It knows how; it gets results; it gets into most difficult situations and then shows its power by finding its way through successfully. It is this aspect that we are all seeking, especially in Systems Engineering, when we try to build a mechanical brain. What *is* the brain, when it achieves its remarkable conquest of nature?

Here a personal note may be helpful. When I started in 1928 I took for granted that the brain had a gimmick: find it, and we would have access to "pure intelligence", that one could then tap in unlimited quantities, like a gusher. Today, we know better; but we know there is no gimmick, not because we have failed to find one but because we now know positively what is there. This new knowledge can be described fairly easily, though one must go cautiously, for the new knowledge fits very peculiarly on to the old ideas, and is certainly not a mere modification of them.

The first fact is that every skill or power that it possesses has been developed, either by natural selection acting on the species, or by the processes of experience in childhood and education as it affects the individual. After a billion years or so of shaping by the environment the living brain is now suited to that environment with quite remarkable precision. But it is shaped to that environment. It no more has general adaptation than the control system of a petroleum refinery has "general" adaptation, or the control system of a rocket.

The living brain, then, has been shaped for survival; and "survival" means the persistent achieving of certain goals – food must be won, injury avoided, thirst satisfied, excessive heat or cold avoided, and so on. "Intelligence" is shown in respect of these goals, or of their subsidiary goals. Thus the first principle of any brain-like or intelligent system, living or mechanical, is:

> *A system can be brain-like or intelligent only in relation to a defined goal.*

In other words, the construction of a "brain-like" system cannot be begun until the question has been answered: What do you want?

Thus, today's view of "brain-like behavior" is that *control* is its touchstone. This is the lesson of 20 years of cybernetics. If one holds on to this view one stays in the world of the scientific, practical, and modern. If one forgets it, one may drift back into the mythical, the superstitious, and the downright impossible.

I say "impossible" because this brings me to the hard core of our present knowledge. Shannon's theorems on the correction of noise were first produced for application to the telephone and radio channels. It has since been realized, however, that as they are quite abstract in content they can be applied to other processes superficially very unlike those of telephone communication. In fact they are fundamentally applicable to all processes of control, the correction of noise now becoming the correction of any deviation from what is desired, from the goal. What these theorems do is to put an absolute upper limit on what can be achieved by any process that respects the usual relations of cause and effect. In other words, if the brain works by the normal relations of cause and effect, (and no serious worker doubts this) what it can achieve as a goal-seeking device has an absolute limit. Just as the human muscles are absolutely limited by the law of conservation of energy, so is the brain limited by Shannon's theorems and the law of requisite variety.

With this new insight we can today deal firmly with the question: Can we make a super-brain? But first we must be sure that we are approaching the question in the right way. The *wrong* way is that of 1928 – to think that the brain has some gimmick that, once discovered, will be like an Aladdin's lamp – one stroke and it will produce what we want. This point of view is dead, killed by Shannon's theorems and the law of requisite variety. There are no super-brains to be obtained by discovery of the brain's gimmick, for it has none.

What then remains? We now realize that the question whether we can make a brain better than the *human* brain is a trivial matter, of no more interest than the day, a hundred years or so ago, when the engineers made the first electric motor to exceed one-seventh horse-power, thereby making a motor of more power than their own muscles. Obviously, once they had discovered how to make motors, the exceeding of the human power was of merely sentimental, not of technical interest. The same unimportance applies today, technically, to the building of a system whose brain-power exceeds that of the man who designed it. *Today the building of a brain-like system of any given intellectual power is as well understood in principle as is the building of a steam-engine of any given horse-power.* The method can, in fact, be sketched. First, there must be defined the goal (which may be complex and provisional); then comes identification of the factors that tend to prevent its being achieved; finally comes the construction, often stage by stage, of the regulator that shall so process the information about the disturbances that the goal is reached in spite of them.

If this process should seem to simple and ordinary, I would ask those who think that the human brain is smarter to bring forward the actual evidence on which such a claim to smartness is based. As I said earlier, the idea that certain people have a "genius" fades away on scrutiny of the facts. I can only say that I know of no fact to suggest that the human brain gets its results in any way other than in the way just sketched.

Just as power-engineering is dominated by the law of conservation of

energy – that every erg of work that an engine can put out must have been put in – so every regulating and control device is dominated by the law that the amount of appropriate selection that it can achieve is bounded by the amount of information that it has received and processed. Thus we can today build an artificial brain as large and expert as we please, provided the total quantity of information processed is adequate. One might say that we have found our Aladdin's lamp, and we now know that through it we can get whatever we desire; there is only one qualification: whatever we take must be paid for!

Let me summarize briefly the position today. Through the ages, the brain like the heart, collected a great variety of merely mythical attributes. As the heart was found eventually to be simply a pump, so has the brain been found to be a regulator, and therefore a processor of information. It has the limitations implied by Shannon's theorems and by the law of requisite variety. The great strength of the human brain lies in its eon-long shaping to the needs of this terrestrial environment, so that problems of *that type* find it particularly well adapted. There is no reason to suppose that any other data-processing device, given the same amount of developmental work, need be inferior to it. Nor is there any reason to suppose that non-living devices are unable to exceed it in performance. The level of performance to which a mechanical brain can be raised depends simply on the quantity of design and experimental work that is devoted to its development.

The Future of Cybernetics

As I said, cybernetics has now crystallized out to a clear, coherent, and firm statement of the basic principles of brain-like mechanisms. Right or wrong, these basic principles will have to serve as our scientific framework for quite a number of years, until, as always happens, they are replaced by something better. But though cybernetics has reached, as I see it, the time of crystallization, this by no means implies that cybernetics will cease to develop. So I would now like to consider its future, in technology, in teaching, and in research.

Let me turn first to the future of cybernetics in industry and technology, for here will be decided whether cybernetics is living and active, or merely an academic set of speculations.

My belief is that in time its contribution to industry and technology will be decisive, though not perhaps along the lines expected 10 or more years ago. What to me is outstanding is the discovery of the *limitation* on the artificial brain – that implied by Shannon's theorems and the law of requisite variety. Here the situation is very similar to what it was a hundred years ago, when the power engineers were inventing new engines almost daily, and every imaginative engineer confidently expected that the machine with perpetual motion would be invented shortly. Then emerged an unpleasant surprise. Evidence began to grow that all machines were fundamentally limited, that energy was conserved, and that every erg of energy that came out had first to be got in. For a time many engineers felt profoundly disappointed. Nevertheless, the *acceptance* of this limitation was the first step to a new and altogether more realistic grasp of the principles of machinery. In the long run, the engineers who accepted the limitation outdistanced those who clung on to the hope of the perpetual motor, precisely because the ideas of the acceptors were more in accord with the actual realities. Thus I feel sure that those who accept the modern idea of the brain, and accept the limitation of it and of all brain-like mechanisms, will eventually go far beyond those who want to hold on to the old idea that the living brain has some mysterious "extra", that can do things in a flash.

How will the new method improve on the old? As soon as it is accepted that all brain-like actions are acts of appropriate selection, achieved by the processing of information, so soon can we get the focus of interest onto the proper place, namely, on to the system's *efficiency*. We work now, not to find a magic process that will give us all in a flash but to develop one that achieves the goal with reasonably high efficiency.

This change may sound small; actually it is far-reaching. We are, in fact, still so ignorant in the matter of efficiency that while the power-engineer would today laugh at a system that was only 10 percent efficient, we are sometimes, quite ignorantly, today using informational processes that are "efficient" only to a tiny fraction of one percent. Lest this seem a wild exaggeration, let me quote an actual example. After 15 years of discussion by hundreds of people (including myself) it has only just been pointed out, by Don Campbell of Northwestern, that if we know what the goal is we can use that knowledge to cut down the time of search for it. By how much? If we know the goal as well as the starting point, and use this information, the number of operations drops to about the square root of the original number. Thus if the original number is a million operations, the improved method drops it to about a thousand. This means that if we know the goal, but do not use the information, our efficiency in this example, is a tenth of one percent! Such are the inefficiencies that are only too common today in our methods for brain-like processes.

We must admit that today our intuitive sense for "quantity of information" is very poorly developed. It is to be hoped that the younger generation, thoroughly accustomed to working with discrete states rather than with continuous functions, will steadily remedy the matter. It is clear that there can be no major improvement in the efficiency of brain-like processes until at every point the synthetic processes are raised to an efficiency that is at least of the

same order of magnitude as those used by the brain.

What then is the future of cybernetics in technology, in brief? Firstly, it will give up chasing the will-o-the-wisp of the brain's supposedly magical powers of intelligence; it will do what the brain actually does, not what it is popularly supposed to do. This means in practice that it will start with a defined goal, one for each job; and it will then settle down to the steady development of the appropriate control-mechanism for the job, focusing simply on ***maximizing the efficiency.*** This may sound hum-drum, but it is what the power engineer does. Started on, say, a gas-turbine or a nuclear rocket, he expects no miracles, but settles down to a steady developmental program that works progressively forward from perhaps a ludicrously low initial efficiency towards the higher. This too is cybernetics' method, in Systems Engineering, for developing systems of really brain-like capacity.

What Must be Taught

What I have said about how cybernetics will be used in industry shows clearly what we must teach our students; and a moment's reflection makes us realize how grossly unsatisfactory is our teaching today. In special branches, of course, the teaching is highly skilled and adequate – in the theory of servo-mechanisms, for instance, or in the theory of telephone communication. But in the theory of brain-like mechanisms it is quite otherwise. We are suffering today from something of a historical accident. The neurophysiologists always expected to be the eventual discoverers of how the brain worked. Then came World War II, and much activity behind closed doors. Suddenly, in 1947, the doors were opened, and the neurophysiologists were confronted with incomprehensible machines that carried out brain-like processes, and an incomprehensible mathematics of information that purported to explain them. With few exceptions, the neurophysiologists abandoned the subject, and buried their heads under a heap of amplifiers. If you think these words too strong I would call your attention to the new *Handbook of Physiology,* which must be regarded as the best that can be produced at the moment. The volumes of Neurophysiology are three and massive, filled to the brim with the most erudite facts; yet they quite fail to give an answer to the intelligent young man who simply asks: What is the brain *for*? The fact is that so far as the higher physiology of the nervous system is concerned, most neurophysiologists are between five and 10 years behind the times.

How then should cybernetics be taught? Answer is not easy, for this is partly a matter of administration, and cybernetics sprawls most irregularly over half-a-dozen disciplines. The specialized developments, of course, are fairly easy to allocate to the departments; what is difficult is to get taught the foundations,

which lie in several departments simultaneously. Here, what is wanted is simplicity, common sense, and a getting of first things first, with a resolute exclusion of technicalities until their necessity has become obvious. Thus the teaching must, I think, commence with a study of evolution, natural selection, and homeostasis; for until these matters are understood the word "brain" has no clear functional meaning. Then would come the general theory of servo-mechanisms and regulators (with only the slightest reference to the linear case), and a gradual introduction to the importance of the *quantity* of information involved, when the system becomes complex. The quantities of information would be measured either combinatorially or by the method of McGill and Garner (as being more suitable than the specialized methods of Shannon). By now the student would be well able to see the form of the subject and to decide in which way he wished to specialize. At the moment, it seems that in most universities the course would have to be arranged to be shared between several instructors, each asked to deal with certain topics so that between them the whole field is covered.

Such a course should give the student the essential basis of knowledge for understanding the higher physiology of the nervous system (if he is going into computing or Systems Engineering). I recommend these lines with some confidence as I have already taught them at Urbana, and found them practical.

Future Research in Cybernetics

Finally, I would like to say a little about the future of research in cybernetics. Here I must speak at first somewhat critically.

It seems to me that some of the present work is characterized more by the elaborateness of its technique than by the soundness of its aim. Sophisticated modern techniques are being used to attack problems that are trivial or even medieval in their conception. What would we say today if a medical research team wanted a mass spectrograph to find which of the four humors was chiefly present in the brain? – we would merely laugh. Yet some of the research going on today can only be described as being of this type. The situation will, of course, cure itself eventually; but at the present time there is a real need for a searching scrutiny of research into cybernetic projects, especially with regard to its primary aims. I must go cautiously on this point, for I have no wish to do harm to honest workers; but I cannot help recommending that all concerned with the subject, sponsors and workers, should persistently ask themselves: Is this research seeing the brain as it should be seen in 1961, or as it was in 1941? Even in this short time there has been a complete revolution in our view of its higher functions, and this revolution must be reflected in the researches of today.

should appeal to some supposed mysterious power of the brain. In 1961 such as appeal is as unjustified, and as discreditable, as would be an appeal, made perhaps in the stock market, to some supposed mysterious power of electricity. Today the rule must be: Every research that refers to brain-like properties must say what the properties are, and just why they should be wanted. The justification must ultimately be in some sense practical, and explainable in common-sense terms. An appeal to the *mistique* of the brain today marks the mountebank, or perhaps, the ignoramus.

Of good problems for research, there is, however, no lack. We know, for instance, as I said earlier, far too little about the *efficiencies* of various regulatory processes, especially of those primitive but fundamental processes used so much by biological systems, such as that of proceeding by trial and error.

Again, we know practically nothing about the general properties of systems with many states of equilibrium, of systems, as one might say, with distributed memory. In the latter class are those with threshold; it seems almost incredible that though we have known for 50 years that many of the cerebral units work by threshold, we know far too little about the general dynamics of systems of that type.

Another great need today is for a formal mathematical theory of simplification. Our future progress with giant brains will depend acutely upon our power to simplify. At present, simplifications are handled merely by intuition and native wit. We require a developed theory of it. The mathematicians' work with homomorphisms is clearly along this line; the work needs re-orienting to the needs of information theory and Systems Engineering. To sum up: If I have spoken somewhat harshly of some aspects of cybernetics you will not think that I condemn it root and branch! In 1928 when I first started to think about the brain and its higher functions, the topic was an impenetrable mystery. Today that mystery has gone, and cybernetics is in the difficult state of transition. From a somewhat science-fiction-like start it has reached cohesion and solidity. Today we are ready to develop and use the new knowledge. Nothing stands in our way but our typically human tendency to hang on to our old ideas, even after they have become obsolete. The future, in the study and control of brain-like mechanisms, lies in the hands of those able drastically to revise their own ways of thinking about the brain, and with the next generation. If they are properly taught they will see the brain realistically, and they will then proceed to demonstrate that the science of brain-like mechanisms is essentially clear, practical, and useful.

CHAPTER SIX

BY W. ROSS ASHBY

Analysis of the System To Be Modeled

I would like to start, not at: How can we make a model?, but at the even more primitive question: Why make a model at all? By getting a clear answer here we shall, I hope to show, obtain a firm foundation for the further developments. I have frequently been impressed recently with how easy it is to jump too quickly to questions of this type, only to find later that the whole elaborate attack has been directed against a basically inappropriate target.

I would like then to start from the basic fact that every model of a real system is in one sense second-rate. Nothing can exceed, or even equal, the truth and accuracy of the real system itself. Every model is inferior, a distortion, a lie. Why then do we bother with models?

Ultimately, I propose, we make models for their convenience. We make models of an aircraft wing and put it in a wind-tunnel because taking a real wing into the air is far more expensive and dangerous. We write equations on paper to represent the flows of traffic through new highways, because it is cheaper than building many highways and then scrapping all but one. And sometimes the avoided "inconvenience" is extreme, as when we trace out, in a model, what *would* happen if the water level of the Great Lakes were raised fifty feet, or if planets were attracted according to the inverse cube.

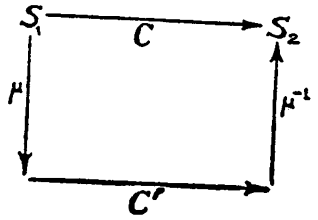
The superlative convenience and accuracy of Newton's way of representing the gravitational actions of planets by a few pencil marks on paper has sometimes led to the idea that Newton's model is in some way inherently superior to the actual system—that in this model he somehow extracted what was valuable, the rest being mere clutter. Here I want to adhere to the other point of view, that the truth is the *whole* system, not any extract from it. I would point out here that however vigorously Newton's equations of motion are defended as the extreme in truth, the astronomer must always reintroduce the discrepancies, such as (until Einstein's work) the rotation of the perihelion of Mercury, when he would make *real* predictions. Ultimately, the raw facts are final.

Taking, then, the practical usefulness of models as a basis, the process of using a model can be formulated in a manner highly congenial to modern mathematics [5]. We start with the assumption that the scientist is interested in some process that will, or may, occur in the real world—a planet goes from its present position to where it will be in a year's time; a pile of steel girders becomes a constructed bridge; the present traffic chaos in a certain city changes to a smooth flow. This change, call it C, can be represented by an arrow going from the one state S_1 to the consequent state, S_2.

$$S_1 \xrightarrow[C]{} S_2$$

In the model there must be a corresponding change C' $\longrightarrow$, from one specification of planetary coordinates to another, or from one arrangement of lines on a drawing board to the final drawing of the bridge.

There must also be a rule of correspondence μ such that if we translate the real state S_1, using μ, into the model, there apply C', and then translate back, using the inverse μ^{-1}, we



arrive again at S_2. Since the correspondence must (in a good model) hold for all pairs S_1, S_2:

$$\mu^{-1} C' \mu = C$$

(with a convention for the order of operation), in the sense that the single, direct operation C, in the real world, is equivalent to the triple operation $\mu^{-1} C' \mu$, going through the model.

The equality stated is necessary (if the model is to be a faithful one). It also makes clear the importance of the purely practical criterion that I have above called "convenience" (using the term very broadly). Use of the model demands three operations (μ, C', and μ^{-1}) instead of the single operation C. What science has found is that many cases exist in which the use of the three operations is actually more convenient than the use of one.

It seems to me that this purely pragmatic reason for using a model is fundamental, even if it is less pretentious than some of the more "philosophical" reasons. Take for instance, the idea that the good model has a "deeper" truth—to what does this idea lead us? No electronic model of a cat's brain can possibly be as true as that provided by the brain of another cat: yet of what *use* is the latter as a model? Its very closeness means that it also presents all the technical features that make the first so difficult. From here on, then, I shall take as a basis the thesis that the first virtue of a model is to be useful.

At this point I may refer to the possible objection that this

formulation is not applicable to a system like the human brain, which may have some "random" or unpredictable aspects. To this I would reply that the scientist is essentially concerned with finding and using such laws as *are* present, the random aspects being usually relegated to such terminal classes as the residual error of an analysis of variance, or the probable error of an astronomical prediction. Again, when randomness is present in the actual events, as at Las Vegas, the *probabilities* may well be invariant and exact for a particular game, such as roulette; and the probabilities associated with a proposed new game may well be exactly predictable.

Thus the formulation given may well be appropriate to such aspects of the real system as have a law-obeying regularity. What I have to say will be relevant to those aspects.

The Multiplicity of Models

From this point of view, there is no such thing as *the* true model of such a complex system as a cat's brain. Consider, for instance, the following four possible models, each justifiable in its own context:

1. An exact anatomical model in wax.
2. A suitably shaped jelly that vibrates, when concussed, with just the same waves as occur in the real brain.
3. A biochemical soup that reacts biochemically just as does the cat's brain when drugs are added.
4. A programmed computer that gives just the same responses to auditory stimuli as does the living brain.

Clearly, complex systems are capable of providing a great variety of models, with no one able to claim absolute priority.

We are in danger, perhaps, of being led astray by the outstanding merits of certain well-known particular models. We

rightly admire Newton's system of equations and laws and, after its great success, are apt to think that he discovered *the* model. I suggest that his model is widely used largely because pencils and paper are widely available, and *his* type of mathematics widely known. Had our circumstances been very different we might well have preferred a different model: had we lived, for instance, in a world where algebra was not a practicable process, but where many point-sources of light and many conical bodies made the geometric development of ellipses instantly available, we might well have found that wholly geometric methods were preferable to the algebraic.

With a multiplicity of models available, our real question becomes (assuming a model to be necessary): which one shall I choose? But before considering this question there is a highly relevant factor to be considered.

Bremermann's Limit

In my opinion, one of the most fundamental contributions to the epistemology of large and complex systems was made by Bremermann's demonstration [3,4] that there is a limit to the rate at which matter can transmit or process information. The limit rests on two of the most basic properties of matter (Einstein's Mass-Energy relation and the Uncertainty Principle of Heisenberg), so it is not likely to be overcome by any merely technical advance. Further, being so general it applies with equal strictness to the matter in a computer and to the matter in a scientist's brain. We are *all* subject to this limit.

Its actual value is 10^{47} bits per gram per second. This number may look large, but its importance comes from the fact that when we, as scientists, look at a complex system we may easily envisage processes that would require far more than this quantity. (Notice first that taking a ton of computer only

adds 3 to the exponent; changing to nanoseconds adds only 9; and taking tons of computer and decades of time will not increase the number of bits processable beyond about 10^{70} bits.) Yet as soon as complex combinative processes are considered, the demand may go far beyond this limit.

As a simple example, suppose we have a square screen of lamps, with 20 a side, and thus 400 lamps in all. Let each lamp be only lit or unlit. The number of illuminated patterns that can be presented on it is 2^{400}, i.e., about 10^{120}. Suppose now that we ask some question about *grouping* these presentations, for instance: What grouping will best correspond with "looking like a cubist picture"? The number of dichotomies of 10^{120} objects is $2^{(10^{120})}$, so to ask for a particular dichotomy is to select one item from this last number of items. Until further information is available, this selection demands bits to the extent of its binary logarithm. Thus our apparently harmless question about the grouping of these pictures, on a basis of only a 20 x 20 screen, has latent in it a demand for the processing of 10^{120} bits—a quantity far beyond what is possible under the limit.

I need not give further examples, for it is very commonly found that as soon as we make an estimate, even the roughest, of how much information-processing is required when the system is complex and combinatorial, the estimate is beyond Bremermann's limit, often beyond to a degree far exceeding that shown by the example above. Far from being merely "theoretical," the limit is in fact outstanding as a practical obstruction.

It is on this basis that the method of "model-making" has an irrefutable claim as *better* than the study of the raw facts. The model, by replacing a system whose study would demand a transgression of Bremermann's limit, makes the study possible. From this point of view we transfer from system to model to *lose* information. When the quantity of information is small we usually try to conserve it; but when faced with the excessively large quantities so readily offered by complex systems, we have to learn how to be skillful in shedding it. Here, of course, model-makers are only following in the footsteps of the statisticians, who developed their techniques precisely to make comprehensible the vast quantities of information that might be provided by, say, a national census. "The object of statistical methods," said R. A. Fisher [8], "is the reduction of data."

Analysis

On this basis, the making of a model may find a firm justification (though making a model is not the only way of lessening the quantity of information).

Within the method of models there are various ways of proceeding. One of the ways that often helps to reduce excessive demands on the quantity of information-processing is that of "analyzing the whole into parts." Its mode of action depends on the following features.

When parts combine to form some whole, we often find that the quantity of information necessary to comprehend the whole increases, not proportionately to the number (n) of parts, but far faster, often exponentially, so that the quantity of information approximates to a^n, where a is some base. Now it is easily verified that when n and a are both large, the operation of dividing n by some quantity k and also multiplying the whole by k, resulting in $ka^{(n/k)}$, is to cause a very great *reduction* in the quantity.

"Dividing a whole into k parts" does just this. (Each is $1/k$th the size, so its complexity will approximate to $a^{(n/k)}$. But

there are k parts, so the total quantity [if the interaction between the parts is negligible], becomes k times this.) When the interaction between the parts is not negligible, the fall is not as great, but it may still be worth while. When the parts are in *full* interaction with one another (every part having a direct and full effect on every other part) then there is commonly no gain at all. Then the method of proceeding by analysis is either futile or must be justified in some other way.

Thus, the method of analysis, sometimes presented as obligatory, is in fact a strategy for taking advantage of the situation (if it occurs) of the whole system's being composed of parts that do not have direct and full effects on one another. Thus it may well be a natural way to take advantage of such facts as:

1. In a big city, not every person knows, or communicates directly with, every other person.
2. In a brain, not every nerve cell is connected directly with every other nerve cell.
3. In a big computer, not every part is directly affected by every other part.

To sum up, the method of analysis has no right to be regarded as the correct one, absolutely. It is, however, often a very powerful resource when one faces the obstruction of Bremermann's limit.

From now on, I shall assume that we are working within some specific problem for which it has been decided that the method of analysis is appropriate.

Finding the variables

The would-be model-maker is now in the extremely common situation of facing some incompletely defined "system,"

that he proposes to study through a study of "its variables." Then comes the problem: Of the infinity of variables available in this universe, which subset shall he take? What *methods* can he use for selecting the correct subset?

To start at first principles, it may be taken as axiomatic that the scientist, selecting a set of variables to be "his system," must either be arbitrary in his selection or must be guided by reasons. The former case need not be discussed here, but the case where he selects for reasons emphasizes that the selection must be guided by some cause, or primary motive. Often there is an obvious goal, but sometimes the worker, not having a goal sufficiently specified, is plagued by uncertainty about what to do towards selecting his variables.

It is here that a practical goal can be invaluable. In the behavioral sciences there are plenty: How reduce the delinquency in this town? How teach these students better? How improve the viewing of this TV screen? Somewhere, then, there must be a "generating question," able to provide a criterion for the various steps taken later. Without this primary selecting factor, selection among variables can only be aimless and arbitrary. The primary question, therefore, is not: What are my variables? but: What do I want? Under what terms of reference am I working? Without a sufficiently practical knowledge of his criterion, the worker is indulging simply in a random intellectual walk, in the hope that something interesting will turn up. As there is no *method* possible here, let us restrict ourselves to the case when some sufficiently well defined generating question exists.

The generating question will usually at once suggest a list of variables; it is, however, merely provisional. The scientist's next task is to bring the list to a condition of "completeness" (to a degree sufficient for the main task). But before discussing the meaning of "completeness" we must make a point clear.

The worker who has had some training in mathematics can only too easily fall into the habit (or trap) of thinking that a "variable" must mean a numerical scale with an additive metric. This assumption is quite unnecessarily restrictive, sometimes fatally so. The meteorologist has long worked with his five "types of cloud," the veterinarian with the various "parasites of the pig," the hematologist with the four basic types of "blood-groups." Modern mathematics, using the method of set theory, is quite able to handle such variables, which are often unavoidable in the behavioral sciences. What follows below will be written so as to be equally applicable to the metric and nonmetric variables (with obvious qualifications at certain points).

