3

Organisms as Causal Systems Which Are Not Mechanisms: An Essay into the Nature of Complexity

ROBERT ROSEN

Department of Physiology and Biophysics
Dalhousie University
Halifax, Nova Scotia

I. INTRODUCTION

The thrust of this essay is that the theory of organisms, and of what we shall call *complex systems* in general, requires a circle of ideas and methods that, from the very outset, depart radically from those taken as axiomatic for the past 300 years.

What we shall conclude can be stated succinctly here at the outset, as follows.

1. Our current systems theory, including all that presently constitutes physics or physical science, deals exclusively with a very special class of systems that I shall call *simple systems* or *mechanisms*.

ISBN 0-12-597280-6

2. Organisms, and many other kinds of material systems, are not mechanisms in this sense. Rather, they belong to a different (and much larger) class of systems, which we shall call *complex.*

3. Thus the relation between contemporary physics and biology is not, as everyone routinely supposes, that of general to particular.

4. To describe complex systems in general, and organisms *a fortiori*, an entirely novel kind of mathematical language is necessary.

5. A simple system can only *approximate* to a complex one, locally and temporarily, just as, e.g., a tangent plane can only approximate to a nonplanar surface locally and temporarily. Thus in a certain sense, a complex system can be regarded as a kind of global limit of its approximating simple subsystems.

6. Complex systems, unlike simple ones, admit a category of final causation, or anticipation, in a perfectly rigorous and nonmystical way.

We shall find many other novel consequences of complexity as we proceed; especially to be mentioned are deeper insights into the nature of "information" and the relation of theory to "experiment."

We thus argue, in effect, that any attempt to deal effectively with the material basis of biological organization forces a revamping of our entire traditional scientific epistemology. As we shall see, this revamping entails a number of dramatic consequences, not only for biology, but also for physics, and indirectly, even mathematics.

II. BIOLOGY AND OTHER SCIENCES

Biology is the linchpin of the sciences. Insofar as organisms are material systems, their remarkable properties stand as a challenge, a reproach, and an inspiration to the sciences of matter (physics and chemistry). Insofar as organisms adapt, perceive, and behave, their properties impinge on our technologies to an ever-increasing extent. And insofar as organisms evolve, develop, constitute communities, and form societies, biology also provides both the material substratum and the metaphorical inspiration for all the social, political, and economic sciences. Thus nothing that happens in any other science can be immaterial to biology; conversely, anything that happens in biology ultimately radiates into every corner of scientific thought.

At root, theoretical biology is concerned with only one great question: What is it about certain material systems that confers upon them the characteristics of life, which makes them living beings? All other problems of biology, both theoretical and "practical," are collateral or subordinate to this central question.

It is a significant fact that, despite generations of trying, there is as yet no list of tests, characteristics, or criteria we can apply to a given material system that can decide whether that system is an organism or not. Stated another way, the decision as to whether a given system is an organism is entirely a subjective, intuitive one, based on criteria that have so far resisted formalization, or even articulation. Thus from a strictly rigorous point of view, the subject matter of the science of biology is *undefined*; it is based entirely on an informal consensus essentially akin to pattern recognition, but that consensus is one we all share to a startling degree.

The problem of "What is life?", to use Schrödinger's phrase, first became acute with the triumph of the Newtonian revolution (about which we shall have much more to say subsequently), for one of the major consequences of the Newtonian world view was the obliteration of any distinction between the organic and inorganic. In this view, every material system could be analyzed down to a population of structureless particles moving in fields of force; in principle, the dynamical equations governing any such population could be written down and solved. Among other things, this picture has become the canonical one for scientific *explanation*; therefore the revolutionary developments in physics that created the problem of "what is life," by obliterating any distinction between living and nonliving, also proclaim themselves the only place to look in trying to solve the problem.

Indeed, insofar as contemporary physics claims to deal with material reality in all its manifestations and insofar as organisms are material systems, it is most natural to seek insight into organic phenomena via *biophysics*. Therefore we shall briefly review the present status of this endeavor.

First, we must remark that, historically, the relation between theoretical physics and biology has never been close. None of the great names of physical science, from Newton to the present, have known or cared much about the properties of organisms, and therefore organic phenomena played no essential part in their science (leaving aside such diversions as Maxwell's Demon, Schrödinger's informal essays, and the like). For the past 300 years, the theoretical physicist has exclusively concerned himself with the formulation of *universal laws.* From his perspective biology is concerned with a small class of very special (indeed, *inordinately* special) systems, clearly not the place to look when seeking universal laws. In this light, what makes organisms so special is not their immunity from universal laws of physics, but the apparent multitude of initial and boundary conditions that must be determined in order to bring the laws to bear upon them. The specification of these conditions is regarded by the physicist as an empirical job, and someone else's job at that. In a word, the physicist has always believed that there is no new physics to be learned from organisms; he has never doubted that he deals in the general and that biology concerns only the particular.

The modern biologist has all too avidly embraced this perspective. In a historical sense, the past 50 years have seen biology finally catch up with the

Newtonian revolution that swept over the rest of the scientific world in the eighteenth century. The three-century lag arose because biology had no analogue of the solar system; no way to make immediate contact with the Newtonian ideas. Not until physics and chemistry had elaborated the technical means to isolate and manipulate minute quantities of matter (including organic matter) in the 1930s could one think of a particulate, mechanical basis for biology, i.e., of a *molecular* biology.

The preceding considerations do not represent solely my own subjective assessment. It is worth digressing for a moment to address this further, because the current state of biological science is not generally seen as the culmination of historical trends going back to Newtonian mechanics, or even further back to Descartes. Rather, it is regarded as objective and axiomatic, even though, as we shall see, its diverse historical roots imbue it with obvious paradoxical properties.

In 1970 there appeared a volume entitled "Biology and the Future of Man," edited by Philip Handler, then president of the National Academy of Sciences of the U.S.A. The book went to extraordinary lengths to assure the reader that it spoke for biology as a science, that in it biologists spoke with essentially one voice. For instance, it was emphasized that the volume was prepared not as a mere academic exercise, but for serious pragmatic purposes:

> Some years ago, the Committee on Science and Public Policy of the National Academy of Sciences embarked on a series of "surveys" of the scientific disciplines. Each survey was to commence with an appraisal of the "state of the art".... In addition, the survey was to assess the nature and strength of our national apparatus for continuing attack on those major problems, e.g., the numbers and types of laboratories, the number of scientists in the field, the number of students, the funds available and their sources, and the major equipment being utilized. Finally, each survey was to undertake a projection of future needs for the national support of the discipline in question to assure that our national effort in this regard is optimally productive.... [p. v.]

To address such serious matters, we are then told that the academy proceeded as follows:

> Panels of distinguished scientists were assigned subjects.... Each panel was given a general charge... as follows:
>
> The prime task of each Panel is to provide a pithy summary of the status of the specific sub-field of science which has been assigned. This should be a clear statement of the prime scientific problems and the major questions currently confronting investigators in the field. Included should be an indication of the manner in which these problems are being attacked and how these approaches may change within the foreseeable future. What trends can be visualized for tomorrow? What lines of investigation are likely to subside? Which may be expected to advance and assume greater importance?... Are the questions themselves... likely to change significantly?... Having stated the major questions and problems, how close are we to the answers? The sum of these discussions, panel by panel, should constitute the equivalent of a complete overview of the highlights of current understanding of the Life Sciences. [p. vi.]

There were 21 such panels established, spanning the complete gamut of biological sciences and the biotechnologies. The recruitment for these panels consisted of well over 100 eminent and influential biologists, mostly members of the academy. How the panelists themselves were chosen is not indicated, but there is no doubt that they constituted an authoritative group.

In due course, the panels presented their reports. How they were dealt with is described in vivid terms:

In a gruelling one week session of the Survey Committee... each report was *mercilessly* exposed to the criticism of all the other members.... Each report was then rewritten and subjected to the *searching*, sometimes *scathing*, criticisms of the members of the parent Committee on Science and Public Policy. The reports were again revised in the light of this exercise. Finally, the Chairman of the Survey Committee... devoted the summer of 1968 to the final editing and revising of the final work. [p. vii.]

Thus we have good grounds for regarding the contents of this volume as constituting a true authoritative consensus, at least as of 1970. There are no minority reports; no demurrals; biology does indeed seem guaranteed here to speak with one voice.

What does that voice say? Here are a few characteristic excerpts:

The theme of this presentation is that life can be understood in terms of the laws that govern and the phenomena that characterize the inanimate, physical universe and, indeed, that at its essence life can be understood *only* [emphasis added] in the language of chemistry. [p. 3.]