With a provisional list of variables the scientist's next task is to examine them in detail. In particular he will want to know of each: Is it relevant?

Tests for "relevance" have many technical forms though they all express the same basic idea. According to the type of investigation we may ask, of two variables X and Y:

1. Is the correlation (suitably defined) between X and Y zero?
2. Is $\delta X/\delta Y$ everywhere zero?
3. Is the transmission (in Shannon and Wiener's sense) between X and Y zero?
4. Are X and Y independent in probability?
5. If we changed Y, would X change?

All these forms (and there may be further variations) seem to be expressions of the basic idea in set theory: Is the relation between X and Y a ***product*** set?

It should be noticed that all these tests are ultimately operational: they can be brought to ***demonstration***, and owe nothing to any argument of plausibility. It is worth noticing here that the test by demonstration is always treated as the ultimate test, let plausibility say what it will. Thus, on an afternoon in 1888, Heinrich Hertz showed two pieces of electrical apparatus with no trace of electrical connection between them. Yet after he had showed the correlation between their behaviors, sparks in the one following a switchclosing in the other, no scientist, whatever his philosophy, denied the validity of the proof of ***effective*** connection. The operational test is the last court of appeal.

With this test (that some aspect of X's behavior is *conditional* on Y's value) available for all pairs of variables, the scientist's task is then logically simple. He looks for a set of variables that (1) is clearly related to his primary generating question, and (2) is closed, or complete, in the sense that for every variable in the set, the variables that affect it are all already included in the set.

(Sometimes he will accept a weaker form, in which some of his variables are affected by other variables that are not in the set but are otherwise acceptable. These other variables are his "parameters." Only the generating question itself can decide what variables may be allowed this special relation to the set. In the theory of "machines," in the general sense, they are its "input.")

In this connexion it may be worth noticing that systems showing "memory" (specially common in the behavioral sciences) may, at least in principle, be treated by exactly the same method. All that is necessary is that some of the variables in the set will be related to others by being the values of these others taken at some earlier time.

Before we leave the topic of "which variables," however, a word may be said on an aspect that must not be left unnoticed: what if some of the variables are relevant only in combination, not individually? Here we open a topic of great difficulty and complexity, on which much remains to be said.

This richer possibility has already been encountered (in the history of science) in all the criteria of conditionality mentioned above:

1. When the correlations between X, Y, and Z are all zero pair-wise, but some partial or third-order correlation shows that linkage is present.

2. When the zero-ness of $\delta X/\delta Y$ depends on the value of Z, (which corresponds to examining the value of $\delta/\delta Z \cdot [\delta X/\delta Y]$, i.e. of $\delta^2 X/\delta Z \delta Y$).

3. When $T(X:Y)$, the transmission between X and Y, is zero but the conditional transmission $T_z(X:Y)$ is not.

4. When X, Y, and Z are probabilistically independent, pair-wise, but not as a triple.

5. When the effect of Y on X (e.g. a switch controlling a light) depends on the value of Z.

6. When the ternary relation R is not a product set, yet all its Z-sections show a product-set relation between X and Y.

These examples show that we are now moving into the topic of the systems that show complex internal interactions. Here the complexities increase with extreme speed as we ascend to interactions of higher order. Bremermann's limit soon puts a stop to such explorations! The subject deserves extensive generalized study, for those who work in complex systems cannot afford not to be well armed with the right ideas. Fisher's general methods for the treatment of high-order interactions [9] are now well known, but they are not generally applicable, as there must be an outstanding variable on which a variance is meaningful.

A possible line of greater generality has recently been commenced [2], using the idea of "cylindrance," but the subject is still largely unexplored.

Beyond this point, so far as I know, it is impossible to go

while treating systems in general. Basically the worker is selecting (a set of variables from the multitude that the universe offers). As a selector he is subject to the rule: Any system that achieves appropriate selection (to a degree better than chance) does so as a consequence of information received [1]. The necessary information may come from past knowledge of the particular type of system, and also be won by trial and error (with "experiment" as a specially efficient form of trial).

The process of finding a suitable set of variables may be summarized:

1. There must exist some "generating question" as primary criterion. Without it, selection can only be arbitrary.

2. The generating question must generate many possible sets of variables. As the relation between question and set is almost never one to one, there are often many sets that satisfy the demand (in various ways and to various degrees): to look for *the* set is often inappropriate. From these various sets further selection can be made only by considering further details of the particular question.

3. The work of selection ultimately becomes one of testing for independence between variables. The process is essentially the same whether it goes on through the statistician's tests of significance, or through measures of the transmission of information, or through direct experiments to test for causes, or through the mental processes of the scientist, or through the operations of a programmed computer.

Finding the Isomorph

Finding a suitable set of variables, however, is (as was said earlier) only a means to an end. The end, in model-building, is some relation (law, structure, pattern) isomorphic with the

relation in the real world. Sometimes the set of variables is almost obvious and the major part of the work is to find the relationship, the isomorphism. It is here that the myth of the "genius" may enter, to the exclusion of the scientific approach.

To the scientist, all selection is assumed to be subject to the postulate that appropriate selection (to a degree better than chance) is possible only on a basis of information received. He does not, in other words, admit the concept of "inspiration" as a factor in selection. How then, when he considers the model-making activities of (say) Newton or Gauss or Mozart is he to explain the known facts?

The answer may well be that a most valuable way of obtaining more information is by trials, and there is plenty of evidence that those who showed unusually high powers of selecting the right law, theorem, or chord did in fact do a great deal of work through trials, either on paper or in their heads. Newton, once asked how he solved so many problems, replied simply: "By always thinking about them." Gauss, in a letter to Olbers [10], wrote:

> Perhaps you remember . . . my complaints about a theorem which had defied all my attempts. . . . This lack has spoiled for me everything else that I found; and for four years a week has seldom passed when I would not have made one or another vain attempt. . . . But all brooding, all searching has been in vain. . . . Finally I succeeded several days ago.

Any idea that the genius goes straight to the solution clearly is inapplicable here, with its four years of "all brooding, all searching. . . ." The origin of this misconception may perhaps be revealed by the end of Gauss's letter [10], when he says:

> . . . when I some day lecture on the topic, nobody will have any idea of the long squeeze in which it placed me.

Without intending any deception, Gauss may well have left some of the audience with the impression that he produced the method of solution instantaneously at the blackboard, and thus quite misled them into thinking that *his* processes (for choosing the appropriate method of solution) were not subject to the ordinary law of selection.

"Finding the isomorph" is, from this point of view, perfectly clear in its *general* principles. The isomorph may be deducible from available knowledge; but this process is not research, only an exercise in the application of the known. If present knowledge is insufficient for deduction of the isomorph, then more information must be obtained. There are many ways of obtaining more information—here I wish only to emphasize the importance of *trials* as a source, especially of those that occur, in a not specially orderly way, in the researcher's brain. Poincaré's well known description [13] tells how he made a discovery after taking an unusually large amount of strong coffee:

> A host of ideas kept surging in my head; I could almost feel them jostling one another, until two of them coalesced, so to speak, to form a stable combination.

My personal belief is that such "trials," whether semi-systematically as Gauss's, or quite randomly as Poincaré's, play a large part in every active research worker's progress. Speaking personally, for years my method of attack was to fill my thoughts freely at bedtime with the topic in hand, so that the problem could be seen in the clearest possible way—so that the utmost possible tension was created, in other words. "Sleep" followed, and I was often able, the next morning, to set the whole matter in a much clearer way. (Speaking psychiatrically, I can also say that the method has dangers and must be used with some knowledge of when to stop; but I do not think that intellectual success and psychiatric danger can be entirely dissociated.)

Such trials must be to some degree "at random," for the worker does not yet know where to try; but it is also clear that a Poincaré, or a Gauss, with a lifetime of experience behind him, will tend to conduct his trials in those ways on topics that offer, on the average, a better prospect for success than those chosen by a beginner.

The general rule for such selections may thus be stated: Use what you know to narrow the field; then, within it, make trials at random. Any rule that claims to be superior to this rule must necessarily involve some appeal to "inspiration," the action of some guiding factor not possessed by the worker.

The subject of "finding the isomorph," which includes some of the great triumphs of science, has been bedeviled by those cases where a guess turned out right, leaving us gasping at its unexpectedness. The subject will be seen in proper balance only when the historians of science are equally careful to record the outcomes of the other, less successful, guesses. William Hamilton, for instance, was certainly one of the outstanding mathematical physicists of the last century, and a maker of many discoveries; his method for treating complex dynamic systems, for instance, is still used by everybody today. He also developed a penetrating eight-dimensional algebraic method ("model") for understanding the nature of polarized light, at that time a great mystery. After several years' work he had his theory perfected, and the first experiment that it predicted showed it to be wrong. This is the *reality* of how the genius gets his results!

Extending the Model

In the behavioral sciences, any isomorph will have been obtained by comparison with facts from some portion of the real word, from some finite number of variables, and over a finite range of each.

Once the model has been made, the work of the model-maker has reached a temporary completeness, but usually he then immediately wishes to see whether the model's range of application may be extended.

The process of extension, if we are to stay within the framework of the ideas expressed in this paper, will be subject to just the same postulate as the other processes of selection; for, of all possible ways of extending, the model-maker naturally wants to select those that have some special property of relevance. Thus, a model of the brain in gelatin, that vibrates just like the brain under concussion, is hardly likely to be worth extension in the biochemical direction.

From this point of view the process of extension is essentially an exploration. So far as the worker does not *know* the validity of the extension, to that degree must he explore *without* guidance, i. e., "at random." Newton himself, after he had found the coherence of the astronomical facts, must have trembled when he first applied his model to earthly mechanics: however confident he may have been, he must have known that confidence can be misplaced.

A clear example of the uncertainties of extension occurred when Einstein, by expanding his relativistic formula for a moving body's mass, found it proportional to

$$m_0c^2 + \tfrac{1}{2}\, m_0v^2 + \ldots \text{ (negligible terms)}.$$

(where m_0 is its resting mass, v its velocity, and c the velocity of light). He noticed that $\frac{1}{2}\, m_0v^2$ was the body's energy due to its movement, and wondered whether m_0c^2 might correspond to some basic or "total" energy associated directly, in some unknown way, with its having mass m_0 [7]. Writing in 1920, Einstein admitted frankly that he had no idea whether it would prove to be significant, or to be just an algebraic artifact that must be ignored. Decades later it proved to be

highly significant, showing that his original theory *was* extensible at that point, but here I wish to emphasize his admitted inability even in 1920 to say whether this extension would or would not sustain the isomorphism: only further explorations could tell. Ultimately, all arguments about plausibility must give way to further tests against the raw facts.

Models in the Future

A word on this topic may be of interest and may be specially important today if, as I believe, "model-making" in the future is likely to differ somewhat from that of the past.

The distinction lies essentially in the facilities available to the model-maker. Until about 1940, every model-maker had little more than the resources of pencil and paper, and of perhaps sixteen hours in the day. Every model that he made was subject to these restrictions and to the fact that his brain, as a material dynamic system, could not get through more than a certain quantity of information-processing in one life-time, with Bremermann's limit as the ne plus ultra.

Sometimes, as Newton found, a comparatively simple model can be isomorphic with an extremely broad range of phenomena; then we speak of his discovering a great "law," (not universal, however, for his "law" fails at the cosmic and nucleonic extremes). When this happens, the event is so striking and worthwhile that whole generations of later scientists take as their aim the finding of another such isomorphism.

I do not for a moment wish to suggest that no more such laws remain to be found, but it is true that most scientists cannot expect to be as lucky: much of their work, especially in the behavioral sciences, will have to be on the construction of isomorphisms that are not only of narrower range but are also much more complex in their structure.

Before 1940, models of really complex structure were both unconstructible, because of the labor involved, and unusable for the same reason, large quantities of information-processing being impossible. Today, however, much greater quantities of information can be processed, and we must expect the construction of "theories" (or models) of much greater complexity. Such models have already been built of such complex systems as the respiratory and cardio-vascular [11], the thermo-regulatory [12], and that of electrolyte distribution in living tissues [6].

I see no reason (in the really distant future) why all model-making, and in this I include all "law-discovering," should not be carried on, as a routine matter, inside computers. The basic processes of search must go on equivalently whether the mechanism is made mostly of copper in a factory-built computer, or mostly of protein in a living brain. In a sense, the problem of finding an isomorphism is trivial, for it has the logically irrefutable solution: generate all possible forms and examine them seriatim. If the set of forms has no structure, no constraint over it, *this solution cannot be improved on*, and computer and genius alike are reduced to simple drudgery. (As example: find a rule, as complex as is necessary, for relating the peoples' names, in the New York Telephone Directory, to their numbers. Unless the company uses some systematic method that imposes a constraint or "structure" on the relation, the "rule" will end by being as bulky, and complex, as the directory.)

Often, however, human questions have human answers, and here the experience imposed on us in a billion years of evolution and a few decades of personal learning may well find us humans able to select much more economically in questions of human types. (Similarly the digital computer can select much more economically when the structure is in

any way related to the scale of 2.) When questions having a large "human" component occur, the human information-processor has a great advantage. But with this special advantage set aside, the processes that must be carried through when information is to be used in the making of models, finding laws, and finding constraints (all equivalent to recoding the information into a more compact form) can be performed alike by all law-abiding mechanisms.

My expectation is that behavioral scientists will have only one model of the human brain (with parameters to allow for age and similar factors). It will be essentially a large general purpose computer, kept at an international center. Initially it will have no specific structure, but more and more will be programmed into it as facts, physiological and psychological, become available. Gradually its behavior will become more and more recognizably brainlike. It will be a total archive of the time's knowledge, and it will, on demand, give (by doing) a prediction of the consequences implied by that time's knowledge. When the predictions are falsified by new experimental facts, the machine's structure will be altered. This is the way, as I see it, that we shall move from "Newton's theory of gravitation" to (300 years later) "the world's theory of the brain."

Literature Cited

1. Ashby, W. R. "Computers and Decision-Making." *New Scientist*, 7 (1960).
2. ——. "Constraint Analysis of Many-Dimensional Relations." *Progress in Biocybernetics*, ed. N. Wiener and J. P. Schadé. Amsterdam: Elsevier Publishing Co., 1965. Vol. 2, pp. 10-18.
3. Bremermann, H. J. "Optimization through Evolution and Recombination." *Self-Organizing Systems, 1962*, ed. M. C. Yovits, G. T. Jacobi, and G. D. Goldstein, Washington, D. C.: Spartan Books, 1962. Pp. 93-106.

4. ——. "Quantum Noise and Information." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. Lucien M. LeCam and Jerzy Neyman. 4 vols. Berkeley and Los Angeles: University of California Press, 1967. Vol. 4, pp. 15-20.
5. Coombs, C. H., Raiffa, H., and Thrall, R. M. "Some Views on Mathematical Models and Measurements Theory." *Psychological Review*, 61 (1954), 132-44.
6. DeLand, E. C., and Bradham, G. B. "Fluid Balance and Electrolyte Distribution in the Human Body." *Proceedings of IBM Scientific Computing Symposium on Simulation Models and Gaming*. New York: IBM, 1966. Pp. 177-93.
7. Einstein, A. *Relativity: The Special and General Theory*, trans. Robert W. Lawson. New York: Henry Holt & Co., 1920.
8. Fisher, R. A. "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society*. Series A. 222 (1922), 309-68.
9. ——. *The Design of Experiments*. Edinburgh, Oliver & Boyd, 1935.
10. Gauss, C. F. Letter to Olbers, Sept. 3, 1805. Quoted in G. W. Dunnington, *Carl Friedrich Gauss, Titan of Science*. New York: Exposition Press, 1955.
11. Grodins, F. S. *Control Theory and Biological Systems*. New York: Columbia University Press, 1963.
12. Milsum, J. H. *Biological Control Systems Analysis*. New York: McGraw-Hill Book Co., 1966.
13. Poincaré, H. *Science and Method*, trans. Francis Maitland. London: T. Nelson and Sons, 1914. Pp. 52-53.

MATHEMATICAL MODELS AND COMPUTER ANALYSIS OF THE FUNCTION OF THE CENTRAL NERVOUS SYSTEM[1,2]

By W. Ross Ashby

Departments of Electrical Engineering and Biophysics
University of Illinois, Urbana, Illinois

The advent of the big computer has so revolutionized our approach to the central nervous system and its functions that this review needs an introduction. It is not merely that we have made advances in our studies of old problems: today our problems are of new types, for we have found that the old problems were often not only badly chosen, but rested on assumptions now known to be false. By "mathematical models", for instance, we mean far more than was understood twenty years ago, when the phrase referred, essentially, to what a man could express in a dozen or so equations, on not more than a dozen sheets of paper. How the idea is to be extended will be considered below; here I wish only to make clear that this chapter, to fulfill the spirit of its title, must depart somewhat from its letter. Its subject will be the interactions between what we know of computers and what we know of the brain. I hope to show that the subject, though seeming diffuse, has in fact a strong internal coherence because both systems handle information in unusually large quantities. The subject today is dominated by these quantities.

In discussing "quantity of information" I shall make use of the thesis [due to R. B. Banerji (1)] that "information theory" is essentially just a branch of combinatorics, ultimately of mere counting. Its philosophy is no more mysterious than: "You can't get ten maneuvers out of that satellite when you've only five signals." Where the technicalities enter is that perhaps the signal-emitter sends a continuous wave form, which is itself being corrupted by noise. The plain man, who can count ten signals, is then unable to say whether this emitter is worth more or less than ten, and technical methods are necessary if he is to be able to say just what it is worth. In this spirit I shall use the words "quantity of information" to refer mostly to the more everyday "number of effective or distinguishable causes", leaving the technicalities to be invoked only if they are necessary.

The chapter has been written as a review, not as a catalog. Much excellent work has not been mentioned by name here, for I have referred to work

[1] The survey of the literature on which this review is based was concluded in May 1965.

[2] Supported partly by the United States Air Force Office of Scientific Research, Grant 7-64, and partly by the National Institutes of Health, Grant GM 10718.

only as it directly illustrates the theme. The last section, on Further Reading, will take the reader to any necessary detail.

The Big Computer—how much difference will be made by its advent may well be a matter of opinion. I rather lean to that of Martin Shubik (2) who suggested that its impact may eventually be as great as that of the telescope, opening up entirely new worlds of fact and idea (after we have learned how to use it appropriately). Be that as it may, the mathematical resources of twenty years ago are today vastly extended, in directions usable by the physiologist. Today "mathematics", to the practical physiologist, can mean any of the following.

(*a*) There is, of course, the "classic" mathematics, in which a few variables and a few convenient and elementary functions are manipulated over a few sheets of paper. This type of mathematics is used by the practical worker because it provides him with a process that can be carried out expeditiously and that is isomorphic (to a useful degree) with his main interest. But the expedition and the isomorphism can today be provided by other processes, all capable of doing what used to require the formal mathematical model. Today he also has available the big computer.

(*b*) The big computer, which can actually compute, and present graphically, trajectories that involve far more variables and conditions than ever can be handled by the classic method. (Whether the computer is digital, or analog, or a mixture, or a specially made form is here irrelevant.)

(*c*) Also possible is the model in hardware, all successors to T. Ross' (3) pioneer "maze-runner", built in 1938. [A list of models is given in (33).]

(*d*) Stochastic forms are as tractable today as the deterministic, for computers can generate random numbers and explore any well-defined process, regardless of whether the classic mathematics has yet managed to find a formal, algebraic solution.

If, by "mathematics", today's physiologist includes these later resources, he is incomparably better equipped to undertake a "mathematical physiology" than was the worker with pencil and paper. At last he can use dozens or hundreds of variables, in numbers truly appropriate to the actual richness of his systems. At least equally important is the fact that instead of being forced, as he was, to replace every actual function (graph) by some fictitious form (usually the linear) that was chosen solely for its suitability for algebra, he can now compute, as freely as he pleases, with the actual physiological data, with all its idiosyncrasies.

Throughout the chapter I shall use the word "model" in the broad sense, referring to any structure that is well defined, expeditious, and usefully isomorphic with some system in physiology (especially in the central nervous system).

One last word of introduction. As the chapter is directed at the central nervous system, and at its higher functions, there is no clear limit to the "height" of the functions to be discussed. On the "mind-body problem", however, and on the "problem of consciousness" (whatever those words may

mean) I shall say nothing, for I have seen no evidence that there yet exists anything worth saying. I shall attempt to confine myself to an outline of today's facts as they are relevant to complexities of observable behavior.

The Computer Model as Archive

The "model" of twenty years ago—a dozen pieces of hardware or a dozen equations—has become quite inadequate to represent the great quantities of information available in physiology today. Consider, for instance, the base of the brain, with effects showing in the vegetative, endocrine, reflex, instinctual, and behavioral levels simultaneously. Not only are there many causes and effects at work but they interact strongly, not merely once but again and again. Twenty years ago the student was expected not only to know the primary facts—A causes B, B causes C, and so on—but also to deduce, in his head, how these causes and effects would combine into longer chains. Today there are many branches in which that possibility has practically disappeared. Specialized aid has become necessary.

The new resources have not, so far as I am aware, been applied in major degree to the activities of the central nervous system, but Grodins' (4) book shows the general method and its application at the vegetative level. After a 100-page outline of the technical methods for transferring knowledge, obtained piecemeal by experiment, to the computer, he takes an actual physiological system and shows how the computer can be built up, stage by stage, till it forms a model, as accurate as the physiologist's knowledge allows, that will behave in the way corresponding to all the subsystems acting in combination. As his example, he takes the elementary (dynamic) facts about the regulation of arterial pCO_2, (H^+), and pO_2, through ventilation. These variables form a system with feedback, whose dynamic equations (difference or differential) can be obtained. (Whether they are obtained by deduction from the laws of physics and chemistry or whether they are obtained directly from empirical graphs is here irrelevant; purely empirical knowledge is here just as valid as knowledge obtained through some elegant theory.) Since, in his example, analog components happen to represent the facts conveniently, he develops an analog network that will represent these primary facts and activities sufficiently well. Had the facts been more awkward computationally—i.e., requiring very unusual functions—, exactly the same general method could have been used in a digital computer, which is totally unrestrictive. He then considers the system of the cardiovascular variables and components and puts these dynamic facts together to form a model of them. Finally, he joins the two models together to form a single system—a dynamic embodiment of the main parts of today's knowledge.

Studies such as these have demonstrated a fact long known in physics but hardly used in physiology: building models from raw dynamic facts is essentially simple. Every primary physiological fact, so far as it is not quite timeless, is ultimately a contribution to one line in a set of ordinary simultaneous differential equations. Any number of such facts accumulates to give a whole

set of equations. Deducing the system's behavior then consists in getting the computer to deduce a chain of values (or states) by the merely repetitive use of the equations. The distinction between "model" and "computation" has ceased to be significant.

Such a model (e.g. Grodins') can now become a dynamic archive. As new facts become known, they can be added to the existing store, or old ones can be corrected. The model is then available, not only to represent the current stock of knowledge about the physiological system but to give, in addition, all those other facts that are logical consequences of the direct observations.

It is also clear, however, as Grodins' book shows unmistakably, that the whole task is too much for the individual physiologist: these tasks must be undertaken by a team, with each man attending to his speciality. If the reader objects that these methods will not enable him to understand personally the processes at work, the fact is that "understanding" is a luxury to be enjoyed only when one faces a system not more complex than one can manage; that is, when the quantity of information to be handled does not exceed one's personal capacity. The systems that were faced by the physiologist in the nineteenth century were far simpler than those faced today. Today the finiteness of the information-processing capacity of the human brain is becoming, in some sciences, a dominating and limiting factor. The aim of personal understanding must, in some branches or problems, be given up. The team may know the answers, not the individual.

In the central functions of the nervous system, so far as the building of complex models is concerned, almost everything remains to be done. The day of the simple model is past. There was a time when, as many people believed that no mechanical model could manifest a "higher nervous" function, the production of any device that manifested one was of interest. But as soon as it was proved (by Shannon in 1938 for nets of relays and by Pitts & McCulloch in 1943 for nets of simple neuron-like elements) that every describable behavior could be produced, not merely by one but by an infinite number of machines, the production of yet another merely showed that the worker had not grasped the significance of Shannon's work. Today, the only model that deserves attention (unless there is a special reason—e.g. for teaching) is the model that processes more information than can be processed by the physiologist's own cerebrum.

When an advanced model, such as Grodins', has been constructed, an obvious consequence is that it can provide information, about cardiovascular events say, while an experiment is actually in progress. Its predictions can then be turned to use during the course of the experiment. This feedback method was used by Warner, Topham & Nicholes (5) during their study of the factors that act on the heart during exercise. They programmed an analog computer to model most of the events and then used it, by letting it control a cuff round the dog's aorta, to monitor and keep constant the resistance to the outflow from the heart while other factors were made to vary. In this way the computer was made to behave as an assistant of quite exceptional

skill. Though this method does not yet appear to have been used to a major degree in studies of the central nervous system, its application will probably not be long delayed.

The Computer as Laboratory

Until the last few years, almost the only way of investigating the nervous system theoretically was by mathematics of the pencil-and-paper type. This type has the essential disadvantage that the complexities increase almost explosively as the number of variables is increased, so that the really interesting cases are effectively unexplorable. The big computer, however, offers a greatly increased range, and advantage is beginning to be taken of it. Sears & Khanna (6), for instance, used it to study how excitation would spread through a large neural net; they were able to give their "neurons" properties that accorded fairly closely with the experimental facts, instead of being constrained to fit mathematical convenience. In addition they were able to consider 800 of them, so as to obtain something like a statistical view of the events. Their paper shows how the computer can be used as an experimental tool for finding what consequences various input-patterns and parameter-changes would have on the system's behavior.

Modeling the environment is also possible. Thus Reiss (7) studied how, in a simple organism, light on the retina might be coded through the nervous system to affect the muscles. He modeled not only the organism but also the environment, by specifying how the action of the muscles would move the body and eye, relative to the light, and would thus modify the position of the image on the retina. In his experiments (conducted in the computer) the environment, with its properties, was allowed to play its full part in the development of the final behavior. Thus he was at last able to consider something of the actual reciprocal action between organism and environment that characterizes real life.

When such models produce behaviors that confirm one's intuitive expectations, one has little to say. Such models, however, have already shown that one's naive intuitions can be grossly in error. An indication that this might occur had been given earlier by the mathematical investigations of Rubin & Sitgreaves (8). They considered what behaviors would be produced by the dynamic system that had been formed at random; i.e. with a transformation that was first formed at random and then held constant while the trajectory developed. They found that, in such a system, the ending of a trajectory in a large and complex cycle was the rule, not the exception. As the system was made larger, so did the probability (that any trajectory would end in a complex cycle) tend to 1. Their report shows some other surprising facts—that peculiar L- and U-shaped distributions for some of the main features tend to occur where one might expect the commoner bell-shaped distribution. (Their report, not formally published, deserves wider circulation, for the facts it presents are fundamental to any study of the generalized, random, dynamic system.) Similarly, a simple study by Ashby, von Foerster & Walker

(9) of the dynamic consequences of "threshold" showed that richly connected systems whose parts react in this way are apt to suffer from a peculiar form of instability. If the general level of excitation is low, any unit of excitation (any pulse) has small chance of joining with other excitations to form a total excitation sufficient to surmount the threshold, so the general level tends to fall even lower. (The converse is true for high levels.) Thus deviations from a central level are self aggravating, and such a system will persistently tend to go either to total inactivity or to total activity.

These investigations were by pencil-and-paper mathematics, but studies in the computer have also yielded surprises. Farley & Clark (10), for instance, copied the elementary properties of the neuron into the computer, formed a net of such units, and then watched its behavior (chiefly the spread of regions of excitation) on the oscilloscope. The power of the computer as a means of "experimental" exploration was demonstrated by its revelation that certain conditions could give rise to peculiar oscillations, the waves of excitation traveling round a central focus like the sails of a windmill. The revelation, hardly to be noticed in the algebraic or numerical forms, was immediate when the computer presented the events visually.

A similar study was undertaken more recently by Walker (11), who continued Rubin & Sitgreaves' work experimentally. He took units with extremely simple transformations (to represent, as it were, the most primitive possible neuron) and then found what types of behavior would occur if numbers of them were joined into a net by random connexions and then set working. He too found that L- and U-shaped distributions occurred where naive intuition might have expected the bell-shaped. He showed, too, that the effect of reducing the richness of connexion (complete in Rubin & Sitgreaves' cases) was to increase greatly the tendency of trajectories to end at a state of equilibrium (rather than in a complex cycle). Evidently, richness of connexion can be expected to have a marked effect on the general style of a system's behavior.

Another somewhat unexpected result, found by Minsky & Selfridge (12), is the "mesa" phenomenon. They were studying the general problem of controlling a system's output by manipulations at its input; in particular, they wanted to see what happened when the system was made more complex, functionally larger. They found it tended to the extremes: either to show no response at the output (as the input parameter was changed) or to show an abrupt and extremely sensitive response.

Experimentation in the computer is thus not merely possible but may give information, e.g. about variations that do not occur in nature, that is otherwise unobtainable.

The Computer as Analyzer

One of the earliest and most obvious applications of the big computer was to the analysis of the many complex records that could be obtained from the

central nervous system, from the EEG for instance. Most laboratories are today using some form of computational assistance, and a few are able to transmit the data directly into the computer. Methods are in fact developing so rapidly and the equipment is changing so fast that the reader who wants details should go to the recent symposium, *Computers in Medicine and Biology* (13). Here I will merely review the subject generally.

Analysis by averaging, or by superposition (which elicits essentially the mode) have long been used. More recent studies (of the EEG for instance) have used methods that count the number of maxima on the oscillations and of the frequency with which the trace crosses the central line. The computer here does merely what an unskilled assistant would do in a longer time.

More penetrating are the studies (mostly on the EEG) of auto- and cross-correlations. Both examine aspects of the dynamic, richly connected system that are profoundly significant. The auto-correlation reflects the degree to which the unit's behavior depends on its past (on its "memory" in some sense), while the cross-correlation reflects the quantity of information transmitted, and its delay, between unit and unit. At present the studies are chiefly laying the factual basis on which any more sophisticated analysis must rest.

Even more penetrating are the studies applying the techniques of multiple regression and factor analysis. Only the big computer can undertake the huge quantities of information-processing that these methods require (if the number of factors is to be more than two or three). Again, most of these studies are now in progress; little more can be said at the moment than that they clearly must be done, as explorations, and that they are laying the foundations for even deeper studies.

The direction for deeper studies is fairly clear. The basic principles of information theory have established that, when the scientist attempts to get "information" from a system, he is, essentially, looking for a constraint (over the totality of possibilities). Thus, correlation, if nonzero, means that the variance from the regression line is less than the total variance; transmission, if nonzero, means that the entropy $H(x,y)$ is less than the entropy $H(x) + H(y)$; a "natural law" (e.g. of gravity) says that, of all imaginable paths, a planet will in fact take a certain one. In the past, the scientist has looked for correlations (e.g. of product-moment type) partly because of their proved usefulness but also partly because of their mathematical convenience. Factor analysis, similarly, was originally developed to search for, and identify, basic factors that combined in the simple additive, or linear, way. In the nervous system, however, one might be much more interested in identifying factors that combine as measures with threshold. Pencil-and-paper mathematics has, at present, little to offer; the digital computer, however, will carry out all the operations as easily in this way as in any other way. It can consider constraints of any well-defined form, and is therefore able to search data for constraints that may be highly meaningful to the physiologist though most

awkward to the classic mathematician. It is therefore to be expected that the newer computer analyses will be more concerned with operations, on direct physiological data, that study specifically the constraints of purely physiological significance.

"Higher Functions"—The Facts

What are the brain's "higher functions" and whether they have been modeled on computers is a subject too contentious for dogmatism. So I shall confine myself, in this section, to as purely factual a description as possible of what has actually been done, in such ways as are thought by some to be relevant to the topic. The significance of the facts will be left to the judgment of the reader.

In one respect, those who would explore the higher functions face an exceptional difficulty. Their colleagues in atomic physics and cosmology explore subjects in which our prior knowledge is nothing, so each new fact is discussed simply on its merits. Man, however, has always been sure that he knows how he thinks, so every new fact has been received into a morass of old speculations, and pseudo-facts (mostly entirely pre-Darwinian in nature), where it has commonly been regarded as a personal insult. In this atmosphere I will attempt to present simply the facts.

It was on July 12, 1962, that Samuel's Program played Mr. R. W. Nealey at checkers (draughts in Britain) and defeated him (14). Mr. Nealey—a former Connecticut checkers champion and one of the United States' foremost players—wrote afterwards: "In the matter of the end game, I have not had such competition from any human being since 1954, when I lost my last game."

A. L. Samuel, who wrote the program, thereby "making" the machine, has no claim as a checkers player. The program is constructed so that the machine forms its own methods of play, tries them in actual games, and then modifies its methods according to their outcomes. This brief description gives, of course, only the outline, as one might sum up human life by saying that it consists of the conversion of chemical energy to muscular. In fact, the program contains many sophisticated methods to ensure that the information obtained by the machine is turned to advantage with high efficiency. Once the program was launched, its development (of its methods of play) has progressed independently of Samuel, and (unless someone looks into the machine to find what it is doing) its methods of play are now known only within the machine. It has discovered general aspects of play (e.g. the value of computing the second moment about the diagonal between the double corners) that were previously unsuspected.

[If the reader is confused at this point between thinking of the computer as an absolutely stupid slave that has to be told everything and the computer as a highly sophisticated and capable agent, he should decide whether he is thinking of the computer as it is initially (as it leaves the factory, say) or whether he is thinking of it already supplied with an elaborate and ingenious program. Throughout this chapter, by "computer" I shall usually imply "armed with an appropriate program".]

Essentially parallel in principle with Samuel's program, though very different in its way of manifesting the principle, is the machine (or system, or method) MENACE, made by Michie (15) to play Tic-Tac-Toe (Noughts and Crosses in Britain). It, too, is given no details of play by Michie, but develops its method of play in accordance with its experience. It progresses steadily towards the faultless game.

In the more complex game of chess, the earlier explorations have been surpassed by Samuel's achievement, but Simon & Simon (16) have formed a program for the process of arriving at "mate" (when it is not too distant). They have shown, by tests on actual positions from classic games, that their program can be as successful as the grandmaster, even when he has achieved a well-known brilliancy. They also show (I record merely the fact here) that certain well-known oversights made by grandmasters (failing to find the quickest way to mate) were also made by their program.

Programs that will produce mathematical theorems, with a rigorous proof, now exist in several forms. Gelernter's program (17) proves the elementary theorems in Euclidean geometry with approximately a schoolboy's skill. It has, however, produced a proof of the *pons asinorum* that the reader may find impressive.[3] The machine cannot claim priority since the proof seems to have been known to Pappus, but the machine discovered it independently.

This proof was obtained by the manipulation of symbols by rules, in the algebraic manner. Evans (18) has shown that if the topology of ordinary Euclidean space be given to the computer, processes that use geometrical analogy, such as the use of diagrams, can also be used by the machine.

The proving of algebraic and logical theorems is also now a well-established possibility. As early as 1957 the program of Newell, Shaw & Simon had proved the first 38 theorems of Russell & Whitehead's *Principia*. [The later GPS program of Newell & Simon (19) is much more powerful, but it has been used so far only in other types of research.] Using more modern

[3] It goes as follows. Given that the triangle ABC is isosceles, with AB = AC, prove that the angles at the base, $A\hat{B}C$ and $A\hat{C}B$, are equal. The machine simply used Euclid's immediately preceding theorem ("two sides and the included angle") by applying it to the two triangles BAC and CAB, treating it as irrelevant (as it is) that they are derived from a single original. The Proof:

ΔB A C		ΔC A B
B A	equals	C A
A C	equals	A B
B $\hat{A}$ C	equals	C $\hat{A}$ B

Therefore the two triangles are equal, and corresponding angles are equal. To $A\hat{B}C$ corresponds $A\hat{C}B$ (Q.E.D.).

methods, which are much more efficient, Wang's program (20) has completed the earlier work by dealing with all those theorems in the *Principia*, nearly 400 of them, that use the predicate calculus: it has found proofs for all.

Another process that has been successfully transferred to the computer [Slagle (21)] is that of finding indefinite integrals; e.g. removing the integral sign from

$$\int \frac{x^4}{(1-x^2)^{5/2}}\,dx$$

The program now carries out such integrations about as well as the average mathematician who is not a specialist.

In pattern-recognition, the program of Greanias et al. (22) recognizes numerals, handwritten by briefly trained subjects, with an error of about 0.1 per cent. Uhr & Vossler's program (23) develops its own criteria from samples given; in a competitive test against human subjects (on arbitrary patterns to minimize these subjects' lifelong experience with numerals and similar shapes) it performed appreciably better.

Without comment is offered the fact that the "Illiac Suite for String Quartet" (24) was produced by the Illiac computer on a program by Hiller & Isaacson (25). The program, of course, specifies only trends, not details.

To keep the review to a suitable length, other achievements in the last decade must be passed by. The loss is not serious, for what we once thought were scientific mountains have sometimes proved to be mere molehills. The "machine that changes its own behavior", for instance (sometimes called an "adapting" machine), is today well understood as simply the system that consists of two subsystems interacting, seen from a particular point of view. Any machine that loses information is a pattern-recognizer; whether the pattern that it recognizes is of interest to any sponsor is another question. The work of the last decade is helping us to discover which are the real problems and which are mere survivals from the questions that agitated the dark ages.

Any idea that the achievements were obtained by exploiting the computer's mere size or speed must be rejected. It was precisely the tendency to rely on mere size and speed ten years ago that made the first essays in these functions so incompetent and unproductive. Since then a great deal has been learned and new methods have been invented or discovered. What they are will be outlined in the next section.

"Higher Functions"—The Methods

How were these behaviors achieved? One way of answering the question is simply to display, *in extenso*, the machine's circuitry and the punched cards of the program, for undoubtedly the answer is contained here. But this answer is no help to the worker in another subject, who has no use for the details, but only for the principles. Are there principles of use to the physiologist?

In addition to the technicalities of the particular computer, today's programmer knows a great deal about machine, control, and coordination that was not known ten years ago. Much of this knowledge has not yet been formulated explicitly and generally, and the book that will display the general principles to the physiologist has not yet (so far as I am aware) been written. Gill's *Introduction to the Theory of Finite-State Machines* (26) is written especially for the mathematician and programmer, but it shows so clearly the nature of the new knowledge that I will abstract it briefly, so that the physiologist can judge whether or not it is likely to help him.

It is primarily epistemological, studying how an experimenter can obtain knowledge about the system that lies before him. It is assumed that the system is subject to the physical rule (true, so far as is known, at all levels above the atomic), that the system's next state is determined completely by its present state and the conditions of its surroundings (here, of its input). It is assumed that he can do things to it (through its input) and can record what happens at its output, but it is also assumed that the system is large, so that not all of what happens in it can be observed directly. Gill then considers what, in principle, is discoverable about the system by comparison of the input-records and the output-records.

One of his questions he calls, by an obvious medical analogy, the "diagnosis" problem: Given a system whose construction is known but whose present state is unknown, to specify a study that shall enable the present state to be deduced—or prove that no such study exists. To attack this question he divides the modes of experiment (the sequence given to the input) into (*a*) the *pre-set*, in which the experimenter declares in advance the whole of what he will do; and (*b*) the *adaptive*, in which he conducts the experiment in stages, modifying the later stages according to what he observes at the earlier. Gill also distinguishes the *simple* experiment (e.g. on a human subject) from the *multiple* (e.g. on rats) in which each subject may be destroyed by the experiment, to be replaced by another version of the original. He proves rigorously that there exist cases in which the diagnosis problem is essentially unsolvable by any pre-set experiment but which are solvable by a suitably adaptive experiment. He gives an example (p. 107) to show that this absolute necessity for experimentation of the adaptive type may occur with quite a simple system. He also shows that if the experimenter is restricted to the simple experiment (subject not expendable) the diagnosis problem may be unsolvable; i.e., there may not exist any procedure, pre-set or adaptive, capable of eliciting the required information. And he then shows (p. 113) that if the experimentation is not restricted to the simple, there always exists at least one pre-set experiment that will elicit the desired information.

He also considers the "homing" problem, one of control: Given a system whose construction is known but whose present state is unknown, to specify a process at the input that will bring the system to some state that is known. This problem proves to be closely related to that of "diagnosis", but there are differences. Thus the homing problem can always be solved by simple

(nonexpending) experimentation, but the diagnosis problem cannot always be solved under this restriction.

So far it has been assumed that the system's construction is known; a problem of high interest to the physiologist is that of "machine identification"—by providing input and observing output to deduce the construction. On this question he proves a number of theorems on what is achievable; in this way one can identify what are the features that are relevant. The topic becomes rather too detailed for further summary, but his results show clearly that experimentation on complex systems has its own science. There is a theory of experimentation that must be understood when the systems studied are complex, if the experimenter is not to run the risk of finding that his work is misdirected.

As an example of how easily one's naive ideas may go wrong when one faces the complex system, I may quote his proof on p. 156. Many people (including myself) have considered as plausible the idea that a finite system, such as a cat's brain, must have some finite duration for a memory. Gill destroys this loose thinking by showing that, however small the system (provided it has at least two states!) there is logically no limit to the distance back in time at which an event may not be significant for present behavior. Clearly, today's knowledge of what is meant by machine, control, and process includes altogether new insights; it is these insights that have made possible the achievements described in the earlier section.

By now, the neuronologist may be uneasy at my neglect of the neuron, but the neglect is only apparent. If the words used above suggested the study of the whole brain, the fact is that the neuron is itself a complex system, with 10^{17} or so molecules, to be studied by experiment and observation, so all the theorems are equally valid for it. All that is necessary is that the words "state", and a few others, be re-interpreted.

In another sense, however, the neuronologist's comment is valid. The advance of knowledge has revealed the enormous intellectual gap between the behavior of the nerve cell and that of the whole person. In physics it is quite accepted that a skyscraper is not to be designed by deduction from the principles of atomic physics, though the skyscraper admittedly stays up by the action of interatomic forces. A chain of sciences—molecular physics, crystallography, metallurgy, strengths of materials, structural engineering—is necessary to span the gap. What is being done in computers demonstrates some of the intermediates that may be necessary in neurophysiology. As a particular example, consider Samuel's program. On what units the program depends physically—vacuum tubes, transistors, relays, etc.—is, for us, simply irrelevant; the significant features in his work are at quite another level. Thus, one of the important problems at his level was: after a game had been lost, where should the correction be applied and in what degree? If too vigorous, the correction, though in the right direction, might be excessive, and lead to play even worse (in the opposite way) than before; if too small, the process might take too long to register a significant improvement. At Samuel's level the question is: where is the optimum?—the componentry is not visible.

Another example of what is significant at this intermediate level concerns information storage. The method of Zato-coding (27) is used commercially, but its principle, which replaces an extremely inefficient method of storage by one of far higher efficiency, may well also be used by the nervous system. Its method may most simply be explained by following an imaginary example. Suppose the State Department receives letters and wishes to file them, when each refers to some 4 topics out of a possible 10 million. One way is to take, for each letter, a card with 10 million sites on it and then to punch the card at the corresponding 4 sites. But with even a square millimeter to a site the card would measure about 5 by 15 feet, and the way is impractical. Zato-coding uses the fact that the number of ways of selecting 4 out of 10 million is approximately equal to the number of ways of selecting 46 out of 92:

$$\binom{10{,}000{,}000}{4} = \binom{92}{46} = 10^{26.6} \text{ approximately}$$

If now we allot each topic some 12 sites (they may be selected at random) from a card of 92 sites (and this can be done, for there are $10^{14.5}$ ways of taking the 12, of which we need only 10^7), then each letter, with its 4 topics, will result in (a little fewer than) 48 punch holes, in the card with 92 sites. Thus, the change to a more efficient method has shrunk the demand for sites from 10 million to less than a hundred—from a card the size of a carpet to one smaller than a postage stamp! The method carries with it the cost of the more complex coding and also the possibility that, by chance overlap of the 12's, a letter is recorded as referring to a topic on which in fact it says nothing. If these costs are acceptable, the method illustrates the enormous difference that very commonly exists in information processing between the obvious method and the sophisticated.

Another discovery of the last ten years that has enormously increased the effectiveness of these processes is the discovery that processes of search (for some known goal) should be conducted, not only in the obvious direction but also from the goal backwards. The reason lies in an elementary combinatorial fact, whose significance went for a time quite unnoticed: the gain is not a mere halving—it cuts the number of operations to approximately its square root. Searching processes are at the core of most of the achievements described above, but searching processes are only too apt to demand quite unpractical numbers of operations (as we shall see later), but it is just when the number is really large that the change of method is most effective. 10^{20} operations, for instance, even at a microsecond apiece, demand 3 million years. 10^{10} operations, however, can be done in a morning. Thus, the improvement in method here converts the completely unpractical to the easily achievable.

Closely related is the discovery by Newell, Shaw & Simon (28), after

working for some years on proving theorems (in the computer) by deduction from the axioms, that this direction, so readily taken over from the books, is in fact far surpassed, for speed and efficiency, by the method of guessing the theorem and then working back to find any set of axioms that will justify it. ". . . the efficacy of working backward may be analogous to the ease with which a needle can find its way out of a haystack, compared with the difficulty of someone finding the lone needle in the haystack." Here again the improvement is commonly by great orders of magnitude, converting the practically impossible to the readily achievable.

A final example will make clear why it could be said above that the achievements of the last ten years were due to new methods rather than to the brute size and speed of the computers. Simon & Simon's mating program (mentioned earlier as having produced results equivalent to grandmaster play) was in fact tested by mere pencil-and-paper simulation. Once the method was well understood, and the inefficient method (the source of the huge demands) replaced by one of far higher efficiency, the huge capacity and lightning speed became unnecessary. Evidently there are principles of information-processing that would interest the student of the higher cerebral functions

Quantity of Information

It will be seen from this survey that we are now far from the conceptual worlds of Sherrington and Pavlov. What has happened, in cybernetic terms? The change is essentially quantitative.

Until recently, all branches of science lived and grew by exploiting the many systems—in physics, chemistry, biology—in which the amount of interaction between the parts was small. Newton's mechanics was applied mostly to only two bodies at a time; chemistry succeeded best with molecules of only a few atoms; physiology seldom considered the interactions of more than two or three variables simultaneously. In the last decade, however, technical advances have made feasible the direct study of systems with far more than a few; the result is that the quantity of information involved in a modern study may be larger than the quantities considered earlier by great orders of magnitude.

The question of these quantities, and just how big they are, has recently become specially insistent. Bremermann (29) has shown that, because matter is inherently coarse grained, it cannot transmit information faster than 10^{47} bits per g per sec. The restriction applies equally, of course, to the big computer and to the physiologist's cerebrum. The number may seem large, but anyone who starts to estimate the quantities of information that may be involved in the topics of this chapter will soon encounter numbers so vastly greater as to make the limit extremely restrictive. To get a sense of proportion, let us glance at an example. Suppose a sense organ has a million receptors, each of which can be in one of only two states, and that an effector organ has a million units, each also of only two states, and that we wish to

identify or consider the possible ways in which they may be related (to be specific, the possible mappings from the states of the sense organ to the states of the effector). Since the number of mappings is the number of output states ($2^{1,000,000}$) raised to the power of the input states (also $2^{1,000,000}$), the quantity of information involved in the specification of any one mapping is easily found to be no less than $10^{300,000}$ bits—a quantity beside which Bremermann's limit is little more than an infinitesimal. (Taking tons of computer and centuries of time merely adds a few units to the 47 and makes practically no difference.)

Even this elementary example shows a fact that seems to have escaped notice. The number of mappings (ways of coding sensory states to motor states) is of the form m^s, where s is the number of sensory states and m the number of motor. Now when these two numbers are large, the exponent s is enormously more powerful as an increaser than is the base m. It follows that considering extra combinatorial complexities at the sensory side leads to a far bigger increase (in the quantity of information that the observer faces) than that caused by an equal change at the motor side. Again, to get a sense of proportion, suppose that the observer who was studying the system above (in which 2 states at each sensory element led to $10^{300,000}$ bits) decided to consider the possibility of nonlinearity in the elements. For "nonlinear" to be meaningful he must now consider at least 3 states at each sensory element. It is easy to verify that the information facing the observer has been jumped, by this innocent-looking suggestion, to $10^{477,000}$ bits. Thus, adding the possibility of nonlinearity (to an already complex study) has not multiplied the quantity by the plausible 3/2—it has multiplied it by the factor $\times 10^{176,000}$!

Studies of information-processing at the sensory side have been made by many workers in the last decade, especially by those studying learning, for the method offers a way into the intricacies of the nervous system. But even the roughest approximations are sufficient to show that the quantities of information demanded by the study may easily go far beyond the resources of both computer and cerebrum.

What classes of study are most apt to generate these huge quantities? Essentially they are those generated by combination—by what happens to one variable being *conditional* on the value of some others, or by *components* combining to form some compound event. As a practical rule, one may expect an explosive increase in the quantity of information if the study uses any of the words

Net	System
Assembly	Subset
Organization	Property
Pattern	Relation
Interaction	Order
Coordination	Mapping

A glance shows that these are just the words that commonly occur as soon as any attempt is made to discuss the "higher" functions of the brain. It is to be expected, therefore, that studies of these functions will demand that the quantities of information be watched continuously, to make sure that the worker is not attempting the impossible.

As a last example to illustrate the peculiar tendency of these combinatorial actions to increase the quantity of information explosively, we can consider the idea—put forward by authorities no less eminent than Eddington (30) and Piaget (31) and taken seriously for a time by the writer—that perhaps the brain reaches the higher relations by first taking the primary sense data, then forming the primary relations between these elements, then forming the secondary relations between the primary, and so on. But what happens quantitatively? It is easy to verify that the number at each level is an exponential function of the number at the lower (to some small-number base). Thus the attempt to explore the whole structure leads to an ascending scale of exponentials!—

Bremermann's limit vetoes this suggestion immediately. Nothing made of matter can execute such a process to a major degree.

The import of these observations is that they show that the problems faced by the brain are not to be solved by a merely multiplicative increase in resources. To go a million times as fast, for instance, in the attempt to process $10^{300,000}$ bits, is to shorten the units of time from that number to $10^{299,994}$—in this context, no effective change. The only way is by a change of *method*. At the heart of the modern studies of "mechanical brains" lies the question: what methods can be used (by computer or living brain) to evade the avalanche of information that threatens as soon as we consider the system whose parts are allowed to interact freely?

The problem has had to be faced in evolution, for the world's parts interact to more than negligible degree, and the brain's parts must interact among themselves if they are to form a whole complex enough to adapt to complex environments. As evolution solved the aerodynamic problems of the wing, it has probably solved the information-processing problems sketched above. What has been found by computer studies in the last ten years thus suggests that one central problem of the "higher" functions in the brain is to discover its methods for processing information. The methods we know today tend to be inefficient to "astronomical" degree: the living brain may well know better.

FURTHER READING

The specialist will already know his own sources. Here I would merely suggest how the reader may best enter the subject.

Outstanding is the recent anthology *Computers and Thought* (32). Editors Feigenbaum and Feldman have collected the best papers of the last twenty years and have included only those that were: (*a*) written as reviews (so as to be easily readable); and (*b*) written by the workers themselves (so as to be authoritative). All the papers are written with sufficient technical detail for the reader to be able to follow the arguments with critical insight, yet are essentially capable of being read straight through (if the reader will change down into bottom gear at times). It concludes with Minsky's very fine Bibliography (33), containing practically every significant contribution from about 1940 to 1962, classified so that the reader can soon find whatever he wants.

Information about what is happening in the physiological world is given extensively and almost completely in the symposium *Computers in Medicine and Biology* (13), in which the emphasis is on the biological side. The reader who wishes the emphasis to be on the computer side should read the collection of articles *Computer Applications in the Behavioral Sciences* (34). These three books omit nothing of importance.

As a broader source, *Behavioral Science* has, since its inception in 1956, been closely associated with our topic. It includes a regular section "Computers in behavioral science" in which a great deal of ancillary information is to be found.

For more general (and more speculative) reading, the Yearbooks of the Society for General Systems Research (35) may be mentioned. Being reprints, they have all passed the elementary test of seeming worth reprinting. Their quality is uneven, however, and they are suitable only for the reader whose critical faculties are alert.