A little further along, we find this:

Until the laws of physics and chemistry had been elucidated, *it was not possible even to formulate* [emphasis added] the important, penetrating questions concerning the nature of life.... The endeavors of thousands of life scientists... have gone far to document the thesis... [that] living phenomena are indeed intelligible in physical terms. And although much remains to be learned and understood, and the details of many processes remain elusive, those engaged in such studies hold *no doubt* [emphasis added] that answers will be forthcoming in the reasonably near future. Indeed, *only two major questions* [emphasis added] remain enshrouded in a cloak of *not quite* [emphasis added] fathomable mystery: (1) the origin of life... and (2) the mind-body problem... yet (the extent to which biology is understood) even now constitutes a satisfying and exciting tale. [p. 3.]

Still further along, we find things like this:

While *glorying* [emphasis added] in how far we have come, these chapters also reveal how large is the task that lies ahead.... If [molecular biology] is exploited with vigor and understanding... a shining, hopeful future lies ahead. [p. 6.]

And this:

Molecular biology provides the closest insight man has yet obtained of the nature of life—and therefore, of himself. [p. 64.]

And this:

> It will be evident that the huge intellectual triumph of the past decade will, in all likelihood, be surpassed tomorrow—and to the everlasting benefit of mankind. [p. 130.]

It is clear from such rhapsodies that the consensus reported in this volume is not only or even mainly a scientific one; it is an emotional and aesthetic one. Indeed, anyone familiar with the writings of Newton's contemporaries and successors will recognize them.

The volume to which we have alluded was published in 1970. But it is most significant that nothing fundamental has changed since then.

Despite this overwhelming commitment to the mechanical, there is at the same time another thread running through contemporary biology, one quite incompatible with mechanics. A good statement of this was given by Jacques Monod (1971):

> We can assert today that a universal theory, however completely successful in other domains, could never encompass the biosphere, its structure and its evolution as phenomena deducible from first principles.... The thesis I shall present... is that the biosphere does not contain a predictable class of objects or events but constitutes a particular occurrence, compatible with first principles but not deducible from these principles, and therefore *essentially unpredictable* [emphasis added]. [p. 42.]

In other words, the important features of organisms are the result of accidents, and hence governed by no laws at all. Biology thereby becomes a branch of history, not of science. All that can be said for this peculiar combination of mechanism and historical accident is that it allows the modern biologist the luxury of enjoying the fruits of both mechanism and vitalism with an unsullied conscience.

As a matter of fact, there is not a shred of evidence supporting any of this received picture. Quite the contrary; as we proceed, we shall see much that is totally incompatible with it. The reader may take my word for it that a great deal of additional contrary evidence could be adduced with ease. Such incompatibilities are simply ignored by most biologists, and by most physicists as well, on the grounds that merely the acquisition of more information, more data, will somehow resolve them. In their view, what fault there is is not a fault of principle but of practice.

Indeed, it is certainly true that if the fault is one of principle, then the fault is something inherent in the mechanical paradigm itself, something present from the outset. To find it, we need to go back to the very beginnings, to the basic epistemological presuppositions that have become so familiar and axiomatic that to challenge them seems today tantamount to challenging science itself. However, this is exactly what we shall do. To make it clear that we are not operating in a vacuum, we shall preface the more specific analysis with a concrete example of an approach that takes us immediately out of the Newtonian, mechanical context. This is the approach to organisms first systematically developed by Nicolas Rashevsky in 1954 and termed by him *relational biology.*

The upshot of these developments, to be reported later, is thus very much in the spirit of a pregnant remark that Einstein is supposed to have made to Leo Szilard: "One can best appreciate, from a study of living things, how primitive physics still is."

III. RELATIONAL BIOLOGY

As noted earlier, the term *relational biology* was coined by Nicolas Rashevsky in 1954 to distinguish it from more familiar approaches, which he characterized as *metric biology*; Rashevsky, who had himself been trained as a theoretical physicist, was the great pioneer in the application of quantitative physical ideas to the elucidation of the material basis of organic behavior. He almost singlehandedly created the field he called "mathematical biophysics," which, in his words, would stand in the same relation to experimental biology as mathematical physics stands to experimental physics.

In modern terms, Rashevsky's initial approach was mechanist and reductionist. His earliest work involved those areas in which physical processes and biological behavior naturally seemed to intersect—the physical basis of cell division (cytokinesis) and reaction–diffusion processes generally (this was in 1930); nerve conduction and nerve excitation, leading to the first network theories of the central nervous system and the brain; the control, form, and dynamics of the cardiovascular system, and many others.

By 1950 the feasibility and fecundity of mathematical modelling as a probe of biological phenomena was well established, owing largely to Rashevsky's own work and that of his school. But Rashevsky himself was beginning to feel dissatisfied. The source of this dissatisfaction lay in the fact that, in the process of modelling its individual features, the organisms *qua* organism seemed to have disappeared. He began to wonder where and why it had disappeared, and how it could be retrieved. Thus in a spontaneous way he began to struggle with his own reductionism.

The answer he came up with, in his pioneering paper of 1954 ("Topology and Life"), was roughly as follows. Heretofore, we have supposed that we can resolve an organism into *physical* subsystems, understand each of these in detail through traditional modes of physical and mathematical investigation, and when we are done, the original biological organization to which these material subsystems belonged will reemerge *of itself*, as a *consequence* of the nature of these subsystems separately. This is still, of course, the reductionistic credo, most firmly ensconced in the field of molecular biology, which hardly existed in Rashevsky's time. It is a modern version of the idea that any *mixture* (i.e., a system of many phases) must be resolved into its constituent *pure phases* and that the properties of the original mixture may then be *inferred* from those of its constituent pure phases, as logical consequences thereof.

However, as Rashevsky realized better than anyone else, this did not seem to be happening; quite the contrary. Moreover, so many utterly diverse kinds of *physical* systems could be organisms that it became a really serious problem to try to understand, on a *physical* level, how all of them in their physical diversity could manifest those commonalities of behavior and organization that make us recognize them as being alive. That is, Rashevsky grasped that what he called the "perceived unity of the organic world" could not, either conceptually or technically, be approached directly through the "metric" approaches he (and everyone else) was then using.

He thus proposed something extremely radical. In effect, he said something like this: We are really interested in the organizational features common to all living systems; and in their material structures only insofar as they support or manifest those features. Therefore, we have heretofore approached organisms in precisely the wrong way; we have abstracted out, or thrown away, all those global organizational features in which we are really interested, leaving ourselves with a purely material system that we have studied by purely material methods, hoping ultimately to recapture the organization from our material studies. This has not happened. Why do we not then *start* with the organization? Why do we not, in effect, *abstract away the physics and the chemistry*, leaving us with a pure organization, which we can formalize and study in completely general abstract terms; and recapture the *physics* later through a process of *realization*? It was basically this endeavor that he called relational biology.

His initial approach was a simple one; he represented biological organization through directed graphs, whose nodes were "biological functions" and whose directed edges were relations of temporal or logical precedence. His notion of an "abstract biology" was a class of such graphs, all canonically generated from an initial "primordial graph," and hence all canonically related through graph morphisms of special types. These "abstract biologies" had many striking properties; e.g., from suitable knowledge of any graph in such an abstract biology, one could reconstruct the primordial and hence the whole biology.

In his relational biology, then, Rashevsky anticipated many more recent mathematical ideas. However, the time was quite wrong for his new relational ideas to find any acceptance anywhere. In biology, the "golden age" of molecular biology was just beginning; experimentalists had no time or use for anything of this kind. Those who considered themselves theorists either were preoccupied with the reductionistic modelling that Rashevsky had earlier taught them or were bemused by seductive ideas of "information theory," games theory, cybernetics, and the like, regarded Rashevsky and his ideas as generally archaic because he did not take direct cognizance of their enthusiasms.

Thus his ideas about relational biology ended in Limbo, unread or forgotten. However, what he had done was, in effect, to propose a whole new way of representing material reality; a way that, when pursued, pointed to glaring defects and omissions in the older, received ways. We shall now see how this comes about, in a rather different and simpler relational context than Rashevsky's original one.