LITERATURE CITED

1. Banerji, R. D. (Personal communication)
2. Shubik, M. In *IBM Sci. Symp. Simulation Models and Gaming*, Dec. 7–9, 1964 (In press)
3. Ross, T. *Psychol. Rev.*, **45**, 185 (1938)
4. Grodins, F. S., *Control Theory and Biological Systems* (Columbia Univ. Press, New York, 1963)
5. Warner, H. R., Topham, W. S., and Nicholas, K. L. In Computers in medicine and biology. *Ann. N. Y. Acad. Sci.*, 115, 669–79 (1964)
6. Sears, R. E., and Khanna, S. M. In *Proc. Joint Computer Conf.*, **24**, 15–24 (Spartan Books, Baltimore, Md., 1963)
7. Reiss, R. F. *Behav. Sci.*, **5**, 343–58 (1960)
8. Rubin, H., and Sitgreaves, R. Probability distributions related to random transformations of a finite set. *Tech. Rept. No. 19A* (Appl. Math. & Stat. Lab., Stanford Univ., Stanford, Calif., 1954 [ASTIA Rept. AD 27350]
9. Ashby, W. R., von Foerster, H., and Walker, C. C. *Nature*, **196**, 561–62 (1962)
10. Farley, B. G., and Clark, W. A. In *Proc. London Conf. Information Theory, 4th*, 242–51 (Cherry, C., Ed., Butterworths, London, 1961)
11. Walker, C. C. *A Study of a Family of*

Complex Systems (Doctoral thesis, Univ. Illinois, Urbana, Ill., 1965)
12. Minsky, M. L., and Selfridge, O. G. In *Proc. London Conf. Information Theory, 4th*, 335–47 (Cherry, C., Ed., Butterworths, London, 1961)
13. Tolles, W. E., Ed. Computers in medicine and biology. *Ann. N. Y. Acad. Sci.*, 115, 543–1140 (1964)
14. Samuel, A. L. In *Computers and Thought*, 71–105 (McGraw-Hill, New York, 1963)
15. Michie, D. *Science Survey*, Part 2, 129–45 (Penguin Books, London, 1961)
16. Simon, H. A., and Simon, P. A. *Behav. Sci.*, 7, 425–29 (1962)
17. Gelernter, H., Hansen, J. R., and Loveland, D. W. In *Proc. Western Joint Computer Conf.*, 17, 143–47 (1960)
18. Evans, T. G. In *Proc. Joint Computer Conf.*, 25, 327–38 (1964)
19. Newell, A., Shaw, J. C., and Simon, H. A. In *Self-Organising Systems*, 153–89 (Yovits, M. C., and Cameron, S., Eds., Pergamon, New York, 1960)
20. Wang, H. *IBM J. Res. Devel.*, 4, 2–22 (1960)
21. Slagle, J. *A Computer Program for Solving Problems in Freshman Calculus* (Doctoral thesis, Mass. Inst. Technol., Cambridge, Mass., 1961); In *Computers and Thought*, 191–203 (McGraw-Hill, New York, 1963)
22. Greanias, E. C., Meagher, P. F., Norman, R. J., and Essinger, P. *IBM J. Res. Devel.*, 7, 14–21 (1963)
23. Uhr, L., and Vossler, C. In *Computers and Thought*, 251–68 (McGraw-Hill, New York, 1963)
24. (Theodore Presser Co., Bryn Mawr, Pa.)
25. Hiller, L. A., and Isaacson, L. M. *Experimental Music* (McGraw-Hill, New York, 1959)
26. Gill, A. *Introduction to the Theory of Finite-State Machines* (McGraw-Hill, New York, 1962)
27. Mooers, C. N. *Aslib Proc.*, 8, 3–22 (1956)
28. Newell, A., Shaw, J. C., and Simon, H. A., *Proc. Western Joint Computer Conf.*, 15, 218–39 (1957); In *Computers and Thought*, 109–33 (McGraw-Hill, New York, 1963)
29. Bremermann, H. J. In *Self-Organising Systems 1962*, 93–106 (Yovits, M. C., Jacobi, G. T., and Goldstein, G. D., Eds., Spartan Books, Washington, 1962)
30. Eddington, A. S., *Philosophy of Physical Science* (Cambridge Univ. Press, Cambridge, 1939)
31. Piaget, J. *Logic and Psychology* (Manchester Univ. Press, Manchester, 1953)
32. *Computers and Thought* (Feigenbaum, E. A., and Feldman, J., Eds., McGraw-Hill, New York, 1963)
33. Minsky, M. L. *IRE Trans.* HFE-2, 39–55 (1961). In *Computers and Thought*, 453–523 (McGraw-Hill, New York, 1963)
34. *Computer Applications in the Behavioral Sciences* (Borko, H., Ed., Prentice-Hall, New Jersey, 1962)
35. Secretary-Treasurer: Milton D. Rubin The Mitre Corp., Bedford, Mass.

BRIT. J. PSYCHIAT. (1968), 114, 1485-1498

THE CONTRIBUTION OF INFORMATION THEORY TO PATHOLOGICAL MECHANISMS IN PSYCHIATRY

Information theory is sometimes presented as a new philosophy; here it will be presented as an essentially practical branch of science. Its essence occurs when an engineer says: "You can't get 10 manoeuvres out of that satellite when you've only five signals." He is thinking along the usual and well understood lines of cause and effect, but using a rather unusual approach: instead of trying to relate each cause to its particular effect (e.g. "what is *the* cause of tuberculosis?"), he is bringing a *set* of (five) causes into some relation with a *set* of (ten) effects. Throughout this article, information theory will be used in accordance with what I believe to be its true nature – that it is the body of knowledge developed to help when we have problems in which large numbers of causes are related in some way to large numbers of effects.

Problems that involve causes and effects in large numbers fall into two natural classes: 1. where the various causes are distributed over space (or equivalent dimensions), as, for instance, *patterns* of lightspots distributed over the retina, and 2. when the various causes are joined in long chains in time, each effect becoming the cause of the next. This second case corresponds to the activities of the modern computer, which might be described simply as a device for performing accurately throughout an extremely long chain of causes and effects. For this reason, computer-theory and information-theory are hardly separable, and I shall refer to both freely. Reference to both is specially necessary in this article, for the brain shows the double complexity of accepting, through the senses, complex *patterns* of stimulation, and then carrying them through long *chains* of processes. Modern theories of information and computers attempt to say something useful about such broad and lengthy processes.

It should perhaps be noticed at this point that these theories, of information and computers, are entirely objective in their methods. Though words such as *information, memory, control,* and *recognize* have an introspective aspect (practically the only aspect considered in the psychologies of the previous century), they are used in these theories of today (and in this article) solely to refer to objectively demonstrable facts of behavior. Thus, the geneticists and molecular biologists today speak freely of the "information" on the DNA molecule: this "information" has no reference, of course, to any "knowing" by the DNA: it is simply a reference to the various causes, exertable by the DNA as a physical system, over the various effects that can be shown on, say, the proteins synthesized.

The theories of information and computers are thus essentially concerned

with general principles that should hold, or that should guide one, when one has to deal with a system in which the processes at work are extremely complex and lengthy. For this reason it seems that psychiatry must inevitably be related in some degree to these theories. Is anyone more likely to say "this complexity is becoming unmanageable" than the psychiatrist? I need therefore offer no apology for attempting to trace some relation, to explore the borderline, between these two sciences.

When the complexity is lacking – when one relates, say, ten manoeuvres to five signals – the theories of information and of computers may give little help, for the worker needs no help. They tend to become increasingly useful as the complexities grow. When does this happen? – the work of Shannon and Weaver (1949) has shown that the main factor leading to the complexities referred to here is *combination:* where many parts have relations to one another. Thus, these theories are likely to be useful whenever we encounter such concepts as:

an organization (of units);
a net of neurons;
a society of persons;
a system of parts;
interactions between parts;
co-ordination of parts to a goal;
integration of parts to form a whole;
a pattern, gestalt (of units).

It is thus just when dealing with the higher functions of the brain that these theories may be specially useful.

Memory as "transmission"

If these theories are to be used, however, one must learn to see some old phenomena from a new angle. As illustration, consider the subject of "memory". A century ago it was defined as the power of evoking past images, or by some similar phrase that appealed essentially to introspection. The man in the street today (and the beginner-student) still tends to think of "memory" in that way. Psychologists, however, long ago found that to study the subject one must treat it as a phenomenon of correlation between past events and present behaviors. Now "cause and effect" may well have the cause in one place and the effect in another (e.g. closing a switch here lights a lamp there). If a change of dimension is allowed we may consider a cause now as having an effect later in time (e.g. setting the alarm clock now will make it ring tomorrow). With this approach, the phenomena of memory (in its objective aspects) can be treated by the theories of information and computers when we regard "memory" as corresponding simply to the existence of demonstrable "transmission" (here understood as "having a correlation") between events that were appreciably separated in time. To illustrate the theme, I will give two examples to show two sentences, one that has "memory" for exactly one word back (so that adjacent pairs of words are related) and the other for exactly three words back (related in fours). The particular span was ensured by the method of generation, which was as follows (in the three-word case):

Three words were written to get started. These were then shown to a person who was asked to add one word that would be related naturally to the visible three. The first word was then covered, and the remaining three words were shown to another person, who was asked similarly to add a natural next word. Then the second word was covered, a third person asked, and so on. In this way each word could be related directly only to just three previous words. (The one-word case differed by having only one word visible as each was added.)

Here is the sentence with each word related to only *one* word back:

> ***paper bag of the time which was coming to be friendly atmosphere was never forget when suddenly he stretched himself and all swollen neck line drawing water hole below deck playing games played with effect on him with only you are well within the nature boy and believing that was already for another pair of pay as before long . . .***

The triple " . . . line drawing water . . ." shows clearly how "drawing" follows as "line drawing", and "water" as "drawing water"; but "water", after "line drawing" shows that, at the moment when "water" was selected, "line" was playing no effective part in the selection.

And here is the sentence with each word related to just three words back:

> ***the costume had holes in my socks I forgot to remind the writer to tell her who would not like the enemy who ran wildly when who should come in handy regardless of the crowd and cameraman seemed pleased with me when suddenly it exploded with great force of gravity was getting lower and lower until at last it gave . . .***

Here the phrase ". . . exploded with great force of gravity . . ." shows that when "gravity" was selected, "force of" was obviously operative, but "exploded" was clearly not operative.

Such examples, as a "synthetic psychosis", show how "memory", as an objective fact showing in behavior, can be treated without any reference to its introspectional aspects. They suggest also that the application of these methods may give insight into some pathological mechanisms.

Determinate or probabilistic?

Another question that has been much clarified by the modern studies of computer theory is whether the brain may be regarded as a machine; in particular, whether it should be regarded as determinate or probabilistic. Shepherdson's recent survey (1967) shows how thorough is today's understanding of just what is implied by the idea of "machine". The experience of the last twenty years has shown that, apart from mathematical subtleties, all the various attempted definitions of "machine" prove to be practically identical, even though the various workers have started from very different branches of science. All have tended to the various forms of "semi-group": a set of states (the machine) and an operator (its laws) such that unlimitedly repeated action by the operator on the states cannot generate a state outside the set. All the usual ideas about machines either lead to this definition or are derivable from it. What has emerged is that the definition is largely indifferent to whether the operator is determinate (the ordinary "law") or probabilistic (with Markovian transitions). And it has been shown (e.g. Shepherdson, 1967) that the distinction between them is, in essence, small. What either type of machine can do, the other can do. This fact makes the writing of this article rather simpler: by referring only to the determinate forms, I shall in fact be including most of what is worth saying of the probabilistic forms.

Recent Advances

The theories of information and computers are, as I have said, essentially those of causes and effects when they occur in great numbers. One way of approaching the application of these theories to pathological mechanisms in psychiatry is to look first at what these theories have led to, for here we can see the theories actually at work. I will therefore review their main achievements over the time since my last review in this journal (Ashby, 1958), omitting those that are hardly applicable to psychiatry.

Perhaps the outstanding event, in its philosophical and theoretical importance, occurred on 12 July, 1962, when the Connecticut champion of draughts ("checkers" in the U.S.), R.W. Nealey, was defeated by the computer programmed by A.L. Samuel (1963). The point is that Samuel has no major skill at draughts: he programmed the machine to develop its own strategies, on the basis of its own experience. The process at work was in no way mysterious; the machine was told: use a random generator to suggest random strategies, test them (either in actual play or against published games by masters), keep those that lead to success, and reject or modify (again at random) those that lead to failure. The process is abstractly identical with that of natural evolution: mutation, with preservation of the better and rejection of the worse leading to even better skill against Opponent Nature. Samuel demonstrated that this process is intrinsically capable of generating the skills involved in draughts-playing: his work lay chiefly in ensuring that the process, as it actually occurred in the IBM 7090, did not excessively waste the resources available.

Studies such as Samuel's, aiming at what may be called "artificial intelligence", have shown (e.g., Feigenbaum and Feldman, 1963) that the fundamental principles of such processes are often basically quite simple. Why they have not been developed faster, in industry say, is because a great deal of selection within the details of the process is necessary if it is not to be of abysmally low efficiency. (One may remember here that the *principle* of the steam engine was known to the Greeks, but it took many centuries for its efficiency to be raised to a point of usefulness.) Thus the studies of today are turning from the question of the principle (which seems to be often simple) to the question, of much greater difficulty and practical importance, of what factors and methods will raise its efficiency. As Minsky (1963) puts it: 'The real problem is to find methods which significantly delay the apparently inevitable exponential growth of search trees."

As was said above, it is precisely when the process is richly combinational that the demands for information-processing are most apt to increase excessively. Prominent among the methods for reducing the demands (i.e. for increasing the efficiency) is that of breaking the process into stages. The importance of this method has been strikingly confirmed by the work of Simon and Simon (1962). They took, in the game of chess, the problem of divising a computer programme that should work through the final moves to mate. They looked for a process that could be specified by a relatively small number of rules and that would then show relatively great power in the terminal game. They showed that a goal (e.g. mate in eight moves) that might be utterly unachievable by its complexity if the process attempted the whole analysis might be quite readily achievable if the final goal could be reached via a few intermediate ("sub") goals. They developed a number of simple rules relating to such sub-goals (e.g.: give priority to a check that adds a new attacker to the list of active pieces), and showed that such a process was capable of paralleling many of the "brilliancies" in the recorded literature. In fact, the improvement in efficiency was so great that the huge modern computer was hardly required: mere hand-simulation with pencil and paper was sufficient in many cases. They also discovered the interesting fact that where their programme showed weaknesses, apt to overlook the best move, some of the historic mistakes of master play had consisted in overlooking exactly the same move. It seems possible that the cerebrations of master play may be carrying out a process similar to that specified in their paper. But in any case, their work showed clearly the importance of the *method* of thinking

(i.e. via sub-goals) and the relative sterility of mere speed and quantity.

Theorem-proving

Another activity commonly regarded as of "higher intellectual" form is that of theorem-proving. Here again the goal can hardly be simpler: one seeks a process, going by steps of accepted validity, to join the available axioms to the final deduction. In the last ten years some major discoveries have been made about the nature of such processes. At first almost all researchers took for granted that the process was deductive, and they devised computer programmes to perform the operations. They achieved some success: Wang's programme (1960) with the aid of a big computer, successfully proved 220 theorems in 3 minutes. But these theorems were all of simple type, and it became clear that such methods would demand utterly unacceptable times if the theorems were to become moderately complex. (Again the process that seemed obvious proved to be extremely inefficient.) It was then discovered, by Newell, Shaw and Simon (1957), that the process could be enormously speeded if it were changed from an imitation of the deductive method given in the textbooks to one of: Guess the theorem (i.e. "Is it true that . . ."), and then search among the axioms (and previously established theorems) for *any* set that justifies the result. This method, unplausible though it seems, proves in practice to be enormously superior to the other. After a discovery, of course, it is easier to see reasons, and they wrote: " . . . the efficiency of working backward may be analogous to the ease with which a needle can find its way out of a haystack, compared with the difficulty of someone finding the lone needle in the haystack." Be that as it may, the brain in evolution has clearly encountered many different *ways* of thinking: it seems likely that today our brain has developed, not merely good biochemical and electrical methods, but also some expertness in the construction of methods of information-processing, constructions that are at an entirely higher level than those of the events in the transistor or neuron.

Is the machine original?

At this point the reader may raise the question whether these chess brilliancies or theorem proofs are to be attributed to the machine or to the programmer-designer. The experience of Minsky (1967) may help to clarify the matter. He developed a computer-programme to prove theorems in geometry. One of its first productions was a proof for the *pons asinorum* that was unknown to Euclid, new to Minsky, and of high mathematical quality. It is so brilliantly simple that it can be given in a few lines. (Triangle ABC has AB = AC; prove that $A\hat{B}C = A\hat{C}B$). The machine proceeded: Compare triangles BAC and CAB!

△	B	A	C		△	C	A	B
	B	A		=		C	A	
		A	C	=			A	B
	B	$\hat{A}$	C	=		C	$\hat{A}$	B

By Euclid's immediately preceding theorem ("two sides and the included angle" the two triangles are equal, so angle $A\hat{B}C$ is equal to the angle that corresponds to it: $A\hat{C}B$ (Q.E.D.). Euclid's proof, with its extensions of sides and constructio of extra triangles, looks absurdly clumsy beside this one, which sets aside as irrelevant (as it is) the fact that the two triangles have been derived from a common source. (The proof cannot be claimed as wholly original, as it was known apparently to Pappus (A.D. 300), but it was certainly unknown to Minsky and those working with him.)

The question may now be asked whether the proof was "really" produced by the computer or by Minsky. In fact, however, there was no computer! – Minsky was trying out his programme by pencil-and-paper simulation when the simulation process led to his writing down the proof! To attribute it to Minsky is true in some obvious sense, but the allocation would be very misleading: his activities were directed at producing a process, not a proof. If the proof is to be attributed to anything it must be attributed essentially to the process, for wherever that *process* occurs, whether in a computer, or in Minsky's brain, or perhaps in Pappus's brain, that proof is capable of emerging. Thus our original question, of "man-versus-machine" type, has been found to be misdirected. The interesting question of today, valid for both man and machine, is of the type: What processes tend to generate what results?

Pattern recognition

Another branch of "higher" information processing that has been much studied recently is that of "pattern-recognition". Some of the work has been explicitly so directed: the ten digits to be read from a cheque, the 26 letters on an envelope, the ten digits when spoken into a telephone, and so on. Other work has involved it implicitly – e.g. is this position at chess suitable for an advance in the center?, will this geometrical proof be helped by drawing a circle?

The opinion seems to be emerging that *every* pattern-recognizer must ultimately be a special purpose one, designed (whether by man, machine, or natural selection) to perform a certain grouping because that grouping is useful.

With this opinion the writer agrees: there can no more be a general-purpose recognizer than there can be a general purpose map (for *all* countires!); but perhaps not all workers in the subject would agree with me. Be that as it may, pattern recognition by machine is today being used industrially in the simpler operations. (Its development to a major degree will depend on the effects of the difficulty referred to later.)

Error correction

Another technique that has grown greatly in the past decade is that of "error-correcting codes". One of Shannon's first discoveries, with his new theory of information, was that no matter how much messages might be disturbed by noise (i.e. by effects due to irrelevant and undesired causes), there always exists a code, i.e. a way of sending the messages, that shall reduce the disturbance to insignificance. The catch is that the code must be matched in its general characteristics to those of the noise, and the code may be very difficult to find. Now the brain is obviously subject, in its work at any moment, to many effects from "irrelevant and undesired causes" – think of the car driven at night in town, surrounded by flickering lights of all colors, only a few of which are traffic signals. Even in more ordinary situations, a large part of what comes to our retina is simply irrelevant to the work in progress, and must be nullified. It is likely therefore that the brain, during evolution, has developed many special methods for combating noise. Unfortunately, most of the studies of error-correcting codes have been made either for telephone/radio channels or for processes in the computer. What has been done to help understand these codes in the brain has been reviewed by Arbib (1964), but it is clear that much remains to be discovered. The psychiatrist might well find that his clinical material, viewed from this angle and suitably interpreted, gives invaluable evidence about the brain's processes.

Adaptive machines

Another aspect of "intelligence" that has quite lost its mystery is the power of the brain to change its organization. From the earliest surgical experiments (e.g. Marina, 1915), to the wearing of reversing spectacles (e.g. Taylor 1962), it has been known that the brain has a remarkable power, when faced by a new external situation (e.g. a reversal of the attachments of the ocular muscles), of itself altering its ways so that it *compensates* for the external reversal. Analysis of the theory of machines, however, (e.g. Ashby, 1940, 1947, 1952) showed that the "mystery" was due only to our thinking of machines in too simple a form. As soon as one considers the case in which the whole is formed of two sub-machines, of which one performs the obviously visible part while the other performs tasks showing only in the structure of the first machine, then one has a whole that may, if one wishes, be regarded as "a machine that changes its own organization". Today, "adaptive controls", as they are called, have a developed theory and a growing technology. Their theory is today simply a part of the modern theory of machines. It is shown most strikingly perhaps in the latest types of computer. In the early forms, the computer simply performed a computation – solving a set of equations, say, – and it was the human programmer who managed the machine's progress from problem to problem. Today, the computer that solves the equations is only a part, becoming even a minor part, of the total machinery. Behind it is another computer that acts only to manage the primary computer. The "second level" computer (the "manager") accepts a variety of problems, arranges them in order of priority, brings forward necessary sub-routines, finds suitable storage locations, tells the primary computer what to compute, and may well order it in the middle of its job to lay that job temporarily to one side so that another job of higher priority can be put through. In fact, the enormously increased power of modern computers is not so much due to faster speed in electronics as to vastly better organization of its work. (We could say, less politely, that modern machines are not so appallingly inefficient as the early machines, which would perform the computation in, perhaps, a second, and would then wait for ten minutes while the human programmer supplied the next problem.) It is not impossible (or perhaps is likely) that the human brain is characterized not only by what it can do in the immediately manifest way, but by its exceptional power of planning what it will do, at what time, and under what conditions. If so, the modern computer, that plans its computations ahead, may be developing along the same lines as the living brain. (The subject is referred to again below.)

The Theory of Machines

The achievements described above have been possible only because the past ten years have seen the emergency of a general "theory of machines". Books have been written in the past with this title, but they have referred only to the purely mechanical. The new theory of machines is based on a property that, though suspected or accepted for two centuries, is only beginning to show its power: the idea that a machine is any system such that its state at one instant determines its subsequent behavior. This property, taken for granted by Laplace, was explicitly denied by the ancients, who held that an event now might be determined by what happened many years earlier (the laying of a curse, for instance) regardless of the events in between. Two centuries of science however, have shown that, in every system adequately studied, its future behavior has been found to depend on just its present state and its present surroundings. The

consequences of this "law" (unnamed but universal) are beginning to be traced.

It is not yet easy to say of this new "theory of machines" to what degree it may be useful in psychiatry; what is sterile to one worker may be rich in possibilities to another. Here I will attempt to sketch some possibilities.

It is now known that all behaviors that are clearly and objectively describable can be produced by a machine, in the sense given (McCul loch and Pitts, 1943; von Neumann, 1951); so the question: "Can a machine do it?" is dead, for the answer is always "Yes". (I exclude here certain purely logio-mathematical complications.) Related to this result is the recent proof by Steiglitz (1965) that the analogue and digital modes of processing information are essentially isomorphic: whatever can be done by one can, perhaps clumsily, be done by the other. In some sense, therefore, any argument about whether the brain is "really" analogue or "really" digital is of minor interest, for all the higher processes of intellectual activity could be achieved by either mode. Further, all the main theorems provable in one mode must have a corresponding form true in the other. Those who are interested only in the higher processes may thus justifiably ignore the distinction: it becomes significant only when one considers the actual working details.

Some of the results to be anticipated from this theory of machines can be seen from the work of Gill (1962). Among the problems he considered were two that obviously may have application to psychiatry. The first he called the Problem of Diagnosis (he was thinking largely of the computer, but the medical parallel was obvious): given that a system, whose laws of behavior are known, is in some one of a *set* of states, devise a sequence of actions on it, and observations from it, that shall enable the observer to identify the state it is (or was) in. (The theory accepts that the making of the observations will usually change the system's state.) To prove his theorems, Gill showed that we must distinguish between the *simple* experiment, where the machine is unique and non-expendable (like a human patient), and the *multiple* (like the laboratory rat), where the system may be returned repeatedly to the same state (by just starting again with a new rat). Also to be distinguished were the *pre-set* experiment, in which the experimenter would declare beforehand all that he would do, and the *adaptive,* in which later stages of the experiment would be dependent on what had been observed in the earlier stages. He proved a number of theorems some of which showed that certain diagnosis problems were essentially unsolvable; others, essentially unsolvable by a pre-set experiment, would become solvable if the experiment were adaptive.

He also considered the Problem of Homing: to so act on the system as to bring it to a desired state (e.g. to one corresponding to "health"). Here again theorems have been proven about when a pre-set "treatment" may succeed and

when the treatment *must* be adaptive, i.e. based on information gathered during progress.

These results are, at the moment, somewhat remote from immediate application. Nevertheless, by making clear the *principles* that must guide the therapist in his interactions with his patient, they may well lay the foundations for a science of the therapy of complex systems, replacing methods based somewhat on intuition and rules of thumb. All the results are, in a sense, dominated by information theory, for they treat the situation of the compound "therapist + patient" as one system subject to basic laws of cause and effect: the patient obviously so, and the therapist also restricted in that he cannot become "knowing" except as the actions or behaviors of the patient make him so.

Consciousness

In this discussion of persons treated as machines, the reader may feel that some essential element of "consciousness" is being ignored. It is true that, in the past, the distinction between man and machine was so obvious that even the slightest resemblance was astonishing; but the point of view has changed much in the last twenty years. By "a machine" is today meant "that which behaves as a machine"; so far as man behaves like a machine, so far *is* he a machine – to other observers. If a woman dislikes her husband coming home late, but is always put into a good mood by being given flowers, then if her husband *is* late and puts her into a good mood by bringing her flowers, it is merely a verbal matter whether we say he is treating her as a machine or as a woman.

From the operational, and from the entirely objective point of view of information and computer theories, all scientific knowledge of dynamic systems is knowledge of the aspect that is machine-like. Nevertheless, the questions are still being asked: Can a machine know it is a machine? Has a machine an internal self-awareness? Can it feel pain?

These questions are of the greatest difficulty. One should notice that "consciousness" is sometimes used in a sense that is not intended here; after a motor accident, say, a victim may be "unconscious" in the sense of being simply non-reactive: pricked with a pin he makes no movement. There is no difficulty about *this* use of the word: any dynamic system may be demonstrably reactive or non-reactive. The difficulty enters when "consciousness" is used to refer to personal introspective awareness, to direct "self-knowledge".

It is sometimes held (e.g. Culbertson, 1950) that we have only to extend our scientific knowledge a little further and all will be explained. I can only say here that my opinion is quite otherwise. The work of the last twenty years seems to me only to have repeatedly emphasized the profound difference between those aspects of a system that an observer can discover from its outside, by

interacting with it (giving it stimuli and receiving stimuli in return from it) and those aspects accessible to the system about itself. The difficulty seems to be that science deals only with what is communicable (to other scientists and thus to the body of collective knowledge). A system can thus yield to science only such aspects of itself as are communicable. Some aspects, e.g. its weight, are readily communicable, but what Eddington described as "my taste of mutton" is not so: he can transmit to another only his reaction to mutton. As soon as one attempts to probe this matter thoroughly one comes, it seems to me, directly at the fact of solipism. If I have no absolute certainty whether a starfish feels pain when it is pricked, or a mimosa, or a balloon, (though all three react), I have to admit that I am exactly as devoid of certainty if what is pricked is a twin brother: of only one object in the universe have I the direct certainty. Self and non-self are, from this point of view, entirely, and not just quantitatively different. It seems to me, therefore, that the last twenty years' work in cybernetics, far from bridging the gap between knowledge of self and knowledge of other, has only strengthened our appreciation of its profundity.

In one aspect, however, the theory of machines helps to support the psychotherapist in his conviction that empathy can be useful. The *isomorphism* of systems has not yet been studied much beyond the elementary cases in engineering and physics, but if patient and therapist have a similar background of childhood and experience, the "structure" of knowledge and normal adaptation in the therapist is clearly available as reference for the differing structure in the patient. The subject is too large to develop here; but it may well provide the possibility of a fully scientific basis for the very high-level interactions between patient and psychotherapist.

Information theory of many variables

This digression to the subjective may give clarity to a brief discussion of whether the methods of Shannon are adequate to represent the many ways in which "information in general" enters into such subjects as psychology, psychiatry, sociology, and everyday life. Here there is space only for me to record my opinion that it *is* sufficient, and that most of the dissatisfaction with it comes either from the wish to introduce introspectional aspects (consistently excluded from scientific work) or from a failure to appreciate the great range of ideas and methods that Shannon's basic work has opened up. Here it must be admitted at once that we in the biological sciences have been little helped of recent years by the mathematicians and engineers, most of whose developments of the theory have been directed at the telephone and the computer. The developments in directions meaningful in the biological sciences have largely yet to be made. As example, take the fact that most developments are from the

case of sender-receiver: two variables. Now two variables is an absurdly small number for most biological systems. McGill (1954) showed that the extension of information theory to any number of variables is straightforward, and his methods have proved capable of further extension to matters of real interest to the biologist (e.g. Garner, 1962; Ashby, 1965). It is, for instance, now possible to measure informational transmission not merely in-and-out, as a passive telephone wire or an optic nerve treats it, but as the amount that is processed *internally,* as a computer works or a man thinks. (It should be remembered here that "internally" may be interpreted not only as "internal to the organism but also, if one wishes, as "internal to the system of organism-and-environment", the interaction between the two being the real focus of interest.)

One consequence of this development is that it provides a direct and objective measurement for the amount of co-ordination or integration in a system (Ashby, 1968a). To make the idea clear, let us consider the pianist who has the skill to play scales in any key. "Co-ordination", muscular and nervous, is clearly involved, and is objectively demonstrable, for any disturbance by drugs or disease would show objectively as a failure to keep to the appropriate eight notes of the possible twelve. Now while he is playing (correctly) in, say, G major, the twelve notes are obviously not being produced at random, i.e. with statistical independence. (An example in detail is given below.) This lack of independence corresponds to a calculable quantity of "transmission" that *must* exist in some form between the various finger movements if the coordination is to be achieved. The transmission may be effected by a great variety of possible mechanisms (and one must not jump to the assumption that it must all be mediated by nerve fibres, for, e.g., mechanical forces may in fact be used) but the total quantity of transmission *must* be demonstrable if the process is not to be achieved by non-material magic. Thus, *every well-defined set of actions showing co-ordination specifies a definite quantity of internal transmission* that must be performed if the co-ordination is to be successful. The quantity is as basic as, say, the quantity of work that a man weighing 150 pounds must do if he is to climb a 20-foot ladder. Because of its fundamental nature, this quantity of information associated with co-ordination and integration may well prove a useful index when co-ordination and integration fail.

Dynamic nets

Computer science, too, is today intensively but narrowly specialized, since it is still largely concerned only with prodigiously long chains of simple addings and multiplyings. In this particular form of "complex cause-and-effect relations" it tends to be of small direct interest to the psychiatrist. Designers, however, are aware that more advanced forms of information processing will require something

more complex than simple chaining: many more parts must be active simultaneously. *Illiac IV* (at the University of Illinois) is now being designed so as to be able to carry on 256 operations simultaneously. A very different style of programming will have to be developed, and the programmers themselves will have to think along somewhat new lines, but the extension will doubtless continue. Little, however, is being done in the direction of exploring the "computer" that is brainlike in the sense of using nearly all its parts nearly all the time. To understand such a system, we need to know much more about what might be called "generalized dynamics" – the dynamics of systems that are supplied freely with energy (and so are not restricted by its conservation) but which have laws to rule them because they are state-determined and either isolated or subject to a determinate input. It has been proved (Ashby, 1959) that ***habituation*** will tend to be shown by a very wide variety of such systems, and some further properties have been identified (Ashby, 1960), but progress is slow.

The study of such systems by the methods of classical mathematics leads quickly to quite unmanageable complexities. Modern studies are turning increasingly to the method of modelling such processes on a computer and simply seeing what happens. The behaviors of nets of randomly connected units were studied in this way by Walker and Ashby (1966). Certain general trends were found, useful perhaps for further studies in the same direction.

Studies of such dynamic systems have repeatedly encountered a phenomenon that may well be of psychiatric interest. It was encountered by Friedberg (1958), and by the writer at about the same time, and was called the "mesa" phenomenon by Minsky and Selfridge (1961). It is apt to occur whenever some change of conditions acts on a large and complex dynamic system: as the system is made larger and larger, so do the consequences of a change in conditions then to pass from the more or less smooth to having either no effect at all or to having a sudden and large consequence at one critical value. In other words, the response curve changes, as the system is made larger, from a steady slope to a step-function.

This tendency seems to be inherent in a very wide class of systems; and although each particular system has its own particular physical mechanism at work, yet the tendency is so wide-spread that it may be a general system property. Gardner (1968), for instance, has found it to occur in linear dynamic systems as the richness of internal connection is increased. He asked: what is the probability that the system will be stable?, and then found how this probability changes with increasing richness of internal connection. He found that when the system was small (five or fewer components), increasing connections caused a ***steady*** fall (in the probability of stability). As the system was made larger (to ten variables or more) the probability tended to stay high as the connections were increased until suddenly it fell to almost zero. What this means is that the large system's behavior will depend critically on the richness of its internal connections, with the dependency very sensitive near the critical value. Thus, in his examples with ten elements, a change of 2 per cent in the richness of connections caused a change of nearly 100 percent in the probability of stability.

It is obvious that a system as complex and dynamic as the brain may provide many aspects at which this "mesa" phenomenon may appear, both in aetiology and in therapy. There is scope for further investigation into this matter, both in its theory and its applications.

The nature of memory

Such studies in the theory of machines have forced into prominence, and have helped to clarify, what is meant by "memory". As was said earlier, this word must be interpreted, under its operational and objective aspects, as equivalent to "transmission between variables significantly separated in time". If the separation is small one may prefer to call it "delay"; if long, "recording". Computer science today ranges over the gamut, lumping them all as "memory", (though of course, it uses the different types of memory with discrimination).

Saying that memory is a form of, or is homologous with, transmission is more than just using a phrase. It implies that all the discipline of information theory, and all its theorems, are applicable. Thus the theorems of error-correcting transmission become theorems about how to store with immunity to specified types of disturbance; the theorems of channel-capacity now hold over storage-capacity, and theorems of transmission by code become applicable to methods of storage in code.

This new point of view is likely to open up entirely new approaches to the old problems of memory. Von Foerster, for instance, has given a most suggestive illustration in his paper "Memory without Record" (1965). The title might suggest that purely mental memory of the Middle Ages, but such is not his intent. He points out that when we ask a child: what is 3 x 7?, and the child answers 21, nothing is more obvious than that the child somewhere has a "record" (? engram) of the fact that 3 x 7 = 21. Suppose now that we want to be able to obtain on demand the product of all 10-digit numbers by all 10-digit numbers. If we assume that the record is to be in the form of an ordinary book, a little figuring soon shows that it will have to be about a billion miles thick! Yet in fact all such products are obtainable on demand from an object about a foot across – called a desk-computer! The point is, of course, that products can be generated actively: passive storage is not the only way of keeping them.

At the time of writing, the topic of "memory and its storage" is attracting

many workers, and research in the future will undoubtedly explore the subject extensively. Yet almost all the work at the moment is envisaging an essentially static trace rather than an active regeneration. Yet everyone knows that few systems have quite such rich facilities for active generation as the brain. No neurophysiologist can say that the suggestion of an *active* process is absurd.

The new possibility has major consequences. Suppose, for instance, a child of six or so sees a chess-board, and then shows later that he remembers (can produce on demand) its characteristic pattern. As 64 squares that may be either black or white, the chess-board has, of course, a very high redundancy, in the sense that after we have seen a few, 10 say, we can predict the colors of the remaining 54. Thus the actual information to be used by the boy is not the full 64 bits but 10 or less, and the memory only *need* use the 10 or less. Here is an obvious occasion for the memory to be regenerated: a few bits' storage can hold the initial conditions, and then the remainder can be regenerated (by the rule that the change to each next square calls for a reversal of the color). Thus the answer to: How does this child store the pattern of the chess-board? would be: He doesn't – he regenerates it. And, of course, an experimenter who looks for anything static that resembles a chess-board would find nothing. Only when the brain acts will the *process* develop the pattern.

Another example can be given showing how this approach to memory, as transmission, can be illuminating (Ashby, 1968b). As was said above, every act of co-ordination implies a certain quantity of transmission. If the co-ordiation extends over times as well as space (later events correlated with earlier) then a calculable quantity of memory is required. This minimal quantity is demanded by the co-ordination as such, and is quite independent of whatever mechanisms may be used to achieve the co-ordination. In the article mentioned (1968b), an example is given of co-ordination in piano-playing, in which is selected the specially simple case in which the player must play some two of three notes A, B, C, and must play on one of two beats and be silent on the other. In this selected example it is easy to show that the total achievement demands a total transmission of 2·92 bits (per bar), of which some must be transmission between the fingers and the remainder between the times. Two different mechanisms are considered in the article and it is shown that they partition this total of 2·92 bits in different ways. One obvious method is to use a "memory store" to record whether the chord was or was not played at the first beat, requiring 1·00 bit, and to provide the other 1·92 bits as transmission between fingers. Another, less obvious, method is to give each finger (or some corresponding nervous center) a memory store of its own, and then co-ordinate the fingers' trajectories. This last method demands 0·25 bit for memory at each finger, and 2·17 bits between trajectories. What is noteworthy is that the second method (or form of mechanism), though it demands three stores, is less demanding on storage (0·75 bit) than the first method (1·00 bit) which uses only one store. If, therefore, storage were very expensive (in some sense), the method with three stores should be chosen, not that with one. The example shows clearly how the ideas of information theory, appropriately used, can give an unusual insight into the fundamentals of "memory theory".

Bremermann's Limit

It remains for me to mention one other fact, a cloud on the horizon no larger than a man's hand, that I think will in time become of dominating importance in the science of information and computers.

Computers today have various limitations. Some are easily removable, with a little more time and money; others are much more profound. Perhaps the most fundamental is that identified by Bremermann (1965). He showed that two of the most basic relations in physics – the mass-energy relation and Heisenbergian uncertainty – together put an absolute limit to the quantity of information that can be transmitted by matter. It certainly covers the industrial computer and if the scientist's thinking is carried on by some material process in his brain (and no physiologist doubts this), his thinking is also absolutely so bounded.

The actual value of the limit is 10^{47} bits per gramme per second. This quantity may seem too large to be of any importance, but in fact examples are easily given (e.g. Ashby, 1963, 1964, 1966a, 1966b) showing how readily quantities exceeding this limit may be demanded. They tend to occur when the information comes from systems having actions in combination (such as were listed earlier). Thus, as soon as one tries to devise processes that will carry out actions of some real complexity – playing a game of chess, driving a car from London to York, writing an article of 10,000 words – one is apt to find, not merely that the demand goes far beyond the limit but that the demand goes beyond it by vast orders of magnitude. The limit in fact is found to be extremely restrictive, so grossly restrictive as to make clear that either our brains are not using ordinary matter (hardly a serious suggestion today) or that they are using methods that are of far higher efficiency than those used in today's computers.

Lest the difficulty be left looking like a hopeless paradox, a few words may be useful to indicate a possible solution. Many processes are known today that compute answers at first sight far beyond even the biggest machine's capacity: they all work by breaking the whole process down into sub-processes (and they can be used only in such problems as *can* be analyzed into sub-processes). Thus, while the game of chess in the strict sense, i.e. with faultless play towards mate, demands quantities of information-processing far beyond

Bremermann's limit, yet quite a good game can be played by machine (and man) as a sequence of sub-processes: 1, mobilize your pieces; 2, control the center; 3, get the rooks working; and so on. Such a division into sub-processes, each of which can be carried out without reference to the other sub-processes, has the effect, when it can be done, of lowering by great orders of magnitude the demands for information-processing.

It is not unlikely that the human brain uses this method extensively, of carrying out the total process *as a sequence of sub-processes.* But here the worker who is devoted to the idea that the brain acts as a whole (no one more so than the writer!) must beware of going too far. The Gestalt psychologists who insisted that the brain acts as a whole were perfectly right to oppose the previous generation's attempt to see the brain as a bundle of independent atomistic reflexes. But the introduction of wholeness can go too far. All the studies of the last twenty years, including those studies of nets mentioned above, show that systems should be only moderately connected internally, for in all cases too rich internal connection leads to excessive complexity and instability. The psychiatrist knows well enough that no one can produce associations so quickly or so wide-ranging as the acute maniac; yet his behavior is inferior, for knowing what associations to avoid, how to stick to the point, is an essential feature for effective behavior.

Some evidence in this direction has come to light as the result of an investigation of the amount of information that is processed by an average person in the course of his everyday activities (in contrast to his maximal capacity when stressed on a special task) (Ashby *et al.,* 1968a). This investigation attempted to assess the quantity of information processed during the following action (with emphasis on the information required for the co-ordination and integration involved): (The human subject is given as being engaged in reading when he encounters an unfamiliar French word.)

> *ACTION: He walks across the room to his book shelf (avoiding a chair that is in his path), finds his French dictionary (among 100 other books), finds the word, reads the English translation, and writes down the corresponding English word.*

Details are given in the paper. What is of interest here is that the final estimate for the rate came out at about 3 bits per second (not likely to be in error by more than a factor of 2). Now this quantity seems at first to be astonishingly low. Each optic nerve alone, with half a million fibres, can transmit at least 500,000 bits per second – where is all this information going to, or why is it collected?

The interpretation of these facts is not yet certain, but there is one interpretation that may be related to the limit (Bremermann's) just mentioned. The estimate of 3 bits per second refers to what is necessary for the defined action. Were a robot made to carry out just this action and nothing more, then fully efficient design should not demand more than the 3 bits per second. But a human being, of course, while carrying out this action, is treating this action as only one of a great number of other possible actions; so there must also be information processing to decide the answer to: should *this* action continue? Thus, should the telephone ring during the course of the action, the normal person will no longer elect to persist in this action but will switch to another. And, what information theory has made abundantly clear, the refraining from going to the telephone when it is *not* ringing demands information-processing capacity just as absolutely as the going when it is ringing.

This difference, between the millions of bits entering by all the sense and the 3 bits per second used directly in the action, suggests that what goes on in the brain may be responsible for the obviously visible ("tactical") actions to a minor degree, and to a far greater degree may be concerned with the less visible ("strategic") question of the choices between the various possible actions. Such a method, separating the details of the tactical action from the processes controlling the strategic organization of many actions would achieve just the reduction of combinations that would be appropriate to Bremermann's limit.

There is a striking parallel here with the developments in computers since they were introduced. As was said above, today's machines are enormously more effective than the earlier, though their operations of computation are little altered. Their vast superiority is due to the fact that they organize their work so much better. The triumph of the last decade is not a faster addition but the development of (say) time-sharing. Today one computer can keep a dozen departments happy, dovetailing all their demands together with a skill like that of a juggler who keeps a dozen balls in the air. The further theory of "higher information-processing" will, I suspect, be at this organizational level rather than at the unit-operational. Psychiatry may perhaps be able to learn something from the computer-scientists; but it is just as likely that the computer-scientists will be able to learn from the psychiatrists. What is chiefly necessary at the present time is that they should learn to speak something of each other's language.

Looking back over the last twenty years one is tempted to think that information theory promised too much, and failed to deliver the goods. Yet the fact remains that information theory is essentially the science of complex dynamic systems, with complex weavings of causes and effects in great numbers. Those who would study such systems need information theory (in some form) just as surveyors need some form of geometry. What has happened, I think, is that so much effort has gone into its development for the telephone and computer that it has developed along lines little suited to the real needs of workers in the

biological sciences. Take for instance the fact that almost all information theory developed so far deals with the very tidy case in which the system is going to use an exactly defined set of symbols – the 26 letters of the alphabet, the 10 digits, the distinct voltages between -3 and +3, for instance – a very useful case in much engineering. In the biological cases, however, the "alphabet" is not sharply limited, but tails off almost indefinitely. Again, Shannon's basic method is to consider one set (e.g. all possible ten-word phrases) with the messages as a population to be sampled. But in "content analysis" one has the essentially opposite situation: a unique message has been received and one wants to discuss the various sets from which it might have come. Thus a patient might utter just "Doctor, I hate you", leaving the real question whether this message is from the set of those expressing opposition, or whether it is from the various ways of saying "At last I can be frank and reveal what is troubling me." Content analysis thus provides a direction which the basic ideas of information theory can be developed in a way of real interest to those in the biological sciences. A start has been made (e.g. Krippendorff, 1967) but the field is almost entirely unexplored.

His work, and the other advances described above, suggest that information theory is at last beginning to be forced in the directions appropriate to the needs of the biological sciences. Twenty years' experience has helped to make the topic more realistic. The time is now ready for the researcher who can appreciate sympathetically the work done by the engineers, and who then, with the needs of the biological worker firmly in mind, can force its development in the directions appropriate to psychiatry. Perhaps this article may help to suggest what these new directions may be.

Acknowledgement. The work on which this article is based was jointly supported by the U.S. Air Force Office of Scientific Research under Grant 7-67, the U.S. Air Force Systems Engineering Group under contract 33(615)-3890, and the National Aeronautics and Space Administration.

REFERENCES

ARBIB, M.A. (1964). ***Brains, Machines and Mathematics.*** New York.

ASHBY, W. ROSS (1940). "Adaptiveness and equilibrium." ***J. ment. Sci.,*** 86, 478-483.

——— (1947). "Principles of the self-organizing dynamic system." ***J. gen. Psychol.,*** 37, 125-128.

——— (1952). ***Design for a Brain.*** London.

——— (1958). "Cybernetics." In: ***Recent Progress in Psychiatry*** (ed. Fleming). 3, 94-117. London.

——— (1959). "The mechanism of habituation." In: ***N.P.L. Symposium on the Mechanization of Thought Processes*** (ed. Cherry). London, 4-4, 1-21.

——— (1960). 2nd edition of 1952.

——— (1963). "Systems and information." ***Trans. IEEE,*** MIL-7, 94-97.

——— (1964). "Modelling the brain." In: ***IBM Symposium on Simulation Models.*** pp. 195-208, New York.

——— (1965). "Measuring the internal informational exchange in a system" ***Cybernetica,*** 8, 5-22.

——— (166a). "Mathematical models and computer analysis of the function of the central nervous system." ***Ann. Rev. Physiol.,*** 28, 89-106.

——— (1966b). "Some consequences of Bremermann's limit for information-processing systems." ***Bionics Symposium,*** Dayton, May 3-5. (In the press.)

——— (1968a). "Information-processing in everyday human activity." ***BioScience.*** (In the press.)

——— (1968b). "Measuring memory." In: ***Festschrift for Prof. P.K. Anokhin.*** (In the press.)

BREMERMANN, H.J. (1965). "Quantal noise and information." In: ***5th Berkeley Symposium on Mathematical Statistics and Probability*** (ed. Neyman), vol. 4.

CULBERTSON, J.T. (1950). ***Consciousness and Behavior.*** Dubuque.

FEIGENBAUM, E.A., and FELDMAN, J. (eds) (1963). ***Computers and Thought.*** New York.

FRIEDBERG, R.M. (1958). "A learning machine." ***IBM J. Res. & Devel.,*** 2, 2-13, and 3, 282-287.

GARDNER, M.R. (1968). Thesis for M.S. University of Illinois.

GARNER, W.R. (1962). ***Uncertainty and Structures as Psychological Concepts.*** New York.

GILL, A. (1962). ***Introduction to the Theory of Finite-state Machines.*** New York.

KRIPPENDORFF, K. (1967). Thesis for Ph.D., University of Illinois.

McCULLOCH, W.S., and PITTS, W. (1943). "A logical calculus of the ideas immanent in nervous activity." ***Bull. Math. Biophys.,*** 5, 115-133.

McGILL, W.F. (1954). "Multivariate information transmission." ***Psychomet.,*** 19, 97-116.

MARINA, A. (1915). "Die Relationen des Palaeoencephalons sind nicht fix." ***Neurol. Centralbl.,*** 34, 338.

MINSKY, M. (1963). "Steps toward artificial intelligence." Reprinted in Feigenbaum and Feldman (q.v.), 406-450.

——— (1967). (Personal communication.)

——— and SELFRIDGE, O.G. (1961). "Learning in random nets." In: ***Proc. 4th London Symp. on Inf. Theory*** (ed. Cherry). London.

NEWELL, A., SHAW, J.D., and SIMON, H.A. (1957). "Empirical explorations with the logic theory machine." Reprinted in Feigenbaum and Feldman (q.v.), 109-133.

SAMUEL, A.L. (1963). "Some studies in machine learning using the game of checkers." In: Feigenbaum and Feldman (q.v.), 71-105.

SHANNON, C.E. and WEAVER, W. (1949). ***The Mathematical Theory of Communication.*** Urbana.

SHEPHERDSON, J.C. (1967). "Algorithms, Truing machines and finite automata." In: ***Automaton Theory and Learning Systems*** (ed. Stewart), 1-22. London.

SIMON, H.A., and SIMON, P.A. (1962). "Trial and error search in solving difficult problems." ***Behavioral Sci.,*** 7, 425-429.

STEIGLITZ, K. (1965). "The equivalence of digital and analog signal processing." ***Inf. & Control,*** 8, 455-467.

TAYLOR, J.G. (1962). ***The Behavioral Basis of Perception.*** New Haven.

VON FOERSTER, H. (1965). "Memory without record." In: ***The Anatomy of Memory.*** (ed. Kimble), 388-440. Palo Alto.

VON NEUMANN, J. (1951). "The general and logical theory of automata." In: ***Cerebral Mechanisms of Behavior*** (ed. Jeffres), 1-32. New York.

WALKER, C.C. and ASHBY, W. Ross (1966). "On temporal characteristics of behavior in certain complex systems." ***Kybernetik,*** 3, 100-108.

WANG, H. (1960). "Toward mechanical mathematics". ***IBM J. Res. & Devel.,*** 4, 2-22.

THE BRAIN OF YESTERDAY AND TODAY

Summary

Every child naturally becomes a Flat-Earther and has to un-learn what at first seemed obvious. We also pick up many "obvious" facts of psychology and may have to unlearn them later. The article presents some of the "facts" that must be unlarned if we are to advance realistically to artificial intelligence. Among those discussed are:

(1) That a regulator *must* have feedback from the error.

(2) That a self-repairing machine, when a fault occurs, must first localize it.

(3) That the first thing to do in pattern-recognition is to store the pattern.

(4) That the brain of the free-living organism can be regarded adequately as "responding to a stimulus."

(5) That if a memory can be produced later, it must have been stored.

(6) That an organism's stored memories must be stored in the brain.

We all want to be up-to-date in our thinking, but being whole-heartedly up-to-date is not easy. Civilization carries with it a great store of old ideas and ways of thinking that are often useful, sometimes horribly obstructive. Worst of all are those ways of thinking buried in what we take for granted: then we go wrong without even the option of a choice.

Most of us, for instance, would repudiate at once the suggestion that we are Flat-Earthers. "Copernicus settled that over 400 years ago," we would retort. Yet how many of us, when asked, "Where are the stars?" point anywhere but upwards? Unfortunately, of course, every child starts by being a Flat-Earther, for he can actually *see* that it is so; later, he not only has to learn, he has also to unlearn.

Unfortunately, our subject of "brains" requires much to be unlearned. Man has always taken for granted that he knows how he thinks, how he sees, how he remembers, just as he knows how he moves his arm, or how he sneezes. In fact, psychological and behavioral studies of the last half-century have shown abundantly how extremely unreliable are the basic ideas we pick up by introspection and casual observation. The psychologist today is as suspicious of introspection as the organic chemist is of his sense of smell: he uses it occasionally on the chance that he may obtain a useful suggestion, but under no circumstances does he *trust* such evidence. Yet even today, for lack of contact with *real* psychology,

---

The work on which this article is based was supported by AF-AFOSR 7-67 and AF 33(615)-3890.

the background thinking on artificial intelligence is largely pre-Darwinian and entirely pre-Freudian. The operations of Boolean algebra tend to be treated as *the* laws of thought – as George Boole considered it to be in 1854, when "psychology" meant "how a man ought to think," not "how he does in fact think".