IV. THE (M, R)-SYSTEMS

The (M, R)-systems comprise a class of relational cell models, first proposed by me in 1958. As with any relational approach, the problem is to try to characterize at least some fundamental organizational feature common to all cells, independent of their specific physicochemical structures. In the (M, R)-systems, the organizational feature taken as central is the distinction between "cytoplasm" and "nucleus." In traditional cytology, the cytoplasm is regarded as the seat of metabolic activity, the province of enzymes that process environmentally derived materials and convert them into metabolically important new forms. The nucleus, on the other hand, is the seat of the genome and in particular plays the guiding role in the synthesis of the metabolic machinery that sits in the cytoplasm. The (M, R)-systems were invented as the simplest and most general class of mathematical systems that embodies this kind of organization. And whatever else a real cell may be, it must also be (or better, must also realize) an (M, R)-system.

This organization may be seen in the simplest imaginable (M, R)-system, which can be expressed as

$$A \xrightarrow{f} B \xrightarrow{\Phi} H(A, B) \qquad (1)$$

Here f is simply a mapping from a set A of *environmental inputs* to a new set B of *environment outputs*. As noted earlier, this is the abstract equivalent of "cytoplasm." The mapping

$$\Phi: B \to H(A, B)$$

is the abstract counterpart of "nucleus"; in effect, it takes the outputs of metabolic activity in the cytoplasm and converts it into more metabolic machinery. The range of Φ is thus a set of *mappings*, as indicated.

Mathematically, then, an (M, R)-system is simply a collection of mappings, together with their domains and codomains. The theory of such systems is a part of category theory, and I believe that the (M, R)-systems provided the first indication of the deep role that category theory could play in theoretical science. Indeed, the (M, R)-systems themselves form a category whenever the category from which their sets and mappings are drawn is specified.

The (M, R)-systems, though simple in concept, possess a remarkably rich theory. Let us indicate one important property. As matters stand, we have embodied in the (M, R)-system only one of the characteristics traditionally associated with the cell nucleus; namely, its role in synthesis of the cellular metabolic machinery. But any nucleus worthy of the name has one other decisive property; it *replicates*. In all other cell models known to me, replication must be superimposed as an additional *ad hoc* property. But in (M, R)-systems, there are replication mechanisms inherent in the organizational features already represented; requiring only a further mathematical condition to be satisfied and no further *ad hoc* hypotheses.

Let us consider the simplest situation. Quite generally, let X, Y be any sets and let $H(X, Y)$ be a set of mappings from X to Y. It is well known that every element $x \in x$ induces a mapping

$$\hat{x}: H(X, Y) \rightarrow Y$$

by writing

$$\hat{x}(f) = f(x).$$

This map $\hat{x}$ is often called an evaluation map; such maps are familiar, for instance, in linear algebra, where they identify a vector space with its second dual space. In particular, let us put

$$X = B, \qquad Y = H(A, B).$$

Then we can regard each $b \in B$ as inducing a map

$$\hat{b}: H(B, H(A, B)) \rightarrow H(A, B).$$

If this map $\hat{b}$ is *invertible*, then its inverse is a map

$$\hat{b}^{-1}: H(A, B) \rightarrow H(B, H(A, B)).$$

Now look at the original (M, R)-system. The repair map Φ is obviously an element in $H(B, H(A, B))$. Thus if $a \in A$, $b = f(a)$, and $\hat{b}^{-1}$ exists, we can extend the original (M, R)-system as follows:

$$A \xrightarrow{f} B \xrightarrow{\Phi} H(A, B) \xrightarrow{\hat{b}^{-1}} H(B, H(A, B)). \tag{2}$$

The new map added to the right-hand side is a *replication map*, and by its very definition, it clearly replicates the "nuclear" mapping Φ.

However, there is a condition to be satisfied, namely, that the evaluation map $\hat{b}$ be invertible. This means

$$\hat{b}(\Phi_1) = \hat{b}(\Phi_2) \qquad \text{implies} \qquad \Phi_1 = \Phi_2$$

or

$$\Phi_1(b) = \Phi_2(b) \qquad \text{implies} \qquad \Phi_1 = \Phi_2.$$

This latter is a condition on the original *category*; it is reminiscent of a unique-trajectory property, or biologically, of a one-gene, one-enzyme hypothesis.

It should also be noted that the final two mappings in Eq. (2) themselves consititute an (M, R)-system, in which Φ, the "nuclear" map for the original system, is now a "cytoplasmic" map, and $\hat{b}^{-1}$, the replication map for the original system, is now the "nuclear" map for the new one. Thus *mathematically* there is no intrinsic difference between these processes; any map in our category could be realized either as a metabolic map, a repair map, or a replication map, depending entirely upon the *context*.

Thus from just this quick overview we can see that the (M, R)-systems possess novel characteristics, bearing on biology in a unique way. But as always when one attempts to do theory, one confronts a banal but unavoidable question: *Is it testable*, and if so, *how*? We have been brought up to believe that a theory that is not testable (i.e., falsifiable) is worthless. And indeed, it is also considered part of the *theorist's* job to make theory testable in this sense, in effect to construct some kind of experimental protocol for this purpose, even if only in principle.

It was this exigency that first forced me into epistemology, for I was convinced that a relational description of an organism is as valid, as *physical*, a description as any conventional physicochemical one. But, like the wave functions of quantum mechanics, it is a description pertaining to a *class* of physically diverse (though functionally equivalent) systems; and as long as *experimental test* exclusively means verifying some kind of specific physicochemical operation on individual systems in such a class, there was in principle no way that the relational descriptions could in fact be tested in the conventional sense. For as we have noted, it is precisely such physicochemical particulars that are abstracted away in the process of generating the relational model.

It should be noted that relational approaches such as the one we are discussing do allow us, even as they stand, to make assertions about the physicochemical nature of biological organizations, but these assertions are all negative ones (i.e., about what cannot happen) and thus are considered empirically unacceptable. For instance, I was able to show that any attempt to add a new component (either metabolic or repair) to a given (M, R)-system would generally destroy the overall pattern of cellular organization, that is, would not result in a new (M, R)-system. In fact, to construct a new (M, R)-system that contains a given one and a new component is not an easy matter. The situation is somewhat analogous to attempting to add a new instruction to a computer program; this generally requires a whole subroutine, which may be much larger than the original program was. Such considerations turn out to have obvious implications for genetic engineering, whose practitioners are presently finding out about such limitations the hard way.

In any case, the most obvious way of making contact with the conventional universe of physicochemical descriptions and hence of generating predictions testable by conventional physicochemical experimental techniques (which, as we have noted, is what *testable* means) is through the process of *realization* referred to earlier. For such a realization must, on the one hand, have the relational features of an (M, R)-system and, on the other hand, be a conventional description of a specific physicochemical system and thus be amenable to traditional notions of testability. In particular, we would seek realizations for which *replication maps are also realizable*, for it is in connection with these that the most novel and interesting predictions can be made. In fact, the successful realization of such an (M, R)-system is tantamount to the synthesis of a novel, autonomous life form.

The strategy to be followed in physically realizing an abstract organizational structure like an (M, R)-system seemed at first to me not too different from that followed by an engineer in designing a real physical structure to meet some given initial set of functional specifications. For here, too, we must reach into a class of physically diverse but functionally similar systems and pick one out. The usual criterion for this selection purpose is one of *optimality* (e.g., least cost). Indeed, I might assert that optimality is the *canonical* way of selecting individual elements from equivalence classes; one may think even of such things as the Jordan canonical form of ordinary matrices (in which the number of 0 entries is maximized).

But the problem did not turn out to be that straightforward after all. The process of determining why it was not is in fact the main purpose of this discussion, and it is this to which we now turn.

V. A FIRST ATTEMPT AT REALIZATION OF (M, R)-SYSTEMS

Let us begin by reviewing an early attempt of mine (1964) to solve the realization problem. It seemed to me that a first step would be to transform *mathematically* the (M, R)-systems to a form in which the various sets and mappings of the (M, R)-system could be interpreted in terms of the states of some system and a set of dynamical laws could be superimposed thereon. This was at least the conventional language in which physical systems were universally described; hence realizing this kind of mathematical object would be much easier than realizing an (M, R)-system directly.

The first idea that came to mind was the language of sequential machines, or finite automata. This is in effect the language of classical dynamical system theory (or better, of control theory) paraphrased to the constraints of discrete time and discrete states. Furthermore, it is closely associated with certain kinds of material systems that are important in biology and elsewhere; in biology, in the representation of neural nets (brains) and genetic control in cells (operon networks); in the general theory of digital computing devices; and elsewhere, including the basic theory of mathematics itself.

In its most general formal terms, a sequential machine is a 5-tuple $(A, B, S, \delta, \lambda)$ consisting of three (finite) sets and two mappings:

1. an *input set* or *input alphabet* A;
2. an *output set* or *output alphabet* B;
3. a set of (internal) *states* S;
4. a *next-state* map, $\delta: S \times A \to S$;
5. an *output map*, $\lambda: S \times A \to B$.

The "sequential" aspect comes from iterations of the next-state map and output map, through which we allow the machine to operate on *strings* $\omega = a_{i_1}, \ldots, a_{i_n}$ of elements of A (i.e., elements of the free monoid $A^{\#}$ generated by A) instead of on A itself; the kth element a_{i_k} of a word $\omega \in A^{\#}$ is then thought of as the one presented to the machine at the kth *instant*; and the indices themselves thus pertain to the instants of a discrete time. It is well known that such devices can themselves always be *realized* by neural nets in many ways.

Now an abstract (M, R)-system can itself be realized by something like a sequential machine. In fact, if we consider the (M, R)-system of Eq. (1), we can define

1. input set $= A$;
2. output set $= B$;
3. state set $S = H(A, B)$;
4. next-state map $\delta: A \times S \to S$; defined by $\delta(a, f) \equiv \Phi(f(a))$;
5. output map $\lambda: A \times S \to B$; defined by $\lambda(a, f) \equiv f(a)$.

At first, this looked extremely promising. Biologically, there were a host of network realizations now available (e.g., operon networks). Mathematically, there were a number of possibilities for passing from discrete to continuous time, i.e., to true dynamical and control systems, and thence to explicit "hardware" realizations, which would comprise "cells" of perhaps utterly novel kinds. In these various realizations, one could explicitly seek those situations in which replication mappings were realizable, for these were of the greatest interest.

There were indeed many interesting conclusions that could be drawn from just these possibilities. But the really fundamental problems remained refractory to this whole approach. In a nutshell, the reason lay in the mathematical dichotomy between *set* (object) and *mapping* in the (M, R)-system. In a network realization, a "state" of the network is a pattern of activation in the elements that constitute the network, while the "next-state mapping" is embodied in the wiring diagram of the network. But intuitively, in the (M, R)-system, both the metabolic map(s) f and the nuclear or repair maps Φ should themselves be embodied in (or realized by) physical structures, and their mapping properties should be a consequence of these structures. When we realize $\Phi(f(a))$, for example, this is abstractly a *mapping* ($f: A \to B$) in the (M, R)-system; it is a pattern of excitation (i.e., a single state) in a network; but it should be a material structure in the kind of realization we are actually seeking. Even more, the map Φ itself in the (M, R)-system is a wiring diagram in a network realization, a pattern of specificities in an operon network, but, in fact, it should be realized itself as a material *structure*, from which all these mapping properties should follow.

These considerations led to a fundamental rethinking of the whole idea of how to go about realizing any kind of abstract relational description of a material system, and thence to the whole problem of trying to invert the process by which any kind of mathematical description, or model, of a material system is obtained in the first place. In particular, the very fact that the same mathematical formalism (e.g., a network) could be *interpreted* in so many disparate physical ways ultimately led me to suspect that something crucial might be missing *from the mathematics itself*. In other words, I began to entertain the possibility that our conventional mathematical descriptions of physical reality, which have essentially gone unquestioned for three centuries, might themselves be fundamentally deficient, that it was this deficiency that was responsible for the problems posed by an attempt to realize physically an abstract functional organization.

Let us then briefly review what is involved in the way we conventionally build mathematical or theoretical pictures of the material world. As we shall soon see, the procedure involves *several* tacit hypotheses about the material world, which show up most clearly when we attempt to invert the procedure and *realize* a description. These hypotheses may not be true; I would now argue that they are not true in general. Removing them leads us, in fact, to a whole new epistemology, with all the attendant implications of that fact. We will describe all this in the subsequent sections, beginning with a brief description of the nature of our current ideas regarding the imaging of material reality and some of their basic historical roots.

VI. THE MODELLING RELATION

All of our thought processes, from the most mundane aspects of daily life to the deepest reflections of theoretical science, are based on two parallel postulates.

First is the belief that the sequence of events that we perceive in the external world is not entirely whimsical, chaotic, and arbitrary but obeys definite laws or relations. The relations that exist between events in the external world, and that govern their succession, collectively constitute what we call *causality*. Without a belief in causal order, there could be no science and, very probably, no sanity.

But a belief in a causal order relating events in the external world is only one part of the story. The other part is an independent belief that this causal order relating events can be, at least in part, grasped and articulated by the mind. It is a belief that, in some deep sense, the causal order relating events can be mirrored in a corresponding relation between propositions that describe these events. Now such propositions belong to an internal, symbolic, linguistic world and hence cannot themselves be related by any kind of "causality." But there is another kind of order through which propositions can be related, and that is a *logical* order or *implication.*

Thus we must also believe that the causal order, relating events in the external world, can be brought into congruence with a logical or implicative order in some appropriate logical, symbolic world of propositions describing these events. When such a congruence is established, *implications* in the logical system become *predictions* about the causal order.

These two beliefs together constitute the idea of natural law. It is the entire task of theoretical science to establish the congruence between causal order in the external world and implicative order in the formal world, which embodies the very idea of natural law.

The preceding remarks can be summed up succinctly in a diagram:

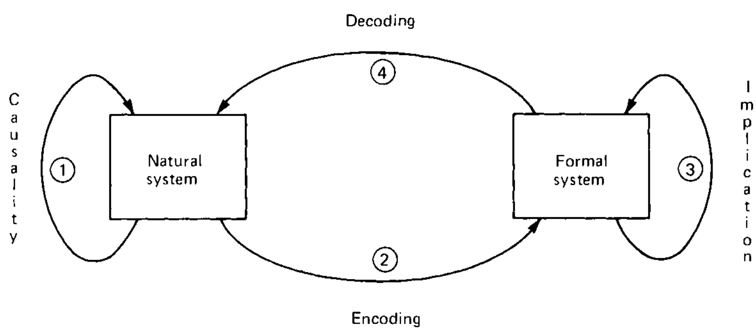


FIG. 1

(cf. Chapter 2). In this diagram the left-hand box represents the external world or some fraction thereof. The sequence of events it exhibits is governed by causal relations, as represented by arrow 1. On the right-hand side sits some formal system, whose elements are governed by relations of implication or logic (arrow 3). We may establish relations between these two diverse worlds by means of "encodings" (arrow 2), whereby attributes of the external world are identified with, or named by, corresponding elements in the formal system; and by "decodings" (arrow 4), whereby elements of the formal system are treated as names of, or symbols for, attributes of events.

We shall say that a *modelling relation* has been established between the natural system and the formal system when the preceding diagram of arrows *commutes*; i.e., when

$$\text{arrow } 1 = \text{arrows } 2 + 3 + 4.$$

In this case, one *always* obtains the same answer; whether one simply sits as an observer and watches the causal order unfold in the natural system (arrow 1) or whether one encodes attributes of that system as propositions (i.e., initial conditions or hypotheses) in the formal system (arrow 2), generates new propositions from these through the inferential structure (arrow 3), and then decodes these back into assertions or predictions (arrow 4).

If a modelling relation exists between a natural system N and a formal system F, we may call F a *model* of N, or N a *realization* of F.

Since mathematics, in the broadest sense, is the study of implication relations in formal systems, or the art of extracting conclusions from premises, it follows that mathematics is integrally involved in the study of natural law. Indeed, many of the deepest questions of theoretical science are concerned with specifying the kinds of formal or mathematical systems that can sit on the right-hand side of the preceding diagram and the kinds of mathematical relations that can exist between them. For instance, the whole problem of *reductionism* involves nothing else.

The equally important inverse or dual problem [i.e. given a formal system on the right-hand side of such a diagram, to determine the class of *natural* systems that can realize it, and the relations that exist between them (variously called *analogy*, *similarity*, or *scaling* relations)] has received no study commensurate with its importance.

We shall be concerned with *both* the class of formal systems that can be put into a modelling relation with a given natural system and with the class of natural systems that can realize a formal system (model). Clearly, these two classes are closely related and should be considered together. For the remainder of this work, then, we shall consider some general aspects of the modelling relation and its ramifications, with a particular eye on the problem with which we started; namely, the construction of realizations of relational models like the (M, R)-systems. Indeed, in the light of the discussion we have just given, encapsulated in the diagram of Fig. 1, we can see clearly how the very idea of a relational model requires *both* (1) a new look at the mathematical structures that can model a natural system (in our case, a "cell") and the mathematical relations between these models and (2) the class of natural systems that realize one (or more) formal structures.