For these reasons, many of today's comparisons between brain and computer are invalid, simply because they are unbalanced and ill-founded. The article by Hubert Dreyfus[1], for instance, is in my opinion largely invalidated, in spite of much skill in the presentation, because of its failure to get man into his proper place for such a comparison. Whatever else man may be, to the Darwinian he is a mechanism shaped in every possible way to match the requirements of his particular planetary environment. He has been shaped to fit not merely its gravitational, optical, chemical and similar properties but also to its less obvious characteristics: that his space is a three-dimensional and Euclidean one, rather than one in which the moves are all like the knight's in chess; that continuity is extremely common; that "change", of almost any type, is likely to be significant; that union makes for strength; and so on through the tendencies recognized in proverbs. To compare man and machine fairly one must start by appreciating that on the human side we have (if our example of a typical man is, say, a graduate aged 25), what is probably the most elaborately pre-programmed thinking device in the world, with many millions of years of pre-programming built into it, while on the digital computer's side we have a device that has been deliberately stripped down to have absolutely no behavioral structure, so that it is an entirely clean sheet on which the new instructions may be written without interference.

Dreyfus points out that these two – the young man with a million plus 25 years of training behind him, and the computer with a stack of cards as its totality of knowledge – have widely different properties. No one, I think, would disagree with him. But the suggestion that there is some essential difference between them can come only from someone who has forgotten his own personal apprenticeship, and who has overlooked, or never even discovered, the long apprenticeship of his species.

Let us make the comparison fair; take the highly programmed human being and diminish his program with disease or senility; take the machine and give it something like the vast input that has come, through evolution and personal learning, to the individual; then *fairly* judge the relative performances. Until this is done, at any rate let us remember that we are in the days after Darwin and Freud, not living among the superstitions of Plato and Pascal.

Perhaps I can reinforce this suggestion (that we may have to un-learn some ideas) by pointing to another extremely common and pervasive mistake: the idea that a self-correcting regulator *must* have feedback from the error. Nothing

is easier than to open the subject of error-correcting regulators by drawing such a diagram as Figure 1.

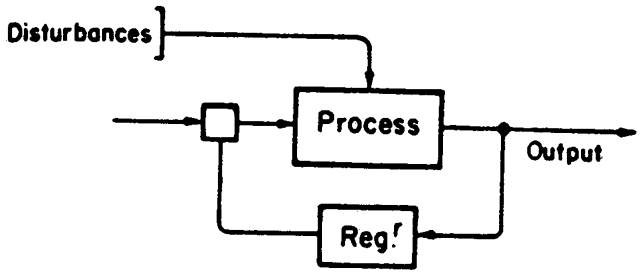


Figure 1.

suggesting that the error-correction *must* be done by a path from the error back to the input. To think of the brain this way is to go wrong at the very first step. The fact is, of course, that most of the biological advance from lower to higher forms of behavior has consisted precisely in changing from this method, which can correct an error only *after* it has occurred, to all those methods of using distance receptors by which the regulatory action is prepared in advance of the actual error, so that the error never occurs. The boxer does not start his evasive action when his opponent's fist begins to exert a rapidly increasing pressure on his jaw! Eyes are there precisely to gain earlier information. Evolutionary advance has consisted largely in getting away from the primitive form of Figure 1 to the form of Figure 2.

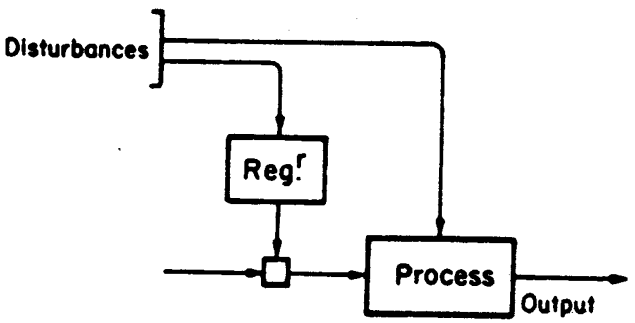


Figure 2.

To forget this fact may be to remain at an unnecessarily inefficient mode of regulation. Suppose, for instance, that the regulating machinery of an oil refinery has to deal with the troubles caused by the crude's content of sulfur. A feedback method might find the problem most difficult: then is the time to remember that there is an alternative – that a cable sent ahead of each tanker, telling the sulfur-content of the arriving oil may make the solution much easier.

This example may justify my belief that at the moment our chief mistakes are occurring right at the beginning. A problem comes forward and we say "Obviously, we must . . .," and find later that our very first step was decisively wrong. Let me give some further examples.

In the theory of self-repairing machines, nothing is easier than to say that the stages are obviously:

(1) localize the fault,

(2) correct it.

Then the worker starts on the difficult theory of how to localize a fault, overlooking the fact that many faults may be cured without localizing them. They are obvious enough as soon as one starts to think that way. An abcess, somewhere in the internal organs, may properly be treated by an antibiotic and may well be cured without the question "where is it?" ever being answered. A bad contact on a chassisful of electrical equipment is frequently cured by a bang on the chassis: the fault is cured though nobody knew where it was. Since there are known to exist methods that will cure without the trouble of first localizing, we first should study this type of process, reserving our more laborious studies for when they are unavoidable.

Another example where the first step may be plausible yet disastrously wrong occurs in pattern-recognition by computer. 'It is obvious that" the first thing to do is to take the picture (say) and get it into the store of the computer. Yet the computer store may offer far more variety, information, entropy than is necessary; so the next stage, that works on what the store offers it, will be faced, not only with the discrimination originally needed, but also the task of eliminating all this superfluous information, or variety, which now has to be treated as noise. As a very simple example, merely to make the point clear, suppose the machine is given the extremely simple task of distinguishing the left "picture" from the right in

1 0	0 1
0 1	1 0

Obviously, this distinction requires merely 1 bit, and any method that uses more is demonstrably inefficient. But suppose we follow the obvious method and first store the picture; "surely this can do no harm," we might say. A plausible way is to number the places as

1 2
3 4

and then to list the numbers of the cells that contain a 1. The left picture has them in cells 1 and 4, so it is recorded as 001100, and the other is recorded as 010011. *Now* the recognizer has to work with functions of six variables, so that the possibilities that the recognizing-process must handle have jumped from

2 to 64. (One also notices that the simple and obvious complementarity has been lost: another sign of how "mere" storage can add complexity.)

The living brain already knows this, as the researches of the last few years have shown. The eye extracts information about such useful properties as edges and local movement in the retina itself, *before* more complex transforations have changed these properties to forms that would require vastly more work later to re-extract them. Codings performed at the right time may be advantageous, but a coding performed at the wrong time may be disastrous.

There is yet another way of being misled by what we have learned. Fifty years ago Sherrington and Pavlov made great advances by artifically restricting their view of the nervous system to the behavioral unit of a stimulus, followed by a response: stop. For the next 30 years the psychologists followed suit. As a result there is today an ever-present danger that we continue in this restricted view, failing to deal with the reality of cerebral life, which is that it is one darn thing after another. In real life, each response is the starting point of the consequent reaction; and real life, of any significant degree of complexity, consists of such responses in sequences; and the properties of the sequences quite transcend those of the mere unit. In other words, the brain is dynamic: every effect is irresistibly the cause for the next step.

So stated, the proposition seems obvious; everyone knows it. But does everyone use it? I would like here to quote von Foerster's[2] recent observation that not merely individuals but almost all neurophysiologists have failed to *use* this fact. It is in connection with the problems of memory and its storage. When we are asked what is 6 times 8, we produce "48"; the phenomenon represents "memory", and today's researches are directing much attention to how it is stored. Von Foerster's paper, however, was entitled "Memory without Storage," and at first sight one wondered whether wholly immeterial agencies, of the type now rejected as "mystical", were to be invoked. His argument, however, is entirely in accord with the modern view. He points out that if one wants to obtain such answers as "48", and wants to range the question over the products of all ten-digit numbers, there is first the obvious method of storing the products. If the set of products were printed into a book, the book would be about 10^{10} miles thick. But it need not be stored. The same output, of all products on demand, can be obtained from a little thing only a few inches thick called a "desk-calculator".

With this example Von Foerster, having provided that memories can be regenerated (as well as stored), forces us to face up to the fact that the brain is dynamic, and that it may well achieve its results by active, as well as by merely passive, units.

Here I can do little more than to make sure that this fact receives the

attention it deserves. Its advent is too recent for us to be able to say that we have really grasped its full implications. But there can be no doubt that practically everything we have thought about "memory" in the past will have to be rescrutinized. This is beyond question a matter on which each worker must be clear whether he is thinking of the brain of yesterday, or the brain of today.

The impact of this new idea is likely to be revolutionary, in the literal sense of "turning to the opposite." Consider for instance, the elementary fact that a boy, having seen a checkers board for the first time, can remember it. Here is "pattern" exemplified most clearly; and we may well ask the obvious question: How does he store it? But let's think again: pattern means redundancy; the more marked the pattern, the more extreme the redundancy. Now the primate brain is hardly likely to achieve pre-eminence by specializing in the storage of information redundantly. In the storage of patterns, then, is an obvious opportunity to achieve the same performance while using the more economical method of regeneration: the natural way to store redundancy is to use one small *dynamic* system, whose output will regenerate the redundant form. (The events are essentially similar to those in the oscillator that puts out a repeated waveform: the redundancy in the repetitions corresponds to the physical persistance of the one material system.) It seems, then, that the proper answer to the question, How does this boy store the memory of the checkers board? is: He doesn't. What he probably does is to form such a dynamic system, from the tangle of neurons available, as will emit, in its output, whatever train is later coded into the checker board. From this point of view, to look into the brain for anything even remotely like the checker board is to look for what is probably not there: to look for what the brain, in evolution, has learned to avoid.

If we really believe that the brain is dynamic, and pay this belief more than mere lip-service, we might as well go on to its logical completion, and take the ultra-modern standpoint of seeing the brain, in its environment, as a steady state.

The proposition can, of course, be defended by weighty arguments of formally scientific type. I propose, instead, to support it by just one trifling anecdote, which I hope you will find not wholly uninteresting. It occurred last summer when I attended a seven-man conference on the shores of Lake Traun in Austria, with the Wenner-Gren Foundation. Each morning I walked to the conference room along the path that ran through the park. One morning I saw, far ahead of me, the lone figure of Carl Helm, one of the participants. I recognized him at once even though the very bluish light made his hair look darker than its usual blond, and the jacket he was wearing was not his usual one. I was then puzzled to know how I could possibly recognize him so confidently when, at that distance, his image on my retina was reduced to little more than a few points. While wondering how the brain performs such miracles of pattern recognition, I gradually overtook him, and then found that it was not Carl anyway This is the reality of the human brain at work, so let us look at the events more closely.

We can now see what had happened. The facts that it was by the lake, on that lonely path, at ten minutes to nine a.m., were already playing an essential part, unconsciously establishing that any person I saw going to that house was one of the other six participants. The sole decision remaining was: which of the six? The answer was then obvious: the man on the path had the conspicuous flapping white collar of an open shirt. Carl was the only participant who had worn such a collar and he had worn it most of the time. The point I want to make here is that while the sensory input from the environment acted in the elementary way to guide me along the path and to keep me from colliding with trees, it was also permeating and playing a part continuously in all sorts of activities, such as recognizing people, in which it might, at first sight, seem irrelevant.

But if the environment is playing a part, and by *its* constancy providing some of the carry-over of information from minute to minute and day to day, then we shall find that some of the human subject's coherence of behavior is due, not to anything carried by him but to the information carried from minute to minute and from day to day by the environment. Phenomenologically this "memory", shown in my behavior, is not localized in my brain at all: it is carried in the environment, at no cost to the human subject and with an increase in his efficiency.

To regard the brain, in its environment, as a steady state might have been regarded, until recently, as impossibly complicated, but my belief is that today our techniques are within reach of this fully realistic view. We have the theory of finite state machines of any richness of internal feedback; we have the theory of dynamic nets, well started by Walker's[3] work; we have communication theory generalized to the flows between any number of variables; and we have the theory of stability developed and completely generalized in set theory. Here it seems to me, is the brain-theory of today and tomorrow.

REFERENCES

1 DREYFUS, HUBERT L., "Alchemy and Artificial Intelligence," Seminar, printed by RAND Corporation, December 1965.

2 VON FOERESTER, H., "Memory without Record," in ***The Anatomy of Memory,*** (ed. D.P. Kimble), Science and Behavior Books, Palo Alto, California, 1954, pp. 388-440.

3 WALKER, C.C. and ASHBY, W.R., "On Temporal Characteristics of Behavior in Certain Complex Systems," ***Kybernetik,*** 3, 1966, pp. 100-108.

VIII.

UNPUBLISHED

CLASSROOM HANDOUTS

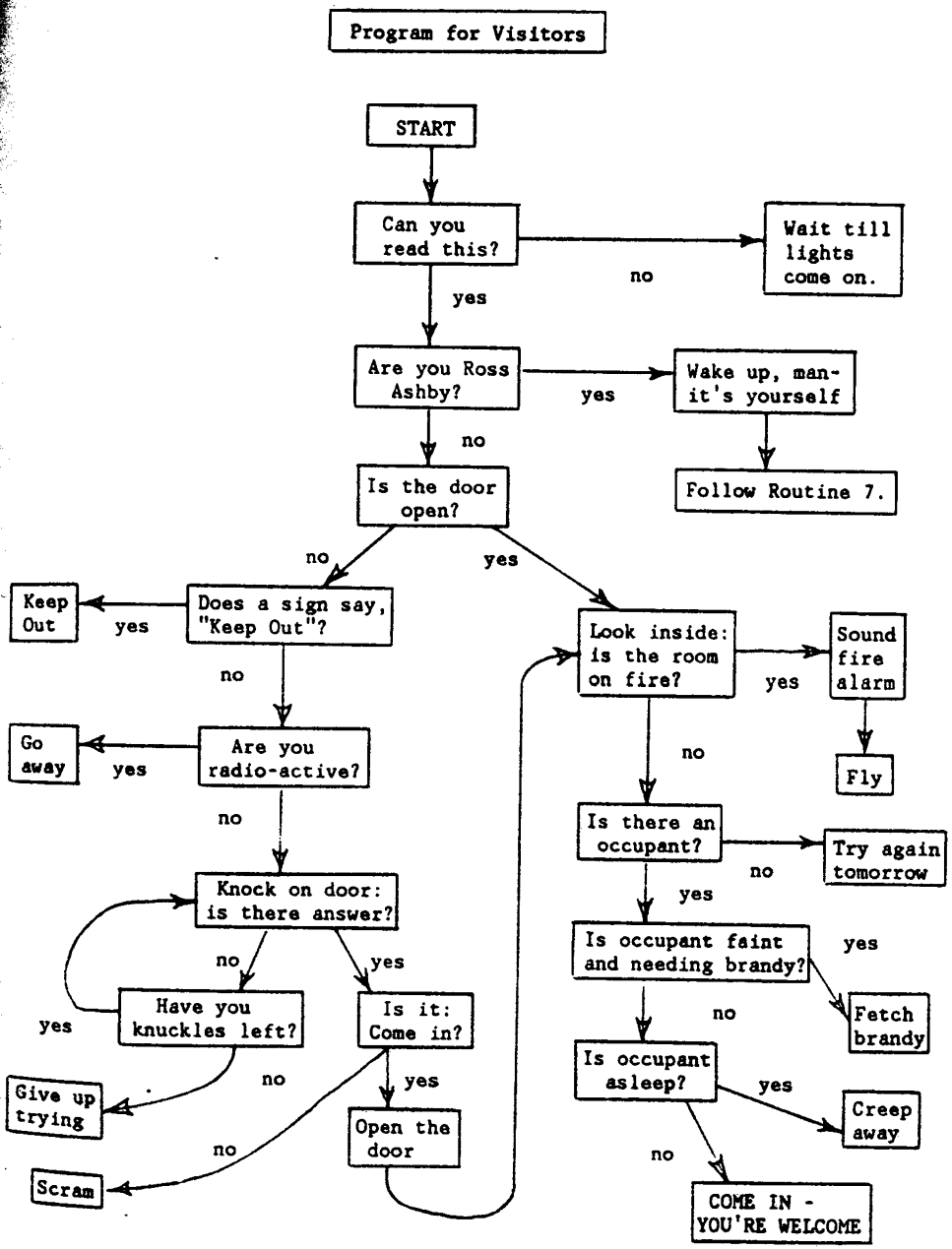
UNPUBLISHED CLASSROOM HANDOUTS

INTRODUCTION

Those of us who were fortunate in having Ross Ashby as a teacher or advisor during his stay at the University of Illinois cannot forget his animated classes, his enthusiastic delivery, some of his "props" such as the little woodpecker-on-a-spring who pecked his way down a pole (illustrating a system going to equilibrium), his famous "black box" which could be wired up to exemplify mappings and properties of state-determined systems ("Wire up the black box to make it spell out ILLIO"), his provocative homework assignments, and his imaginative handouts. In this chapter are given several of the best handouts, none previously available publicly. Ashby's British spellings are copied here faithfully.

The first was actually posted on his office door for a time. The next three provided practice in understanding the dynamics of state-determined systems (and carried some messages besides). His handout on ASS (Automatic Self-Strategizer) will allow the reader to construct a very simple game-playing machine which learns from its experience. One of the most interesting aspects of ASS is that it is never told explicitly what the object of the game is, yet it eventually learns to play flawlessly.

"How Wrong Can You Get?" is a continuation of "The Brain of Yesterday and Today" from the last chapter and will give the reader some good exercise. "Ashby Says" is a collection of Ashby's bits of cybernetic wisdom. The book concludes with Ashby's list of research possibilities for the future. Not many of these problems have been solved in the decade since it was written! So, Gentle Reader, help yourself.



The Dynamics of Personality

HIS personality includes the following traits, foibles, and responses:-

* If startled, he is apt to jump and knock the ash-tray to the floor.
* If asked to take a woman out to a dance he always says yes.
* If he feels sorry for a woman he takes a bunch of flowers to her.
* Going to a dance makes him feel good-natured and affectionate.
* If he is given burnt food he always points out the fault.
* When he sees a woman in tears he feels sorry for her.
* If anyone admits their incompetence to him he replies that they ought to be ashamed of themselves.
* He thinks anyone in an irritated state is best cured by being told to control their temper.
* If anyone should throw things at him he would hit back.
* He thinks a wife who could leave her husband is hopelessly bad, and should be told so.

HER personality includes the following traits, foibles, and responses:-

* Going to a dance makes her feel good-natured and affectionate.
* If a fault in her cooking is pointed out she admits to being an incompetent cook.
* The sight of tobacco ash on the floor irritates her.
* When flowers are given to her, her response is "Let's go out to a dance."
* To be told that she ought to be ashamed of herself would make her burst into tears.
* Being told to control her temper makes her really mad.
* She could never live with a man who thought her hopelessly bad.
* When she gets really mad she throws things at whatever annoys her.
* If her husband were to hit her she would go back to her mother.

Trace the two sequences of events that follow the two initial events:-

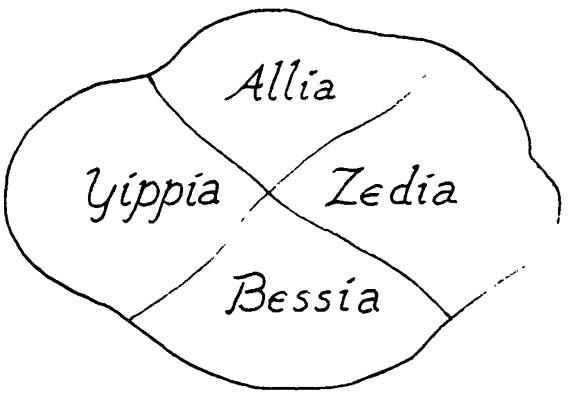
* She comes in unexpectedly while he is smoking and startles him;
* The dinner she has prepared proves to be burnt.

A Brief History of Amasia

- a study in the logic of events

by W. Ross Ashby

Map of the Continent of Amasia and its Four Countries



Words of standardised meaning in Amasian diplomacy:-

A Frontier: one of the four lines of contact.
A Protest.
Refer to Arbitration.
Declare its Support (of another country).
To Mobilise (on a frontier) against (the other country).
A Defenceless frontier.

It is with a feeling of sadness that I write this History, for I have to record how four countries tried to avoid war, did all they could to avoid it, yet nonetheles fell into it. The reader, I hope, may find it of interest and not entirely uninstructive.

First I must say something of the geography and basic military ideas of the four countries.

The continent of Amasia had been divided since the earliest times by two straight lines, giving four Frontiers that divided the continent into four countries: Allia, Bessia, Yippia, and Zedia. Each country thus had two neighbors (for the central point was of no practical importance.)

None of the countries really trusted the others, but as peace had lasted

for some time, all the four national armies were scattered each about its own country. The countries were sensitive, however, and if a soldier of one country were to enter another country, a Protest was sure to be addressed to the soldier's Government; but no one regarded this as serious, for a Protest was not an invasion.

None of the countries kept a large army. In fact, each army was only large enough, when mobilised, to extend along one of its two frontiers; and "mobilisation" was in fact always directed specifically to one of the two. This left the other frontier "defenceless" (a technical word), and much of Amasia's politico-military problems were concerned with the defenceless frontier. All four were agreed that if a country found its neighbor mobilising on their common frontier, then, in self-defence, it must also mobilise on the same frontier. Considering what to do if a neighbor should invade on the side that had become defenceless, all four General Staffs agreed that the attempt to turn the army round would lead only to chaos, so all it could do was to attack forwards, invade the country it faced, and thus perhaps find room for manoever.

All four Governments (with public opinion behind them) generally held that, if war broke out, two countries fighting a common enemy would automatically become allies. (So much for the basic conditions of the countries.)

In the course of time, with the occurrence of small frontier disputes tending to antagonise neighbors, Allia and Bessia had become somewhat hostile toward Yippia and Zedia, and friendly to each other (having no common source of friction.) Yippia and Zedia had similarly tended to draw somewhat closer, but nothing had been arranged officially between either pair.

In 1973, however, a considerable amount of vague political restlessness was evident, though no country wanted war. The government of Allia, under strong pressure from its own Peace Party, decided that it would not reply directly to any Protest addressed to it by its neighbors but would refer it to Bessia for arbitration.

The restlessness continued. Alarmist rumors of mobilisations began to appear in the newspapers. To forestall any such rash act, Zedia announced publicly that if any country mobilised against Yippia, that country would at once be mobilised against by Zedia.

Talk of mobilisations faded somewhat, but threats of Declaration of support began to be heard. (None had been made before.) Bessia tried to suppress these rather provocative threats by announcing that if Yippia declared its support of Zedia, or vice versa, then Bessia would at once declare its support for Allia. To reinforce Bessia's warning, Yippia added that if any country issued a Declaration of Support, Yippia would at once mobilise on its common frontier with that country (if it had one).

These matters were all public, but a private poll taken in Zedia at that time showed that were Bessia to be asked to arbitrate in a matter involving Yippia, public opinion would overwhelmingly demand that the Zedian Government declare its support for Yippia. This fact, however, was little suspected at the time; it was only the events themselves that converted the potential to the actual.

It was at this uneasy time that the United Nations decided to ensure peace in a really positive way. So was passed Resolution B, binding upon all parties:

> No country may mobilise on its neighbor's defenceless frontier. Should any country so act, its two neighbors will immediately declare war on it and invade it.

So was the danger of war removed for all time, apparently.

After recording these high and statesmanlike activities, I find it almost ridiculous to have to report that the next most significant event of the century was that Private Aldunk, of the Allian army, staggered half-drunk over the Yippian frontier and had to be forcibly ejected by Private Yipsky. But so it was. What followed, of course, stands to reason.

The Egyptian Steam Engine
(A Machine Is That Which Behaves As a Machine.)

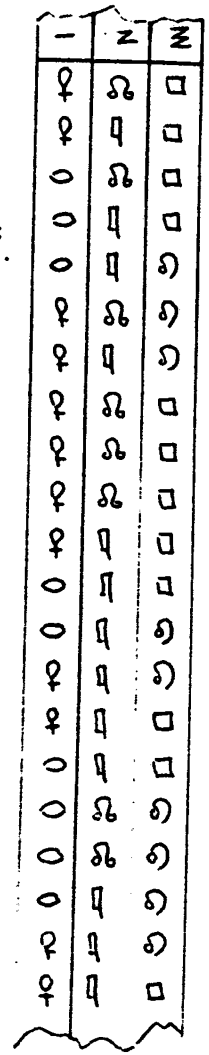
The strip of papyrus on the right is believed to be the protocol of an Ancient Egyptian steam engine. All that is known is that it reads from above downwards. Apparently there are three variables recorded: —, ≍, and ≋, each of which can take (or is recorded at) only two values:

— takes only or

≍ takes only or

≋ takes only or .

From the protocol, however, a good deal can be deduced. We can test, for instance, whether or not it is behaving as a "machine with input", and if so, which the input is. Thus we can identify the throttle (as one of the three variables), and we can then say which of the two symbols means "open" and which means "closed".



ASS - Automatic Self-Strategizer

- a mechanical game-player that develops its own strategy and ends by winning always.

- a proof by demonstration of the theorem that a dynamic system can be, at one and the same time, a blind runner-down to equilibrium and also an intelligent discoverer of how to play successfully.

ASS will master any one of a whole class of games (as the Patience-games are a class). Though devoid of special knowledge or construction for the game when it starts playing, it improves its skill with experience until it wins always. Some of the games in the class have goals that are just the opposite of those of other games and demand opposite strategies: ASS always develops the strategy that is appropriate.

Vocabulary and definitions:

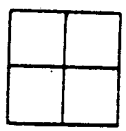
One PLAY = a succession of MOVES according to the laws of one Game, towards a defined Goal, ending in a win for one player (e.g. one Hand in Bridge).

One GAME = a set of laws and a defined Goal; many Plays may occur within one Game. (Chess = one Game; Bridge is another.)

Rules of the YU-ME class of games.

(1) There are two players: White and Black. (We shall assume that ASS plays White always; the other case is merely less interesting.)

(2) There is a Board of four cells:



(3) White moves first by marking one of the cells 0. Black moves next by marking one of the remaining cells X. White then moves by marking a remaining cell 0. Black puts X into the remaining cell.

(4) The Board, now for instance perhaps at

X	O
X	O

is compared with the particular Game's Goal, This might be, for instance, that White wins if he can make the final position one of

0	X	or	X	X	or	0	X
X	0		0	0		0	X

In this case, White has lost, for he has failed to bring the Board to one to one of the Goal positions.