In the process, we will see what radical profundities Rashevsky really unleashed with his apparently innocent ideas about a "relational biology."

VII. THE NEWTONIAN PARADIGM

In this section we shall briefly review the salient features of the class of mathematical or formal systems that are now accepted as models of material reality (what we have called "natural systems"). That is, these are the formal systems that can sit on the right-hand side of a commutative diagram such as Fig. 1. As we shall see, our basic ideas on this subject go back, in one way or another, essentially unchanged, to the mechanics of Newton's *Principia.* Despite enormous *technical* variations in mathematical language (e.g., from classical to relativistic or quantum; from continuous time to discrete time; from continuous state to discrete state; from deterministic to stochastic; from autonomous to forced; from finite-dimensional to infinite-dimensional; etc.), the basic epistemological presupposition remains the same, untouched and all but unnoticed.

This basic presupposition, as we shall see, is that systems have *states* and that upon these states some kinds of *dynamical laws*, or *equations of motion*, are superimposed. The states represent in a sense what is intrinsic, while the dynamical laws reflect the nature of the impinging environment in acting on what is intrinsic. Thus the dichotomy between states and dynamical laws embodies a distinction between *system* and *environment*. Also, in a formal way, the dualism of states and dynamical laws exactly parallels the purely *mathematical* dichotomy between propositions and inferential laws, or production rules, which is nowadays considered as the anatomical foundation of any mathematical formalism whatsoever.

But this basic presupposition, so familiar and axiomatic to us all, involves tacit hypotheses, not just one but several, about the natural world and its mathematical images. We will examine these in detail in the following sections. For the moment, simply review some of the salient formal and historical roots of what we shall call the *Newtonian paradigm.*

First, it must be recognized that Newton's *Principia* was certainly one of the most influential works of human history. In its own time, it was regarded as the capstone of the Renaissance; the culmination of the rational mind and its power to grasp natural law, a fountainhead of optimism and enlightenment.

In subsequent times, it unleashed successive waves of scientific advance that are still going on. It has set the standards for scientific investigation and for scientific explanation to such a degree that alternatives have become essentially unthinkable. This is equally true for modern developments, like quantum theory, which have transcended some aspects of Newton's original formalism, but still, as we shall see, subscribe to exactly the same epistemological presuppositions.

The influence of the *Principia* has radiated in two distinct directions: a reductionistic direction and a paradigmatic direction. We will consider them in turn.

From a reductionistic point of view, we consider the thrust of the *Principia* as the dynamics of systems of mass points. The Newtonian particles are devoid of internal structure; their only attributes are constitutive parameters like mass, which are time independent, and position, which is time dependent. Indeed, the *only* temporally variable attributes of a Newtonian particle are its position and the temporal derivatives of position. Thus the basic problem of Newtonian mechanics (and indeed the *only* problem associated with systems of Newtonian particles) is to tell where the constituent particles are located at any given instant, i.e., to specify configuration as a function of time.

Newton recognized that the arbitrary specification of configuration at an instant placed no restriction on the velocities of the constituent particles; i.e., both configuration and the first temporal derivative of configuration could be chosen completely arbitrarily. One might think that the same would be true for second time derivatives of configuration (i.e., acceleration) and for all higher time derivatives. But here Newton interposed his deep insight. He said, in effect, that the rest of the world (i.e., the *environment* of our system of mass points) exerts *forces* on the particles. *What* these forces are, intrinsically, cannot be (and need not be) specified, but the *effect* of these forces is to determine the acceleration of the particles of our system. That is, insofar as the "force" experienced by a particle is determined by where it is and how fast it is going, the *acceleration* of the particle, and hence all higher temporal derivatives of configuration, are then completely determined by configuration, rate of change of configuration, and the "forces" then imposed by the rest of the world. *Mathematically*, this amounts to expressing acceleration *recursively* as a function of the *lower* temporal derivatives. This expression is the *dynamical law* governing the system; by a mathematical process of *integration* we can convert this dynamical law into a relation giving *configuration* as a function of *time*.

This beautiful conception has a number of sweeping implications:

(1) Insofar as *any* material system can be resolved into a system of structureless particles, Newtonian mechanics seems to provide a recipe, or algorithm, for the study, modelling, and representation of *any* system. That is,

it in principle enables us to construct the diagram of Fig. 1, for *any* material system on the left-hand side. It establishes simultaneously the nature of the formal, mathematical image *and* the encoding and decoding arrows that convert the formal system into a model.

(2) Once the Newtonian picture is accepted, it becomes a purely *empirical* problem to determine, for a given system of interest, what are its constituent particles and what are the forces imposed on them. Thus in the light of the Newtonian approach, there followed historically an enormous empirical shift. We can in fact directly see this shift, in our own time; molecular biology is essentially the result of the belated percolation of Newtonian concepts directly into biology, something that only became technically feasible within the past 30 or 40 years.

(3) The encoding and decoding arrows in Fig. 1 that Newton posited have become axiomatic, and thus in effect invisible. They are no longer recognized as the pivots on which the whole picture turns, but have become as necessary a part of scientific thought as Euclidean geometry was prior to 1800.

In summary, the mathematical image that Newtonian mechanics gives us of a family of structureless particles acted upon by forces is as follows. Such a system is represented by a manifold of possible *phases* (configurations plus velocities) on which a set of equations of motion, representing *forces*, is superimposed. It must be emphasized that this mathematical image is *not* an abstraction; every shred of physical reality of such a system has a corresponding mathematical image somewhere in this picture and requires only a *technical* mathematical exercise to make it visible. All this, by the way, is completely preserved in the transition to quantum physics; the only novelty (though it is a major one) resides in the replacement of the Newtonian phases by a more general space of states, related in a much more complicated way to what we actually measure.

With this as background, we turn now to a brief consideration of the paradigmatic aspect of the Newtonian conception. This involves the use of the *language* in which Newtonian mechanics is couched, to describe systems that have not been, or perhaps cannot be, described as systems of mass points—ecosystems, economic systems, chemical reactors, and so on. The *states* of such a system are the *analogs* of the Newtonian phases; the dynamical laws the *analogs* of Newtonian forces. In this light, as we have noted earlier, *every* mode of system description known to me is nothing more than a paraphrase or an adaptation of the Newtonian *language*, interpreted by Newton initially in terms of particulate mechanical systems, but now interpreted in a wider context.

These two aspects of Newtonian ideas, the reductionist and paradigmatic, come together in the fundamental reductionist assumption that, among all possible encodings of a natural system, there is a *biggest* one, which maps *effectively* on all the others. If every material system is indeed a system of material particles, this would be the original Newtonian one. Mathematically, this biggest description is like a free object in the set of all encodings of a given system. Traditional reductionist ideas, especially in biology, rest entirely on the posited existence of such a biggest encoding and on the assumption that it indeed maps effectively on every other encoding.

We now remark that the relational models described earlier are not of this Newtonian character; they fall outside the paradigm from the very beginning. We will now turn to a study of what this means, and in the process we will find that the Newtonian encodings, far from being universal, are very special indeed.

VIII. MAPPINGS AND "SYSTEM LAWS" IN THE NEWTONIAN PARADIGM

In the previous sections we posed the problem of realizing relational models like the (M, R)-systems. The basic ingredients of such relational models are abstract sets and mappings between them. We take up that discussion again now, in the context of what we have said about the Newtonian paradigm. In particular, we will look more closely at the ways in which mathematical mappings, or functional relations, appear as images of physical reality in Newtonian modelling relations described by Fig. 1. By seeing explicitly what attributes of material reality in natural systems are actually encoded by mappings via the Newtonian paradigm, we will be in a stronger position to talk about realizations of mappings in the (M, R)-systems. But as we shall see, there are several surprises in store.

To see what is involved in the dual activities of encoding events into mappings, and realizing mappings as events, we will consider a few typical examples.

Our first example will be one that I have analyzed in detail elsewhere, a rather simple and degenerate (i.e., nondynamical) thermodynamic situation but one that illustrates clearly some of the basic issues. This is the van der Waals equation, a typical thermodynamic equation of state describing the equilibrium points of a class of nonideal gases. As originally formulated, this equation of state could be written in the form

$$(p + a/v^2)(v - b) = rT. \tag{3}$$

Here p, v, T are interpreted as the thermodynamic variables of state (pressure, volume, and temperature, respectively), and a, b, and r are *parameters*.

We shall not pause to discuss here the rather lengthy arguments required actually to allow us to consider Eq. (3) as an encoding of material reality; this requires, among other things, characterizing the manner in which events are converted to numbers through specific transducers (meters). Rather, we shall consider Eq. (3) as (1) a mathematical expression and (2) as an encoding, a symbolization, or a description of some material situation.

In purely mathematical terms, Eq. (3) is a functional relation involving six arguments:

$$\Phi(p, v, T, a, b, r) = 0. \tag{4}$$

Mathematically, all the arguments of this relation are equivalent; they are simply arguments. But as *encodings*, there is the most profound distinction in the *interpretation* of these arguments. This distinction is *lost* if we only look at the encoded mathematical version Eq. (3); it has been abstracted away. It is precisely what is being lost in the encoding process that we shall examine here.

Intuitively, the arguments in Eq. (3) fall into three classes:

(1) The *parameters* (a, b, r). These have to do with the particular kind (*species*) of gas under consideration (e.g., O_2, CO_2, NO, air...). At this point, it is immaterial whether we say that the parameters (a, b, r) determine or are determined by this species. For this reason we shall call these parameters the *genome* of Eq. (3).

(2) A pair of the remaining arguments [say, (p, T)], whose values are determined by the character of the *environment* with which our gas is interacting. Normally, these variables are interpreted as those under the experimenter's *control*. The point is that the specific values or numbers assumed by these quantities are determined by processes that *do not obey* Eq. (3). We will therefore call these arguments (p, T) *environments* of Eq. (3).

(3) The remaining arguments of Eq. (3); in this case the volume v. At equilibrium, the value of v is completely determined by Eq. (3) when the genome (a, b, r) and the environment (p, T) are specified. For this reason we shall call this value the *phenotype* of Eq. (3).

Thus by looking informally at the nature of the encoding Eq. (3), we see the greatest possible distinctions among its arguments, distinctions that are entirely missing from the mathematical structure of Eq. (3) itself.

Let us see if we can get them back. We will rewrite Eq. (4) in two steps. First, instead of regarding Eq. (4) as a single function of six arguments, let us rather regard it as a three-parameter family of functions of three arguments:

$$\{\Phi_{abr}(p, v, T) = 0\}. \tag{5}$$

Here we are using the parameters we identified as *genomic* as *coordinates in a function space*, and not as arguments. This introduces a purely mathematical distinction into our encoding, destroying the misleading symmetry among the arguments of the van der Waals equation with which we started. Of course, *which* arguments we use in this fashion must be made part of the initial encoding.

Next, we shall rewrite the three-parameter family of (5) of *relations* as a three-parameter family of *mappings*, from environment to phenotypes:

$$\Phi_{\mathrm{abr}}: (p, T) \to (v). \tag{6}$$

In this form, the distinction between *environment* and *phenotype* is manifested mathematically, as an integral part of the encoding itself. But, of course, *which* arguments are chosen as environment and which are chosen as phenotype must also be an integral part of the initial encoding process itself.

We have already noted that the van der Waals equation is a degenerate situation, pertaining to equilibrium values (i.e., limiting values for long times). We can inch toward dynamical encodings if we make one further reinterpretation of (6). Namely, we recall that the situation described by the van der Waals equation is: *whatever the initial volume* v_0, once genome and environment are specified, the ultimate phenotype is that value v given by (3). Thus we can finally rewrite (6) as a three-parameter family of *operators*, mapping "initial phenotypes" v_0 into phenotypes v:

$$v_0 \xrightarrow{\Phi_{\mathrm{abr}}(p,T)} v. \tag{7}$$

If we do this carefully enough, we can actually use the map Φ_{abr} to generate a vector field on the space of phenotypes (i.e., a true dynamics). However, this dynamics is extremely degenerate in the present situation, a consequence of the fact that thermodynamics from which an equation of state like the van der Waals equation is taken is dynamically as singular as it can get (e.g., the manifold of equilibria, or critical points, where nothing can happen, is of the highest possible dimension). This extreme nongenericity is in fact typical of the situations with which conventional physics deals and is one of the reasons (we shall see others) that biology (among many other things) falls outside it and that it is most ironic for the physicist to fancy himself as a purveyor of "universal laws."

By these rewritings, we have embodied explicitly, in the mathematics into which our physics is encoded, at least some of the basic distinctions that were originally lost. In the process, we have gone from a single manifold with a relation on it to something like a fiber space, with genomes as base space and state spaces, or operators on phenotypes, as fiber. We shall not pause here to discuss the rather far-reaching ramifications of this picture, particularly those associated with the *stability* of the parameterized families Φ_{abr}, with evolution and development. We shall, however, touch on these matters, in another guise, in subsequent sections.

For the moment, we merely wish to mention that the forms (5)–(7) allow us to define a variety of *partial* maps. For instance, if we keep genome fixed, we get a mathematical relation between phenotype and environment. If we keep environment fixed, we get a mathematical relation between genome and phenotype. Each of these is mathematically *also a mapping*. But the *interpretations* or realizations of these partial mappings are completely different from one another, even in this utterly simple context. We begin to glimpse, then, some of the subtleties involved in the problem of realizing an abstract mathematical mapping. To do so, we need to know the whole encoding from which it comes.

This is, in fact, only the tip of one of the icebergs implicit in the Newtonian encoding. We shall look at this particular iceberg in more detail, from a new angle, in the next section. We shall then turn our attention to another, even bigger one.

IX. CAUSALITY

Since the distinctions that are our major interest become obliterated as a consequence of the abstractions inherent in the Newtonian encoding, we need another language to make them manifest. The only language that I have found appropriate for this purpose came from a most unexpected quarter, from the old Aristotelian doctrine of the categories of causation. It is a language that everyone believes has been made utterly obsolete by the advent of the Newtonian paradigm. However, we have already implicitly used this language heavily; as we shall see, it in fact permeates the discussion of the preceding section. Let us pause then to review briefly some of its essential features.

To Aristotle, all science is animated by a single question: Why? Science must answer this question "Why?"; it must say, "Because." In so doing, depending on the context, science becomes the vehicle for both explanation and prediction. Aristotle's basic contribution was to recognize that there are *different* and *inequivalent*, but equally valid, ways of saying "because." If we single out some event or thing in the external world and ask why it is what it is, then what we have singled out is the *effect* of its *causes*, and these causes are embodied in the different ways that Aristotle distinguished, in which we can answer the question we have asked.

Aristotle's explicit discussion of the categories of causation is, by modern standards, superficial and incomplete and, oddly, is couched primarily in terms of material artifacts. Following this discussion, let us suppose we are interested in "understanding" something like a house. The four Aristotelian categories of causation for the house are then:

(1) *Material cause*: The house is what it is *because* of the wood, bricks, glass, metal, and so on, of which it is composed. These constitute the material cause of the house and comprise one way of saying why the house is what it is.

(2) *Formal cause*: The house is what it is *because* of the blueprint or plan that it realizes. This blueprint or plan constitutes the formal cause of the house, and provides another way of saying why the house is what it is; a way different from, and inequivalent to, but equally valid as, the material cause.

(3) *Efficient cause*: The house is what it is *because* of the labor of its builders, who manipulated the constituent materials in accord with the blueprint or plan. We now have a third way of saying why the house is what it is; again different from, inequivalent to, but equally valid as the other ways.

(4) *Final cause*: The house is what it is *because* someone required shelter. This way of saying why was for Aristotle the most important because it involved *telos*; volition, goal, end. The study of the final causes of things is accordingly called *teleology*. Largely as a result of the Newtonian paradigm, the whole concept of teleology has become anathema; the corresponding adjective is as close to a defamation as there is in science.

What Aristotle was essentially doing in his discussion of the categories of causation was giving names to, and thereby distinguishing between, certain kinds of *relations between events*. As we noted in Section V, this is precisely the province of causality, one of the twin pillars supporting our belief in natural law. Nevertheless, it is widely believed that causality is not a scientific concept; that *true* science only began when Aristotelian ideas about causality were discarded and replaced by the Newtonian paradigm. Indeed, Bertrand Russell wrote an influential article pointing out precisely that the word *cause* never appears, as a technical term, in any "advanced science" such as "gravitational astronomy." For Russell, such "advanced sciences" consist *entirely* of mathematical relations (of the kind we have seen in the preceding section), which are *deterministic* without being *causal*. What Russell had done, in effect, was to forget about the whole left-hand side of the diagram in Fig. 1, including the encoding and decoding arrows. He tacitly took it as axiomatic that the Newtonian encoding was the only one; thus, we need never look again at the events themselves, and can content ourselves with looking exclusively at the mathematical images provided by that encoding. And of course it is perfectly true that, in this *formal* world of mathematical images, the word *cause* never appears; it has been lost in the encoding process, along with many other important things.