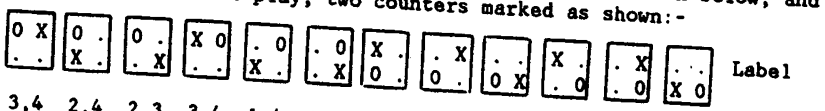
(5) The Goal (the set of positions aimed at by White) is specified before the play starts; it is known to both players throughout the Play. It

defines which one of the many possible YU-ME class of games is being played.

Construction of ASS

It has one "First-move" Box: a cup containing 4 counters labelled 1,2,3,4.

It has 12 "Second-move" Boxes. Each has a label, as shown below, and contains, at the start of play, two counters marked as shown:-

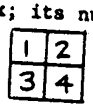


Label

3,4 2,4 2,3 3,4 1,4 1,3 2,4 1,4 1,2 2,3 1,3 1,2 Counters

Behavior of ASS

(1) ASS (playing White) decides on its first move by emitting a counter at random from the First-move Box; its number names the cell of ASS' first move according to the convention



The counter is left outside the Box.

(2) (Black, a human opponent of extreme ingenuity and malevolence, makes his reply.)

(3) The state of the Board now specifies a Second-move Box: ASS decides on its second move by emitting a counter at random from that Box; the number names the cell of ASS' second move. The counter is left outside its Box.

(4) (Black makes the final move.)

(5) (The Goal now determines whether ASS has won or lost.)

Feedback Mechanism

(activated by the 1-bit information
of whether ASS has won or lost.)

WON: If ASS has won, the two used counters are replaced in their respective Boxes.

LOST: If ASS has lost:-

(i) Remove permanently (for the duration of the Game and learning-process) the counter outside the Second-move Box.

(ii) If the removed Second-move counter was the last in the Box, remove also the counter outside the First-move Box; otherwise return it to the First-move Box.

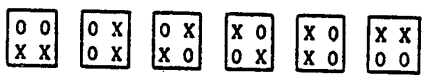
If a new Game is started, by specification of a new Goal, then ASS must be restored to its original state with its full complement of counters, as defined in "Construction".

The specification of ASS is now complete. Notice that nothing is specified to suit any particular Goal. Its moves are determined by the distribution of counters in the boxes; this distribution corresponds to, and specifies, its "strategy" (in the technical sense). The distribution changes as the plays succeed. It is easily verified that:-

ASS' strategies change only for the better;
they change till it wins always.

The Versatility of ASS

To demonstrate that ASS will adapt to any set Goal, notice first that there are only six possible endings to any play:



For simplicity, let us restrict the Goals to those that offer three of the six as wins for White, i.e. in 20, ways. Of the 20, eight are degenerate and give an excessively simple game; the remaining twelve can conveniently be generated by taking first a random choice among the six lines (by rolling a die, say), and then spinning a coin to decide whether ASS shall aim for the right-hand three or the left-hand.

1	0 0 X X	0 X 0 X	X 0 X 0		0 X X 0	X 0 0 X	X X 0 0
2	0 X 0 X	X 0 0 X	X 0 X 0		0 0 X X	0 X X 0	X X 0 0
3	0 X X 0	X 0 0 X	X 0 X 0		0 0 X X	0 X 0 X	X X 0 0
4	0 0 X X	X 0 X 0	X X 0 0		0 X X 0	0 X 0 X	X 0 0 X
5	0 X 0 X	X 0 X 0	X X 0 0		0 0 X X	0 X X 0	X 0 0 X
6	0 0 X X	X 0 0 X	X X 0 0		0 X 0 X	0 X X 0	X 0 X 0

Comments

(1) The versatility of ASS is shown by the fact that whether a given set of positions is to be aimed at or avoided, ASS will develop the appropriate strategy.

(2) The designing of ASS draws no details from the Goal that ASS will adapt to (when learning to play a particular Game.)

(3) No information comes to ASS other than what it can deduce from the 1-bit statement that it has won or lost.

(4) The machine is at one and the same time a blind automaton going helplessly to a steady state, at which it sticks; and a player that works up its own strategy till it wins.

(5) At the time of writing, ASS has found correct strategies for Games that no living person has yet analysed. (That people may perhaps analyse them later and find how right ASS was does not affect the proof that ASS is, in its own field, an intellectual leader.)

Acknowledgements

ASS is merely a much simplified version of MENACE, invented by Donald Michie, that automatically develops the proper strategy for Tic-Tac-Toe (in England: Noughts and Crosses). Martin Gardner simplified Menace to HER, which similarly develops the correct strategy for a microscopic version of Chess. NIMBLE was made by Stuart C. Hight similarly to learn to play Nim. ASS is an attempt to get the system so simple that the strategy can actually be developed before a spectator without boring him.

Michie, D. "Trial and error." Science Survey 1961, Part 2. Penguin Books, London, 1961.

M. Gardner, "Mathematical Games." Scientific American, 206, 138, #3, March 1962.

How Wrong Can You Get?

Our subject is still, alas, riddled with errors. Every one of the phrases below is in error to an almost ludicrous degree, to a degree likely to imperil any work based on it. Yet many have been taken from apparently reputable sources. (Notice that the error in each is gross, not a mere hair-splitting.)

A system that is regulatory (error-correcting) must have feedback from the error.

For a memory to be effective later it must have been stored. A million independent items (of information) demands a million sites for storage.

The problem of information-retrieval is to get the information first to its proper place and then to get it out again.

(If fractional signals are impossible:) The transmission of 1.5 bits requires at least 2 signals.

In pattern-recognition, the first step is to take the picture (or other presentation) and record or store it.

(On self-repairing machines:) The first step is to locate the fault.

"As this behavior is determinate, the system producing it must be determinate."

The purpose of the brain is to think.

"No action is performed by man without his anticipating (predicting) the results of that action."

Natural selection acts to further the individual's chance of survival.

Natural selection acts to further the species' chance of survival.

"Memory" is our power of recall.

(Assuming that human beings are goal-seeking:) "... so the society they form must be goal-seeking."

The wise man foresees the future and takes the appropriate action.

(Written in 1628, but still current:) "A child's soul is yet a white paper unscribbled with observations of the world, wherewith at length it becomes a blurred note-book."

(...enough to go on with...)

Ashby Says

On Science

* Science is the Observer's Digest.
* The Cyberneticist observes what might have happened but did not.
* A System is a set of variables sufficiently isolated to stay constant long enough for us to discuss it.

On Man

* The division of the World's system into Natural and Man-made died with Darwin.
* Man is not the measure: First comes the measure, then we see where he falls; so far the result has always been humiliating.

On Evolution

* The brain is merely nature's latest means of self-preservation.
* The goals of a species, such as homo, are what natural selection has driven it to.
* Poor M. Jourdain! He now has to understand that he has been behaving homestatically all his life, when he thought he was merely minding his business.

On Psychology

* For two thousand years psychology was a simple description of man's highest faculties - most of which he does not possess.
* The scientist does not believe in effects without causes; not even when they happen in the brain.

On Introspection

* A man no more knows how he thinks, just because he has a brain in his skull, than he knows how he makes blood, because he has marrow in his bones.

On the Brain

* The brain has no brain inside to guide it.
* To think is to act - inside the brain.
* The brain controls nothing - it transmits.
* The brain organizes nothing - it acts.
* The brain has no gimmick, just five billion years of research and development.

On Neurophysiology

* Natural selection insists that the nature of the parts shall be irrelevant for the behavior of the whole.
* The neuron is the one unit that, in behavior, is quite devoid of interest today; it is too small to be noticeable in a man's action, and too gross to carry a trace of memory.
* No mammal will ever understand the mammalian brain completely.
* The man who talks today of probability in the brain is usually trying to return to the days when everything in the brain was so delightfully vague.

On Learning

* No man knows what to do against the really new.
* All wisdom is wisdom after the event.
* Every system changes its mind by breaking.
* The educated brain is the wreckage left after the experiences of training.

On Memory

* Don't appoint, as the President's driver, and Englishman who has spent thirty years learning to drive on the left.
* A system that stores its memories away from their site of action must do much work remembering where it put that memory.

On Intelligence

* Today, those who don't know what "intelligence" means must give way to those who do.
* A mechanism is "brain-like" so far as it is effective.
* The only people who talk today of "real" intelligence are those who hope to find a meaning for the adjective later.
* Intelligent is as intelligent does.
* That the brain matches its environment is no more surprising than the match between the two ends of broken stick.
* Change the environment to its opposite and every piece of wisdom becomes the worst of folly.
* For every bump of the phrenologist there exists an environment that demands a depression.
* Everyone is world champion at some game (although some of the games have not yet been recognized.)
* An intelligence test measures the degree to which tester and subject think alike.
* Is there a general intelligence? - A universal weapon is as likely.

On Logic

* A man can be a pure logician only if it makes him feel good.
* Every skilled dramatist understands the inexorable logic of the emotion.

On Artificial Intelligence

* He who would design a good brain must first know how to make a bad one.
* Pattern-recognition is a throwing away of information.
* Any Device that can lose information can generalize.

On Deduction

* Deduction is the running-down of a determinate mechanism.
* Newton arrived at F=MA after a part-random search; the apple arrived at the ground by pure deduction.

On Computers

* The general purpose computer is freer than the trained brain.
* Today's digital computer is organized like an army of a million men that can only get two into action at a time.

On Organization

* It is an open question which has the richer organization: A living cow or a working silo.
* Which biological organization proved more resistant to the Spaniards: The Aztecs of Mexico, or the jungle of the Amazon?
* Can a system be self-organizing? - No system can permanently have the property that it changes its properties.

On Requisite Variety

* Man adapts by conquering the reducible; the irreducible is impregnable.
* We have broken into the Aladdin's cave of brain-like mechanisms; we find we can have anything we like - provided we pay for it!

Unsolved Problems in Cybernetics and Subjects for Exploration

Cylindrance

* Find sets of relations with corresponding sets of operators such that cylindrance is not increased.
* The relation between cylindrance and homomorphisms: what happens to cylindrance when a relation is "simplified"?
* Develop a theory of high-order interactions, whether in sets, communications, analyses of variance, differential calculus, etc.
* Relate the cylindrance of the environment to that of the brain that adapts to it.

Epistemology

* Can an additive measure be defined for the "complexity" of a machine, by starting with the axiom that two identical and unconnected machines should have a measure just twice that of the one?
* The Diagnosis and Homing problems (A.Gill) when the system studying them is itself a finite state machine.
* Develop a rigorous theory of how to "simplify" a system or machine, and apply it to the machine whose variables are continuous.
* Why, when there are such a vast number of possible binary relations, do we pick on a special few for consideration -- equivalence, order, identity, single-valued, etc. --? These evidently have some special property: what is it?
* May the "meaning" of a message properly be identified with what the recipient does about it?

The Subjective

* What special relations does a system have to itself?
* What facts are relevant to the question whether a machine can feel pain?
* What facts are relevant to the question whether the Roman Empire was self- conscious?

Topology

* Construct a topology of machines in which "A is near B" corresponds to "machine A is 'like' machine B". Develop a suitable metric.
* How many topologies, defined by their open sets, are possible on a finite number of points? What is its order of magnitude as $n \to \infty$?
* What relations hold between the local rules that determine connexion in a random net and the consequent large-scale patterns of connexion?

Finite State Machines

* The theory of complex equilibria and steady states, and their internal structures.
* Extend the theory (of the finite state machine) to the statistical case, over a population of machines.
* Are there properties peculiar to the Markovian machine (with input), or is it just a determinate system blurred?

The Polystable System

* Its properties, when isolated, as seen in various ways.
* Its responses, when disturbed by an input, as seen in various ways.
* The distribution of activity over a network of parts that show much clamping.
* The effects on a system of some variables being able to admit noise into the system.
* The properties of systems that store each memory at the site of its action.

The Random Net

* The relation between the specification of the parts and (after the parts have been joined) the shapes of the resulting confluents.
* Is there a relation between the distribution of loops formed by the connexions and the distribution of cycles it shows in its behavior?
* What are the conditions for the occurrence, in a random net, of cycles whose length is a large prime number?
* What invariants are there over such nets as Walker's? -- like energy and angular momentum over Newtonian systems.
* Translate the properties of the wholly discrete random net to the case in which both variables and times are continuous.
* What are the differences between the two types:
  (a) when the net's k internal connexions are unchanging in position?
  (b) when there are k connexions at every moment, but their distribution changes with time?
* Extend the theory of why random nets tend to show habituation.
* Extend the theory of why random nets tend to change towards internal disconnexion.
* Are there critical degrees of connectedness between which it shows a behavioral analog of the "liquid" state?
* Will a random net show the sudden change associated with Reynold's number?
* Explain the "mesa" phenomenon of Minsky and Selfridge. (Proc. 4-th London Symp. on Information Theory; ed. Cherry, C. Academic Press, N.Y., 1961.)

Information Theory

* The non-ergodic and non-stationary cases.
* Characterise the error-controlled feedback regulator by the quantities of information flowing internally.

* Apply Shannon's tenth theorem (law of requisite transmission) to game theory.

Probability

* What is the probability that a linear system will be stable?
* (of 1) - To what order of magnitude does the probability tend as n tends to infinity?
* Extend Rubin & Sitgreaves' results to special cases with more constraints.

Physics

* Is it true that a steady thruput of energy tends to change a heterogeneous system toward such forms as maximally delay the escape of energy?
* Do the laws of physics force the systems governed by them to be of low cylindrance"?
* Make clear the relation between "order" as seen by the physicist and "order" as seen by the biologist.

Neuro-topology

* How can the topology of ordinary space (a three-dimensional Euclidean continuum) be represented on a network of neurons connected partly at random?

Physiology

* Does the brain have generators of "random" values, to get originality when making trials?
* What methods does the brain use, in serial adaptation, to ensure that later adaptations do not over-write, and spoil, earlier adaptations?
* Does pain, and the restlessness provoked by it, correspond to activity in second-order feedback (ultrastability)?
* What is the dynamic consequence of the layering so commonly seen in neural (especially sensory) tissue?

Psychology

* Can regression to simpler or earlier behavior be shown as a general property of dynamic systems?
* How fast does adaptation occur when it is composed of many small adaptations occurring simultaneously? Extend Lerner's result (velocity reduced to $1/\sqrt{n}$) from the linear and additive to the discrete and arbitrary. (Lerner, I.M., Population Genetics and Animal Improvement. Cambridge Univ. Press, 1950.)
* Follows up various lines in:
  (a) Recent Progress in Pscyhiatry; ed. Fleming, G.W.T.H., Churchill, London vol. 2, pp. 94-110, 1950.
  (b) Ibid., vol. 3, pp. 94-117, 1958.
  (c) Journal of Mental Science, 100, 114-124, 1954. - by W.R. Ashby.

Sociology

* Apply uncertainty analysis to find conditional transmissions in social activities.
* Can such "cerebral" activities as habituation, anticipation, conditioning, be demonstrated in contemporary social systems? -- not as a personal but as a social or system event.
* (as 2) -- in history?

Constructions

* Build a machine or organization that will beat the world Champion at chess.
* Build a model that will adequately represent Freud's theory of the neuroses.

PUBLICATIONS OF W. ROSS ASHBY

A. Books

B1. Design for a Brain, John Wiley and Sons, New York, 260 pp. (1952); Second Edition (1960); Russian Edition (1958); Spanish Edition (1959); Japanese Edition (1963), and others.

B2. An Introduction to Cybernetics, John Wiley and Sons, New York, 296 pp. (1956); Third Imprint (1958); Russian Edition (1957); French Edition (1957); Spanish Edition (1958); Czech Edition (1959); Polish Edition (1959); Hungarian Edition (1960); German Edition (1965); Bulgarian Edition (1966); Italian Edition (1966), and others.

B. Articles

1. "A Cell for the Measurement of the Specific Conductivity of the Blood Serum," Biochemical Journal, 24, 1557-1559 (1930).
2. "O-agglutinins in Enterica Carriers," J. of Pathology and Bacteriology, 34, 109 (1931).
3. With R.M. Steward, "Angiema Arteriale Racemosum in an Acollosal Brain," J. of Neurology and Psychopathology, 11, 289 (1931).
4. "The 'Path' Theory of Cortical Function," J. of Neurology and Psychopathology, 12, 148-157 (1931).
5. "The Physiological Basis of the Neuroses," Proceedings of the Royal Society of Medicine, 26, 1454-1460 (1933).
6. "Size in Mental Deficiency," J. of Neurology and Psychopathology, 13, 303-329 (1933).
7. With R.M. Steward, "The Corpus Callosum in its Relation to Intelligence," J. of Neurology and Psychopathology, 14, 217-226 (1934).
8. "On the Nature of Inhibition," J. of Mental Science, 80, 198-223 (1934).
9. With A. Glynn, "The Chemistry of the Brain in the Mental Defective," J. of Neurology and Psychopathology, 15, 193-209 (1935).
10. The Thickness of the Cerebral Cortex and its Layers in the Mental Defective, M.A. Thesis, University of Cambridge (1935).
11. "The Width of the Convolutions in the Normal and Defective Person," J. of Neurology and Psychopathology, 16, 26-35 (1935).
12. With G. de M. Rudolf, "Nasal Diptheria Carriers," J. of Hygiene, 36, 129-139 (1936).
13. "Tissue Culture Methods in the Study of the Nervous System," J. of Neurology and Psychopathology, 17, 322 (1937).
14. With R.M. Steward and J.H. Watkin, "Chondro-osteo-dystrophy of the Hurler Type: A Pathology Study," Brain, 60, 149-179 (1937).
15. "Adaptiveness and Equilibrium," J. of Mental Science, 86, 478-483 (1940).
16. "The Effect of Controls on Stability," Nature, 155, 242-243 (1945).
17. "The Physical Origin of Adaptation by Trial and Error," J. of General Psychology, 32, 13-25 (1945).
18. "Principles for the Quantitative Study of Stability in a Dynamic Whole System," J. of Mental Science, 92, 319-323 (1946).
19. "The Behavioral Properties of Systems in Equilibrium," Amer. J. of Psychology, 59, 682-686 (1946).

20. "Principles of the Self-Organizing Dynamic System," J. of General Psychology, 37, 125-128 (1947).
21. "Interrelations between Stabilities of Parts within a Whole Dynamic System," J. of Comparative and Physiological Psychology, 40, 1-7 (1947).
22. "The Nervous System as Physical Machine, with Special Reference to the Origin of Adaptive Behavior," Mind, 56, 44-59 (1947).
23. "Dynamics of the Cerebral Cortex: Automatic Development of Equilibrium in Self-Organizing Systems," Psychometrika, 12 (2), 135-140 (1947).
24. "Existence of Critical Levels for the Actions of Hormones and Enzymes," J. of Mental Science, 93, 733-739 (1947).
25. "Adrenal Cortical Steroids and the Metabolism of Glutamic Acid in Gluconeogenesis," J. of Mental Science, 95, 153-161 (1949).
26. "Effects of Convulsive Therapy on the Excretion of Cortins and Ketosteroids," J. of Mental Science, 95, 275-324 (1949).
27. With M. Bassett, "The Effect of Leucotomy on Creative Ability," J. of Mental Science, 95, 418-430 (1949).
28. "The Cerebral Mechanisms of Intelligent Action," in Perspectives in Neuro-Psychiatry, D. Richter (ed.), H. Lewis & Co., London (1950).
29. "A New Mechanism which Shows Simple Conditioning," J. of Psychology, 29, 343-347 (1950).
30. With M. Bassett, "The Effect of Prefrontal Leucotomy on the Psychogalvanic Response," J. of Mental Science, 96, 458-469 (1950).
31. "Cybernetics" in Recent Progress in Psychiatry II, G. Fleming (ed.), London (1950).
32. "Stability of a Randomly Assembled Nerve Network," Electroencephalography and Clinical Neurophysiology, 2, 471-482 (1950).
33. "The Homeostat," in Les Machines a Calcular et la Pensee Humaine, Paris (1951).
34. "Statistical Machinery," Thales, 7, 1-8 (1951).
35. "Can a Mechanical Chess-player Outplay its Designer?," British J. for the Philosophy of Science, 3, 44-57 (1952).
36. "Adrenal Cortical Function and Response to Convulsive Therapy in a Case of Periodic Catatonia," J. of Mental Science, 98, 81-99 (1952).
37. "The Mode of Action of Electro-convulsive Therapy," J. of Mental Science, 99, 202-215 (1953).
38. "The Application of Cybernetics to Psychiatry," J. of Mental Science, 100, 114-124 (1954).
39. With M. Bassett, "The Effect of Electro-convulsive Therapy on the Psychogalvanic Response," J. of Mental Science, 100, 632-642 (1954).
40. "The Effect of Experience on a Determinate Dynamic System," Behavioral Science, 1, 35-42 (1956).
41. "Design for an Intelligence-amplifier" in Automata Studies, C.E. Shannon and J. McCarthy (eds.), Princeton University Press, Princeton, N.J., pp. 215-233 (1956).
42. "Cybernetics," in Recent Progress in Psychiatry III, Fleming (ed.) London, pp. 94-117 (1958).
43. "Requisite Variety, and Its Implications for the Control of Complex Systems," Cybernetica, 1 (2), 1-17 (1958).
44. "General Systems Theory as a New Discipline," General Systems, 3, 1-6 (1958).
45. "The Mechanism of Habituation," in N.P.L. Symposium on the Mechanism of Thought Process, C. Cherry (ed.), London, pp. 1-21 (1959).

46. "Applications of Cybernetics to Biology and Sociology," in (trans Vosprosii Philosophii, Moscow (1959).
47. "The Brain as Regulator," Nature, 186, 413 (1960).
48. "Computers and Decision-making," New Scientist, 7, 746 (1960).
49. "Cybernetics in Medicine," in Proceedings of the First Internation Congress on Cybernetic Medicine, A. Mastruzo (ed.), Naples, Italy, p 179-180 (1960).
50. "The Relativity of 'Meaning'," Nature, 187, 532, (1960).
51. "Homeostasis," in Encyclopedia of Biological Sciences, P. Gray (ed. Reinhold Publishing Co., New York (1961).
52. "The Avoidance of Over-writing in Self-organizing Systems," J. Theoretical Biology, 1, 431-439 (1961).
53. "What is an Intelligent Machine?," Proceedings of the Western Joi Computer Conference, Los Angeles, pp. 275-280 (1961).
54. "Brain and Computer," Third Conference of the International Associ tion of Cybernetics, September (1961).
55. "Cybernetics Today," The 1961 FIER Distinguished Lecture, brochu printed by the Fdn. for Education and Research, N.Y., pp. 1-16 (1961
56. "Brain" in the Grolier Encyclopedia, Grolier Society, Inc. (1961).
57. "General System Theory and the Problem of the Black Box," R gelungsvorgange Lebender Wesen, R. Oldenbourg Verlag, Munich, (1961)
58. "Principles of the Self-organizing System," in Principles of Sel organization, H. Von Foerster and G.W. Zopf, Jr. (eds.), Pergam Press, New York, pp. 255-278 (1962).
59. "Frontiers of Integrated Automatic Control - What Can We Learn fr the Brain?," in Proceedings of the Joint Automatic Control Conferenc A.E.E.E., New York, pp. 1-3 (1962).
60. "The Self Reproducing System," in Aspects of the Theory of Artifici Intelligence, C.A. Muses (ed.) Plenum Press, New York, pp. 9- (1962).
61. With H. Von Foerster and C.C. Walker, "Instability of Pulse Activi in a Net with Threshold," Nature, 196, 561-562 (1962).
62. "Simulation of a Brain," in Computer Applications in the Behavior Sciences, H. Borko (ed.) Prentice-Hall, Englewood Cliffs, New Jerse pp. 452-465 (1962).
63. "Induction, Prediction, and Decision-making in Cybernetic Systems," Induction: Some Current Issues, H.E. Kyburg and E. Nagel (eds. Wesleyan University Press, Middletown, Connecticut, pp. 55-66 (1963)
64. With H. Von Foerster and C.C. Walker, "The Essential Instability Systems with Threshold and Some Possible Applications to Psychiatry in Nerve, Brain and Memory Models, N. Wiener and J.P. Schade (eds. Elsevier Press, Amsterdam, The Netherlands, pp. 236-243 (1963).
65. "Systems and Information," IEEE Trans. on Military Electronics, MIL- 94-97 (1963).
66. "Modeling the Brain," in Proceedings of the IBM Scientific Symposi on Simulation Models and Gaming, IBM Corp., New York, pp. 195-2 (1964).
67. "Constraint Analysis of Many-Dimensional Relations," in Progress Bio-Cybernetics II, N. Wiener and J.P. Schade (eds.), Elsevier Pub lishing Co., Amsterdam, pp. 10-18 (1965). Also published in Genera Systems, 9, 83-99 (1964).
68. With H. Von Foerster, "Biological Computers," in Bioastronautics, K. Schafer (ed.), The Macmillan Co., New York, 333-360 (1964).