Nevertheless, we can restore the Aristotelian ideas by, as we did before, superimposing upon the Newtonian paradigm an additional informal layer of interpretation to compensate for the missing or unencoded properties we need. As we have seen, the essence of the Newtonian paradigm involves a *genome-parameterized family of environmentally determined operators, acting on a space of phenotypes*. If we now identify the space of phenotypes with the space of "internal states," or mechanical phases, it is not hard to show that the

Newtonian paradigm can be expressed in conventional mathematical language as a system of equations of motion for the phenotypes or states:

$$\frac{d\mathbf{z}}{dt} = \psi_{\mathbf{g}}(\mathbf{z}, \boldsymbol{\beta}(t)) \tag{8}$$

Here $\mathbf{z}$ is a state vector or phenotype vector; $\mathbf{g}$ is a genome vector, and $\boldsymbol{\beta} = \boldsymbol{\beta}(t)$ is a vector of environments, variously called *inputs* or *forcings* or *controls.* Relations such as this are precisely what Bertrand Russell *identified* with "advanced sciences."

Dynamical relations such as (8) can be converted into a mathematically equivalent form through a process of *integration*:

$$\mathbf{z}(t) = \int_{t_0}^{t} \psi_{\mathbf{g}}(\mathbf{z}, \boldsymbol{\beta}(\tau))\, d\tau + \varphi(\mathbf{z}(t_0)) \tag{9}$$

Although (8) and (9) are *mathematically* equivalent, they are *epistemologically* very different. The relation (8) is a local relation, relating the values of observable quantities at any given instant. The relation (9), on the other hand, relates the values of observable quantities at *different* instants. However, we shall not pursue the implications of this fact here.

Now let us look at the relation (9), which is one way of expressing the Newtonian paradigm, in terms of the Aristotelian ideas regarding categories of causation. Specifically, if we regard $\mathbf{z}(t)$, the phenotype or state of a system at some instant of time, as the *effect*, then we have:

Material cause ≡ initial state $\mathbf{z}(t_0)$;

Formal cause ≡ genome $\mathbf{g}$.

Efficient cause ≡ operator $\int_{t_0}^{t} \psi_{\mathbf{g}}(\ldots, \boldsymbol{\beta}(\tau)\, d\tau$

We note explicitly that there is no *final cause* visible in this picture. In retrospect, it is this fact more than any other that has led to the profound belief that finality is incompatible with science. Indeed, any attempt to impose a category of final causation onto the Newtonian encoding destroys it completely. Insofar as finality involves the effect of future inputs, or future state, upon present change of state, the idea of *anticipation* has been expunged from serious science without further thought. The only exceptions I know involve some interesting discussions of, for example, advanced potentials as physically meaningful solutions of classical wave equations, and the interpretation of variational principles, which specify paths between prior and subsequent configurations, in anticipatory terms. But these are never taken seriously, for the reasons we have sketched already.

From what we have already said, innocent as it may appear, we can now reformulate the whole Newtonian paradigm in a way that manifests more clearly than any other how very special it in fact is. In brief, the essential fact is that *the Newtonian paradigm pertains to only those systems for which the categories of causation can be segregated into mathematically independent structures.* For instance, under the interpretation we have given, the very idea of a state space tacitly means that the category of material causation can be split off from the other causal categories as a completely independent unit; the same is true for all the other causal categories. Stated another way: Any observable quantity pertaining to a system, or to its environment, can be assigned exclusively to one or another of the causal categories, once assigned, it stays in that category, and can play no other causal role.

When cast in this light, we can see in a most vivid way that the whole Newtonian paradigm, which as noted earlier has persisted essentially unchanged from its inception as the only and universal mode of system description, is not that at all. Indeed, the class of special systems that it describes should be given a corresponding special name; we shall call such systems *simple systems* or *mechanisms.* The Newtonian paradigm tacitly says that *every system is a mechanism*; from the discussion we have given earlier, it is manifest that this need not be so. It will be the task of the following section to consider the question of whether there are natural systems that are not mechanisms and if so, how they are to be described.

For the remainder of the present section, however, we shall stay within the confines of the Newtonian paradigm (i.e., within the class of simple systems) and consider briefly a few of the implications of the fact that the categories of causation, in that paradigm, while segregated into independent mathematical structures, are nevertheless *inequivalent.* As far as I know, these implications have never been addressed, or even explicitly recognized, largely because of the intrinsic and historical peculiarities of the Newtonian paradigm itself.

A first conclusion that we can draw pertains to the idea of *simulation.* For instance, suppose we wish to simulate some dynamical system of the form (8), on an analog or a digital computer. This involves the construction of a diagram like that of Fig. 1 that encodes the sequence of state transitions in our original system (call it the *prototype*) into corresponding elements of the simulator. In terms of our preceding discussion, the sequence of state transitions in the prototype involves *material* causation in that system. However, in the simulator, state information about the prototype is always encoded as *data* or inputs to the simulator. Thus in the simulator the data are related to *efficient causation.* In fact, the whole idea of computation pertains essentially to the manipulation of efficient causation, with formal cause as *program* and material cause as "*hardware.*" Thus although a modelling relation can thereby be established between prototype and simulator in a *formal* way, the causal structures in the two are entirely different (and so too, incidentally, are the corresponding *inferential* structures).

It is generally believed that any dynamics can be simulated in this sense, i.e., that a modelling relation can always be established between material causation in some prototype and efficient causation in an appropriate simulator. Whether this is true or not I do not know. However, I do know that the reverse is false, that we cannot in general *realize* efficient causality in terms of material causality (hence their inequivalence). One well-known instance of this fact may be mentioned briefly here. Long ago (cf. Burks, 1966) John von Neumann gave a discussion of a putative "self-reproducing automaton." This discussion was based on the idea of a "universal constructor," which was in turn derived from Turing's earlier (1936) idea of a universal digital computer, or universal Turing machine. Basically, von Neumann's argument was that following a blueprint to construct something was just as much of an algorithmic process as following a program to compute something, and *therefore* that anything true of computation was necessarily equally true of construction. But of course a constructor must manipulate material causation, while a computer manipulates efficient causation; it does not follow, and is in fact false, that a universal simulator implies anything about a universal constructor. Indeed, for the same reason, one must be very careful in extrapolating from the properties of formal systems such as neural networks or automata back to the material properties of such processes as biological development or cellular control; for these too are exercises in the realization of efficient causation by material causation in some prototype. It was also for this reason that my first attempt to realize dynamically the (M, R)-systems (cf. Section V) did not work.

The same inequivalence between causal categories is manifested in the genotype–phenotype dualism, which animates the entire theory of evolution in biology and which in its essential features underlies any equation of state generated by a Newtonian encoding of a material system. That indeed is why we chose the terminology (genome, environment, phenotype) we used in classifying the arguments of such equations, and which as we have seen eventually manifests itself directly in causal terms. For instance, the D'Arcy Thompson "Theory of Transformations" (Thompson, 1917) asserts fundamentally that all closely related phenotypes are also similar; here "*closely related*" pertains to genotypes and *similar* pertains to phenotypes (or better, to environment–phenotype relations). Here the relation is between *formal cause* (the genomes) and efficient cause. The inequivalence of these two categories of causation means precisely, in modern mathematical language, that there exist bifurcating genomes, in any neighborhood of which there will be dissimilar phenotypes. This simple fact goes a long way to help us to understand the basis for what evolutionists call *macroevolution*, but that again is another story.

The inequivalence of causal categories manifests itself in many other ways. It is intimately involved, for instance, in the inability to infer anything much about, for example, the material structure of an enzyme from a knowledge of its substrates and products, and of a knowledge even of its kinetic parameters in the substrate-product conversion. The latter involves again efficient causation; the structure is material causation. More generally, it manifests itself in what is usually called the *system identification problem*, which can be translated precisely into an attempt to predict material cause from knowledge about efficient or formal cause. Going the other way, we cannot predict, for example, that a given material structure will play any particular kind of functional role (e.g., that a particular protein structure is an enzyme, let alone what its substrates are).

And even in pure mathematics itself, results like Gödel's incompleteness theorems are manifestations of this same inequivalence, now expressed in completely formal terms.