"Principles of the Self-Organizing Dynamic System," J. of General Psychology, 37, 125-128 (1947).
"Interrelations between Stabilities of Parts within a Whole Dynamic System," J. of Comparative and Physiological Psychology, 40, 1-7 (1947).
"The Nervous System as Physical Machine, with Special Reference to the Origin of Adaptive Behavior," Mind, 56, 44-59 (1947).
"Dynamics of the Cerebral Cortex: Automatic Development of Equilibrium in Self-Organizing Systems," Psychometrika, 12 (2), 135-140 (1947).
"Existence of Critical Levels for the Actions of Hormones and Enzymes," J. of Mental Science, 93, 733-739 (1947).
"Adrenal Cortical Steroids and the Metabolism of Glutamic Acid in Gluconeogenesis," J. of Mental Science, 95, 153-161 (1949).
"Effects of Convulsive Therapy on the Excretion of Cortins and Ketosteroids," J. of Mental Science, 95, 275-324 (1949).
With M. Bassett, "The Effect of Leucotomy on Creative Ability," J. of Mental Science, 95, 418-430 (1949).
"The Cerebral Mechanisms of Intelligent Action," in Perspectives in Neuro-Psychiatry, D. Richter (ed.), H. Lewis & Co., London (1950).
"A New Mechanism which Shows Simple Conditioning," J. of Psychology, 29, 343-347 (1950).
With M. Bassett, "The Effect of Prefrontal Leucotomy on the Psychogalvanic Response," J. of Mental Science, 96, 458-469 (1950).
"Cybernetics" in Recent Progress in Psychiatry II, G. Fleming (ed.), London (1950).
"Stability of a Randomly Assembled Nerve Network," Electroencephalography and Clinical Neurophysiology, 2, 471-482 (1950).
"The Homeostat," in Les Machines a Calcular et la Pensee Humaine, Paris (1951).
"Statistical Machinery," Thales, 7, 1-8 (1951).
"Can a Mechanical Chess-player Outplay its Designer?," British J. for the Philosophy of Science, 3, 44-57 (1952).
"Adrenal Cortical Function and Response to Convulsive Therapy in a Case of Periodic Catatonia," J. of Mental Science, 98, 81-99 (1952).
"The Mode of Action of Electro-convulsive Therapy," J. of Mental Science, 99, 202-215 (1953).
"The Application of Cybernetics to Psychiatry," J. of Mental Science, 100, 114-124 (1954).
With M. Bassett, "The Effect of Electro-convulsive Therapy on the Psychogalvanic Response," J. of Mental Science, 100, 632-642 (1954).
"The Effect of Experience on a Determinate Dynamic System," Behavioral Science, 1, 35-42 (1956).
"Design for an Intelligence-amplifier" in Automata Studies, C.E. Shannon and J. McCarthy (eds.), Princeton University Press, Princeton, N.J., pp. 215-233 (1956).
"Cybernetics," in Recent Progress in Psychiatry III, Fleming (ed.) London, pp. 94-117 (1958).
"Requisite Variety, and Its Implications for the Control of Complex Systems," Cybernetica, 1 (2), 1-17 (1958).
"General Systems Theory as a New Discipline," General Systems, 3, 1-6 (1958).
"The Mechanism of Habituation," in N.P.L. Symposium on the Mechanism of Thought Process, C. Cherry (ed.), London, pp. 1-21 (1959).

46. "Applications of Cybernetics to Biology and Sociology," in (trans.) Vosprosii Philosophii, Moscow (1959).
47. "The Brain as Regulator," Nature, 186, 413 (1960).
48. "Computers and Decision-making," New Scientist, 7, 746 (1960).
49. "Cybernetics in Medicine," in Proceedings of the First International Congress on Cybernetic Medicine, A. Mastruzo (ed.), Naples, Italy, pp. 179-180 (1960).
50. "The Relativity of 'Meaning'," Nature, 187, 532, (1960).
51. "Homeostasis," in Encyclopedia of Biological Sciences, P. Gray (ed.), Reinhold Publishing Co., New York (1961).
52. "The Avoidance of Over-writing in Self-organizing Systems," J. of Theoretical Biology, 1, 431-439 (1961).
53. "What is an Intelligent Machine?," Proceedings of the Western Joint Computer Conference, Los Angeles, pp. 275-280 (1961).
54. "Brain and Computer," Third Conference of the International Association of Cybernetics, September (1961).
55. "Cybernetics Today," The 1961 FIER Distinguished Lecture, brochure printed by the Fdn. for Education and Research, N.Y., pp. 1-16 (1961).
56. "Brain" in the Grolier Encyclopedia, Grolier Society, Inc. (1961).
57. "General System Theory and the Problem of the Black Box," Regelungsvorgange Lebender Wesen, R. Oldenbourg Verlag, Munich, (1961).
58. "Principles of the Self-organizing System," in Principles of Self-organization, H. Von Foerster and G.W. Zopf, Jr. (eds.), Pergamon Press, New York, pp. 255-278 (1962).
59. "Frontiers of Integrated Automatic Control - What Can We Learn from the Brain?," in Proceedings of the Joint Automatic Control Conference, A.E.E.E., New York, pp. 1-3 (1962).
60. "The Self Reproducing System," in Aspects of the Theory of Artificial Intelligence, C.A. Muses (ed.) Plenum Press, New York, pp. 9-18 (1962).
61. With H. Von Foerster and C.C. Walker, "Instability of Pulse Activity in a Net with Threshold," Nature, 196, 561-562 (1962).
62. "Simulation of a Brain," in Computer Applications in the Behavioral Sciences, H. Borko (ed.) Prentice-Hall, Englewood Cliffs, New Jersey, pp. 452-465 (1962).
63. "Induction, Prediction, and Decision-making in Cybernetic Systems," in Induction: Some Current Issues, H.E. Kyburg and E. Nagel (eds.), Wesleyan University Press, Middletown, Connecticut, pp. 55-66 (1963).
64. With H. Von Foerster and C.C. Walker, "The Essential Instability of Systems with Threshold and Some Possible Applications to Psychiatry," in Nerve, Brain and Memory Models, N. Wiener and J.P. Schade (eds.), Elsevier Press, Amsterdam, The Netherlands, pp. 236-243 (1963).
65. "Systems and Information," IEEE Trans. on Military Electronics, MIL-7, 94-97 (1963).
66. "Modeling the Brain," in Proceedings of the IBM Scientific Symposium on Simulation Models and Gaming, IBM Corp., New York, pp. 195-208 (1964).
67. "Constraint Analysis of Many-Dimensional Relations," in Progress in Bio-Cybernetics II, N. Wiener and J.P. Schade (eds.), Elsevier Publishing Co., Amsterdam, pp. 10-18 (1965). Also published in General Systems, 9, 83-99 (1964).
68. With H. Von Foerster, "Biological Computers," in Bioastronautics, K.E. Schafer (ed.), The Macmillan Co., New York, 333-360 (1964).

20. "Principles of the Self-Organizing Dynamic System," J. of General Psychology, 37, 125-128 (1947).
21. "Interrelations between Stabilities of Parts within a Whole Dynamic System," J. of Comparative and Physiological Psychology, 40, 1-7 (1947).
22. "The Nervous System as Physical Machine, with Special Reference to the Origin of Adaptive Behavior," Mind, 56, 44-59 (1947).
23. "Dynamics of the Cerebral Cortex: Automatic Development of Equilibrium in Self-Organizing Systems," Psychometrika, 12 (2), 135-140 (1947).
24. "Existence of Critical Levels for the Actions of Hormones and Enzymes," J. of Mental Science, 93, 733-739 (1947).
25. "Adrenal Cortical Steroids and the Metabolism of Glutamic Acid in Gluconeogenesis," J. of Mental Science, 95, 153-161 (1949).
26. "Effects of Convulsive Therapy on the Excretion of Cortins and Ketosteroids," J. of Mental Science, 95, 275-324 (1949).
27. With M. Bassett, "The Effect of Leucotomy on Creative Ability," J. of Mental Science, 95, 418-430 (1949).
28. "The Cerebral Mechanisms of Intelligent Action," in Perspectives in Neuro-Psychiatry, D. Richter (ed.), H. Lewis & Co., London (1950).
29. "A New Mechanism which Shows Simple Conditioning," J. of Psychology, 29, 343-347 (1950).
30. With M. Bassett, "The Effect of Prefrontal Leucotomy on the Psychogalvanic Response," J. of Mental Science, 96, 458-469 (1950).
31. "Cybernetics" in Recent Progress in Psychiatry II, G. Fleming (ed.), London (1950).
32. "Stability of a Randomly Assembled Nerve Network," Electroencephalography and Clinical Neurophysiology, 2, 471-482 (1950).
33. "The Homeostat," in Les Machines a Calcular et la Pensee Humaine, Paris (1951).
34. "Statistical Machinery," Thales, 7, 1-8 (1951).
35. "Can a Mechanical Chess-player Outplay its Designer?," British J. for the Philosophy of Science, 3, 44-57 (1952).
36. "Adrenal Cortical Function and Response to Convulsive Therapy in a Case of Periodic Catatonia," J. of Mental Science, 98, 81-99 (1952).
37. "The Mode of Action of Electro-convulsive Therapy," J. of Mental Science, 99, 202-215 (1953).
38. "The Application of Cybernetics to Psychiatry," J. of Mental Science, 100, 114-124 (1954).
39. With M. Bassett, "The Effect of Electro-convulsive Therapy on the Psychogalvanic Response," J. of Mental Science, 100, 632-642 (1954).
40. "The Effect of Experience on a Determinate Dynamic System," Behavioral Science, 1, 35-42 (1956).
41. "Design for an Intelligence-amplifier" in Automata Studies, C.E. Shannon and J. McCarthy (eds.), Princeton University Press, Princeton, N.J., pp. 215-233 (1956).
42. "Cybernetics," in Recent Progress in Psychiatry III, Fleming (ed.) London, pp. 94-117 (1958).
43. "Requisite Variety, and Its Implications for the Control of Complex Systems," Cybernetica, 1 (2), 1-17 (1958).
44. "General Systems Theory as a New Discipline," General Systems, 3, 1-6 (1958).
45. "The Mechanism of Habituation," in N.P.L. Symposium on the Mechanism of Thought Process, C. Cherry (ed.), London, pp. 1-21 (1959).

46. "Applications of Cybernetics to Biology and Sociology," in (trans Vosprosii Philosophii, Moscow (1959).
47. "The Brain as Regulator," Nature, 186, 413 (1960).
48. "Computers and Decision-making," New Scientist, 7, 746 (1960).
49. "Cybernetics in Medicine," in Proceedings of the First Internation Congress on Cybernetic Medicine, A. Mastruzo (ed.), Naples, Italy, p 179-180 (1960).
50. "The Relativity of 'Meaning'," Nature, 187, 532, (1960).
51. "Homeostasis," in Encyclopedia of Biological Sciences, P. Gray (ed. Reinhold Publishing Co., New York (1961).
52. "The Avoidance of Over-writing in Self-organizing Systems," J. Theoretical Biology, 1, 431-439 (1961).
53. "What is an Intelligent Machine?," Proceedings of the Western Joi Computer Conference, Los Angeles, pp. 275-280 (1961).
54. "Brain and Computer," Third Conference of the International Associ tion of Cybernetics, September (1961).
55. "Cybernetics Today," The 1961 FIER Distinguished Lecture, brochu printed by the Fdn. for Education and Research, N.Y., pp. 1-16 (1961
56. "Brain" in the Grolier Encyclopedia, Grolier Society, Inc. (1961).
57. "General System Theory and the Problem of the Black Box," R gelungsvorgange Lebender Wesen, R. Oldenbourg Verlag, Munich, (1961)
58. "Principles of the Self-organizing System," in Principles of Sel organization, H. Von Foerster and G.W. Zopf, Jr. (eds.), Pergam Press, New York, pp. 255-278 (1962).
59. "Frontiers of Integrated Automatic Control - What Can We Learn fr the Brain?," in Proceedings of the Joint Automatic Control Conferenc A.E.E.E., New York, pp. 1-3 (1962).
60. "The Self Reproducing System," in Aspects of the Theory of Artifici Intelligence, C.A. Muses (ed.) Plenum Press, New York, pp. 9- (1962).
61. With H. Von Foerster and C.C. Walker, "Instability of Pulse Activi in a Net with Threshold," Nature, 196, 561-562 (1962).
62. "Simulation of a Brain," in Computer Applications in the Behavior Sciences, H. Borko (ed.) Prentice-Hall, Englewood Cliffs, New Jerse pp. 452-465 (1962).
63. "Induction, Prediction, and Decision-making in Cybernetic Systems," Induction: Some Current Issues, H.E. Kyburg and E. Nagel (eds. Wesleyan University Press, Middletown, Connecticut, pp. 55-66 (1963)
64. With H. Von Foerster and C.C. Walker, "The Essential Instability Systems with Threshold and Some Possible Applications to Psychiatry in Nerve, Brain and Memory Models, N. Wiener and J.P. Schade (eds. Elsevier Press, Amsterdam, The Netherlands, pp. 236-243 (1963).
65. "Systems and Information," IEEE Trans. on Military Electronics, MIL- 94-97 (1963).
66. "Modeling the Brain," in Proceedings of the IBM Scientific Symposi on Simulation Models and Gaming, IBM Corp., New York, pp. 195-2 (1964).
67. "Constraint Analysis of Many-Dimensional Relations," in Progress Bio-Cybernetics II, N. Wiener and J.P. Schade (eds.), Elsevier Pub lishing Co., Amsterdam, pp. 10-18 (1965). Also published in Genera Systems, 9, 83-99 (1964).
68. With H. Von Foerster, "Biological Computers," in Bioastronautics, K. Schafer (ed.), The Macmillan Co., New York, 333-360 (1964).

Principles of the Self-Organizing Dynamic System," J. of General sychology, 37, 125-128 (1947).
Interrelations between Stabilities of Parts within a Whole Dynamic ystem," J. of Comparative and Physiological Psychology, 40, 1-7 1947).
The Nervous System as Physical Machine, with Special Reference to the rigin of Adaptive Behavior," Mind, 56, 44-59 (1947).
Dynamics of the Cerebral Cortex: Automatic Development of Equilibri- n in Self-Organizing Systems," Psychometrika, 12 (2), 135-140 (1947).
Existence of Critical Levels for the Actions of Hormones and nzymes," J. of Mental Science, 93, 733-739 (1947).
Adrenal Cortical Steroids and the Metabolism of Glutamic Acid in Glu- oneogenesis," J. of Mental Science, 95, 153-161 (1949).
Effects of Convulsive Therapy on the Excretion of Cortins and Ketos- eroids," J. of Mental Science, 95, 275-324 (1949).
ith M. Bassett, "The Effect of Leucotomy on Creative Ability," J. of ental Science, 95, 418-430 (1949).
The Cerebral Mechanisms of Intelligent Action," in Perspectives in euro-Psychiatry, D. Richter (ed.), H. Lewis & Co., London (1950).
A New Mechanism which Shows Simple Conditioning," J. of Psychology, 9, 343-347 (1950).
ith M. Bassett, "The Effect of Prefrontal Leucotomy on the Psychogal- anic Response," J. of Mental Science, 96, 458-469 (1950).
Cybernetics" in Recent Progress in Psychiatry II, G. Fleming (ed.), ondon (1950).
Stability of a Randomly Assembled Nerve Network," Electroencephalog- aphy and Clinical Neurophysiology, 2, 471-482 (1950).
The Homeostat," in Les Machines a Calcular et la Pensee Humaine, aris (1951).
Statistical Machinery," Thales, 7, 1-8 (1951).
Can a Mechanical Chess-player Outplay its Designer?," British J. for he Philosophy of Science, 3, 44-57 (1952).
Adrenal Cortical Function and Response to Convulsive Therapy in a ase of Periodic Catatonia," J. of Mental Science, 98, 81-99 (1952).
The Mode of Action of Electro-convulsive Therapy," J. of Mental Sci- nce, 99, 202-215 (1953).
The Application of Cybernetics to Psychiatry," J. of Mental Science, 00, 114-124 (1954).
ith M. Bassett, "The Effect of Electro-convulsive Therapy on the sychogalvanic Response," J. of Mental Science, 100, 632-642 (1954).
The Effect of Experience on a Determinate Dynamic System," Behavioral cience, 1, 35-42 (1956).
"Design for an Intelligence-amplifier" in Automata Studies, C.E. Shan- non and J. McCarthy (eds.), Princeton University Press, Princeton, N.J., pp. 215-233 (1956).
"Cybernetics," in Recent Progress in Psychiatry III, Fleming (ed.) London, pp. 94-117 (1958).
"Requisite Variety, and Its Implications for the Control of Complex Systems," Cybernetica, 1 (2), 1-17 (1958).
"General Systems Theory as a New Discipline," General Systems, 3, 1-6 (1958).
"The Mechanism of Habituation," in N.P.L. Symposium on the Mechanism of Thought Process, C. Cherry (ed.), London, pp. 1-21 (1959).

46. "Applications of Cybernetics to Biology and Sociology," in (trans.) Vosprosii Philosophii, Moscow (1959).
47. "The Brain as Regulator," Nature, 186, 413 (1960).
48. "Computers and Decision-making," New Scientist, 7, 746 (1960).
49. "Cybernetics in Medicine," in Proceedings of the First International Congress on Cybernetic Medicine, A. Mastruzo (ed.), Naples, Italy, pp. 179-180 (1960).
50. "The Relativity of 'Meaning'," Nature, 187, 532, (1960).
51. "Homeostasis," in Encyclopedia of Biological Sciences, P. Gray (ed.), Reinhold Publishing Co., New York (1961).
52. "The Avoidance of Over-writing in Self-organizing Systems," J. of Theoretical Biology, 1, 431-439 (1961).
53. "What is an Intelligent Machine?," Proceedings of the Western Joint Computer Conference, Los Angeles, pp. 275-280 (1961).
54. "Brain and Computer," Third Conference of the International Association of Cybernetics, September (1961).
55. "Cybernetics Today," The 1961 FIER Distinguished Lecture, brochure printed by the Fdn. for Education and Research, N.Y., pp. 1-16 (1961).
56. "Brain" in the Grolier Encyclopedia, Grolier Society, Inc. (1961).
57. "General System Theory and the Problem of the Black Box," Regelungsvorgange Lebender Wesen, R. Oldenbourg Verlag, Munich, (1961).
58. "Principles of the Self-organizing System," in Principles of Self-organization, H. Von Foerster and G.W. Zopf, Jr. (eds.), Pergamon Press, New York, pp. 255-278 (1962).
59. "Frontiers of Integrated Automatic Control - What Can We Learn from the Brain?," in Proceedings of the Joint Automatic Control Conference, A.E.E.E., New York, pp. 1-3 (1962).
60. "The Self Reproducing System," in Aspects of the Theory of Artificial Intelligence, C.A. Muses (ed.) Plenum Press, New York, pp. 9-18 (1962).
61. With H. Von Foerster and C.C. Walker, "Instability of Pulse Activity in a Net with Threshold," Nature, 196, 561-562 (1962).
62. "Simulation of a Brain," in Computer Applications in the Behavioral Sciences, H. Borko (ed.) Prentice-Hall, Englewood Cliffs, New Jersey, pp. 452-465 (1962).
63. "Induction, Prediction, and Decision-making in Cybernetic Systems," in Induction: Some Current Issues, H.E. Kyburg and E. Nagel (eds.), Wesleyan University Press, Middletown, Connecticut, pp. 55-66 (1963).
64. With H. Von Foerster and C.C. Walker, "The Essential Instability of Systems with Threshold and Some Possible Applications to Psychiatry," in Nerve, Brain and Memory Models, N. Wiener and J.P. Schade (eds.), Elsevier Press, Amsterdam, The Netherlands, pp. 236-243 (1963).
65. "Systems and Information," IEEE Trans. on Military Electronics, MIL-7, 94-97 (1963).
66. "Modeling the Brain," in Proceedings of the IBM Scientific Symposium on Simulation Models and Gaming, IBM Corp., New York, pp. 195-208 (1964).
67. "Constraint Analysis of Many-Dimensional Relations," in Progress in Bio-Cybernetics II, N. Wiener and J.P. Schade (eds.), Elsevier Publishing Co., Amsterdam, pp. 10-18 (1965). Also published in General Systems, 9, 83-99 (1964).
68. With H. Von Foerster, "Biological Computers," in Bioastronautics, K.E. Schafer (ed.), The Macmillan Co., New York, 333-360 (1964).

69. "Introductory Remarks at Panel Discussion" in Views on General Systems Theory, M.D. Mesarovic (ed.) John Wiley & Sons, New York, 165-169 (1964).
70. "The Set Theory of Mechanism and Homeostasis," in Automaton Theory and Learning Systems, D.J. Stewart (ed.), Academic Press, London, pp. 23-51 (1967). Also in General Systems, 9, pp. 83-97 (1964).
71. "Measuring the Internal Informational Exchange in a System," Cybernetica, 8, 5-22 (1965).
72. "Mathematical Models and the Computer Analysis of the Function of the Central Nervous System," Annual Review of Physiology, 28, 89-106 (1966).
73. "The Cybernetic Viewpoint," IEEE Trans. Systems Sci. Cyb. SSC-2, (1), 7-8 (1966).
74. With C.C. Walker, "Genius," in Textbook in Abnormal Psychology, P. London and D. Rosenhan (eds.), Holt, Rinehart, and Winston, Inc., New York (1966).
75. With C.C. Walker, "On Temporal Characteristics of Behavior in Certain Complex Systems," Kybernetik, 3, (2), 100-108 (1966).
76. "The Place of the Brain in the Natural World," Currents in Modern Biology, 1, (2), 95-104 (1967).
77. "The Brain of Yesterday and Today," 1967 International Convention of IEEE, Part 9, pp. 30-33 (1967).
78. "Homeostasis," in Encyclopedia of Biochemistry, J.R. Williams and M. Lansford (eds.) Reinhold, New York (1967).
79. "Information Processing in Everyday Human Activity," BioScience, 18, (3), 190-192 (1968).
80. "Some Consequences of Bremermann's Limit for Information Processing Systems," in Cybernetic Problems in Bionics, H. Oestreicher and D. Moore (eds.), Gordon and Breach, New York, pp. 69-76 (1968).
81. "The Contribution of Information Theory to Pathological Mechanisms in Psychiatry," British J. of Psychiatry, 114, (517), 1485-1498 (1968).
82. "Principles of the Self-Organizing System," in Modern Systems Research for the Behavioral Scientist, W. Buckley (ed.), Aldine, Chicago, pp. 108-122 (1968).
83. "Regulation and Control," in Modern Systems Research for the Behavioral Scientist, W. Buckley (ed.), Aldine, Chicago, pp. 296-303 (1968).
84. "Measuring Memory," (in Russian) in Systems Organization of Physiological Functions, V.V. Parin (ed.), Moscow, pp. 239-243. (1969).
85. "Two Tables of Identities Governing Information Flows within Large Systems," Communications of the Amer. Soc. for Cybernetics, 1, (2), 3-8 (1969).
86. "Self-Regulation and Requisite Variety," in Systems Thinking, F. E. Emery (ed.), Penguin Books, Baltimore, pp. 105-124 (1969).
87. "Adaptation and the Multistable System," in Systems Thinking, F. E. Emery (ed.), Penguin Books, Baltimore, pp. 230-240 (1969).
88. "Variety, Constraint and the Law of Requisite Variety," in Modern Systems Research for the Behavioral Scientist, W. Buckley (ed.), Aldine, Chicago, pp. 129-136 (1969).
89. "Energy and Signal," Intern. J. Neuroscience, 1, 95-98 (1970).
90. With M.R. Gardner, "Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability," Nature, 228, 784 (1970).

91. With R. Conant, "Every Good Regulator of a System Must Be a Model of that System," Int. J. Systems Sci., 1, (2), 89-97 (1970).
92. "Information Flows Within Co-ordinated Systems," in Progress of Cybernetics: Proceedings of the International Congress of Cybernetics, London, 1969, J. Rose (ed.), Gordon & Breach Science Publishers Ltd., London, pp. 57-64 (1970).
93. With H. Tuttle and K. Kokjer, "A Table of the Entropies Associated with Simple Partitions," Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 127 pp. (1970).
94. "Chance Favors the Mind Prepared," Science, 168, 777 (1970).
95. "Analysis of the System to be Modeled," in The Process of Model-Building in the Behavioral Sciences, Ohio State University Press, pp. 94-114 (1970).
96. "Systems and Their Informational Measures," in Trends in General Systems Theory, G. Klir (ed.), Wiley-Interscience, New York, pp. 78-97 (1972).
97. With R.F. Madden: "The Identification of Many-Dimensional Relations", Int. J. Systems Science, 3, (4), 343-356 (1972).
98. "Setting Goals in Cybernetic Systems," in Cybernetics, Artificial Intelligence and Ecology, H.W. Robinson and D.E. Knight (eds.), Spartan Books, New York, pp. 33-44 (1972).
99. "Some Peculiarities of Complex Systems," Cybernetic Medicine, 9, (2), 1-8 (1973).
100. "Editorial," Behavioral Science, 18, (1), 1-6 (1973).
101. "Feedback," in Collier's Encyclopedia Yearbook.

REFERENCES TO PAPERS BY OTHERS
(cited in chapter introductions and in the Foreword)

102. Broekstra, G., "On the representation and identification of structure systems." Int. J. of Systems Science, Vol. 9, No. 11, pp. 1271-1293 (1978).
103. Cavallo, R. and G. J. Klir, "Reconstructability analysis of multidimensional relations; a theoretical basis for computer-aided determination of acceptable systems models." Int. J. of General Systems, Vol. 5, No. 3, pp. 143-171 (1979).
104. Conant, R. C., Structural modelling using a simple information measure. Int. J. Syst. Sci., Vol. 11, No. 6, pp. 721-730 (1980).
105. Conant, R., "How to ignore Bremerman's Limit." Proc. Southeastern Regional Meeting of Soc. for General Systems Research, Louisville, April 21-23 (1981).
106. Klir, G. J., "Identification of generative structures in empirical data." Int. J. of General Systems, Vol. 3, No. 2, pp. 89-104 (1976).
107. Klir, G. J., An Approach to General Systems Theory. Van Nostrand Reinhold, New York (1969).
108. Klir, G. J. (ed.), Applied General Systems Research. Plenum Press, New York (1978). Appendix B, pp. 985-988.
109. Klir, G. J., "General systems problem solving methodology." In: Methodology in Systems Modelling and Simulation, ed. by B. P. Zeigler et al., North-Holland, Amsterdam, pp. 3-28 (1979).
110. Klir, G. J. and M. Valach, Cybernetic Modelling. ILLIFFE books, London (1967).
111. Krippendorff, K., "On the identification of structures in multivariate data by the spectral analysis of relations." 23rd annual meeting of the Society for General Systems Research, Houston, Texas, January 3-8 (1979).
112. Krippendorff, K., "Q: an interpretation of the information theoretical Q-measure," Proceedings of the 5th Int. Meeting on Cybernetics and Systems Research, Vienna, April (1980).
113. Porter, B., "Requisite variety in the systems and control sciences." Int. J. of General Systems, Vol. 2, No. 4, pp. 225-229 (1976).
114. Simon, H., "The architecture of complexity." Proc. Amer. Phil. Soc., Vol. 106, pp. 467-482 (1962).
115. Zeigler, B. P., "A conceptual basis for modelling and simulation." Int. J. of General Systems, Vol. 1, No. 4, pp. 213-228 (1974).
116. Zeigler, B. P., "The hierarchy of system specifications and the problem of structural inference." In: F. Suppe and P. D. Asquith (eds.), PSA 1976, Philosophy of Science Assoc., East Lansing, Mich., pp. 227-239 (1976).
117. Special Issue of Reconstructability Analysis, Int. J. of General Systems, Vol. 7, No. 1 (1981).

INDEX

INDEX