Thus we can now appreciate better some of the real difficulties involved in attempting to realize physically a relational structure such as an (M, R)-system. Indeed, if we look again at the (M, R)-system (1) in the light of the discussion of the past few sections, we can see just how complicated a little structure it is, in terms of its causal correlates. More precisely, the sets and mappings that comprise it, though all *mathematically* of a common character, have completely different causal interpretations; sometimes the same thing (e.g., the metabolic map f) is involved in several of these categories simultaneously. Indeed, it may well be (although I cannot prove it) that even such an elementary relational structure as this (M, R)-system in fact cannot be realized within the confines of the Newtonian paradigm at all.

We shall now turn to a consideration of the following question: If indeed the Newtonian paradigm is so special, what alternative is there? If the Newtonian paradigm describes only the limited class of systems that we have called mechanisms, or simple systems, how are we to describe those systems (if any) that fall outside that class? This will be the subject of the next section.

X. COMPLEX SYSTEMS

We now turn our attention to the question of what a mathematical image of a complex system should be like. That is, we suppose that there is a complex system sitting on the left-hand side of the diagram of Fig. 1 and ask what kind of mathematical object can go into the right-hand side so that the diagram will commute.

Before considering some explicit possibilities, we can already draw some general conclusions regarding these new mathematical images. Some of these are:

(1) There can be no such thing as a "state space" in such an image, which can be fixed once and for all. More generally, the causal categories (which become much more subtle in this context) cannot be segregated into disjoint classes; at least some elements of our image play several causal roles simultaneously. Moreover, these causal roles can shift in the course of time as a consequence of system dynamics.

(2) A complex system will have a multitude of partial images of the Newtonian type, which can in some sense "approximate" to the behavior of the system. But this approximation of complexity by simplicity is only local and temporary. This means that, as the complex system develops in time, any such simple approximation ceases to describe the system in the sense of Fig. 1; the *discrepancy* between what the complex system is actually doing (arrow 1 in the diagram) and the behavior of the simple approximation (arrows 2 + 3 + 4) grows in time. When the discrepancy becomes intolerable, we must replace our initial simple approximation by another. The discrepancy between the behavior of a complex system and any such simple approximation is, depending on the context, called *error* or *emergence*.

(3) Even though a complex system has a multitude of partial simple descriptions, we cannot construct from them a single "largest" description that is also simple. In this sense, the reductionistic paradigm fails for complex systems.

Just from these few properties, which follow essentially only from the non-Newtonian character of complexity, we can see that this kind of world of mathematical images must have very different properties from the one we are used to.

There are now two questions to answer: (1) Is there a world of mathematical structures with these characteristics, that can be put into a modelling relation with a complex system, and (2) are there complex systems in nature, which realize such mathematical structures? We shall turn to the first of these questions now, leaving the second for our final comments.

To motivate this discussion, let us return to a consideration of conventional Newtonian images of the form (8). More specifically, let us look at a dynamical system

$$dx_i/dt = f_i(x_1, \ldots, x_n) \tag{10}$$

leaving out of consideration for the moment the genomic and environmental aspects. Originally motivated by an attempt to establish some relation between dynamical and informational ways of treating such a system, and following an earlier treatment of Higgins (1967), I considered the (observable) quantities

$$u_{ij}(x_1, \ldots, x_n) = \partial/\partial x_i(dx_i/dt).$$

A great deal about the stability of (10) can be inferred from these quantities; in fact, most of the significance of the u_{ij} lies in their *signs*, and not so much in their specific values.

These quantities, as Higgins noted, have informational correlates. For instance, if u_{ij} is positive in a state, it means, for example, that an increase in x_j increases the rate of production of x_i (or that a decrease in x_j decreases the rate of production of x_i). It thus makes sense to call x_j an *activator* of x_i under these circumstances. Likewise, if u_{ij} is negative in a state, it is reasonable to call x_j an *inhibitor* of x_i. In fact, on this basis, we can convert the dynamical system (10) into an informational network, quite analogous to a neural net (cf. Rosen, 1979).

There are many situations in biochemical, morphogenetic, ecological, and neural theory in which the activation–inhibition language seems more natural than the dynamic one. Thus the obvious question was whether, given such an activation–inhibition network, we could in effect invert the preceding discussion and recover a set of rate equations (10). What we must do is straightforward; construct the differential forms

$$\omega_i = \sum_{i=1}^{n} u_{ij}\, dx_j. \tag{11}$$

If ω_i is exact, there is an observable f_i such that $df_i = \omega_i$. Put this observable equal to dx_i/dt and we are done. However, if these forms are *not exact*, we have an activation–inhibition network that cannot be realized by a set of rate equations.

For a differential form to be exact, if $n > 2$, is a most nongeneric situation. There are some standard necessary conditions for exactness, which may be written as

$$(\partial/\partial x_k)u_{ij} = (\partial/\partial x_j)u_{ik}.$$

Now these quantities

$$(\partial/\partial x_k)u_{ij} = u_{ijk}$$

also have an informational connotation. In brief, it is easy to see that, if u_{ijk} is positive in a state, it means that x_k enhances or potentiates the effect of x_j on x_i. Under these circumstances, we can call x_k an *agonist* of x_j. Likewise, if u_{ijk} is negative, x_k attenuates the effect of x_j on x_i, and we can call x_k an *antagonist* of x_j. We see then that the conditions for exactness of (11) become $u_{ijk} = u_{ikj}$ for all indices i, j, k; the activator–inhibitor relation and the agonist–antagonist relation are completely symmetrical. This too is a highly nongeneric situation. Thus we can conclude that the informational description is more general than that given by dynamical laws like (10).

In a nutshell, we can continue iterating the process whose first two steps we have described, constructing successive networks $u_{ij}, u_{ijk}, u_{ijkm}, \ldots$, each of which modulates the properties of its predecessors. If we start from a set of rate equations (10) all of these layers of networks are derivable from the rate equations; from any one we can reconstruct all the others in the obvious fashion. On the other hand, if any of the differential forms in these networks are inexact, the networks become independent of each other, and there is no set of rate equations from which all the layers follow.

Networks of this kind provide the first concrete examples of a class of mathematical images satisfying the requirements we have indicated earlier, for representing systems that are not simple. We cannot go into technical details in this short space, but we can set forth certain conclusions about these systems of layers of informational interactions:

(1) The class of all such images can be converted into a general mathematical structure called a category. In this category, the category of Newtonian images (i.e., of dynamical systems or state-determined systems, which are the images of simple systems or mechanisms sits as a very small subcategory, just as the rational numbers sit as a subset of measure zero in the set of real numbers. Moreover, just as in the case of the rational numbers, every object in the big category can be regarded as the limit of a sequence of elements in the small one. Thus we have the notion of the "approximation" of a complex system by a simple one; but as noted previously, this "approximation" is only local and temporary.

(2) The causal structure of the objects in the big category turns out to be much more complicated than is true in the subcategory of dynamical systems. The infinitely greater richness of the causal structures possible in complex systems provides one way to understand the growth of the deviation between what a complex system will do and what a simple approximation does. Moreover, this greater richness of causal structure makes the problem of interpretation or explanation of experimental observation very different from what we are used to.

(3) In complex systems, an ideal of *final* causation or *anticipation* can be introduced in a perfectly rigorous, nonmystical way. Briefly, a complex system may contain predictive models of itself and/or its environment, which it can utilize to modify its own present activities.

(4) Because complex systems ultimately depart from the behavior predicted on the basis of *any* simple approximation, their behavior appears to us to be surprising and *counterintuitive*.

There are many other conclusions to be drawn from the class of mathematical images that we have briefly described and their relation to the Newtonian ones that approximate to them. This relation, on the one hand, explains why we have been able to go as far as we have with the Newtonian paradigm and why, on the other hand, we can in many areas get no further. The relation between complex systems and their simple approximations may be likened to the situation faced by the early cartographers, who were attempting to map the surface of a sphere while armed only with pieces of planes. Here the sphere plays the role of a complex system, while a piece of (tangent) plane is like a simple approximation. As long as we only map local regions, the planar approximation suffices, but as we try to map larger and larger regions, the discrepancy between the map and the surface grows as well. As noted earlier, this discrepancy can be called either *error* (which can be located either in the sphere or in the planar map) or *emergence* (of a new property of the surface; namely, its curvature). Thus if we want to make accurate maps of large regions of the sphere, we have to keep shifting our tangent planes. The surface of the sphere is in some sense a limit of its planar approximations, but to specify it in this way requires a new global concept (the topology of the sphere; i.e., its curvature) that cannot be inferred from local planar maps alone.

XI. AN ALTERNATE APPROACH: "INFORMATION"

Because it is interesting and important in its own right and because it leads to an alternate mode of entry into the universe of complex systems that is of some independent interest, we shall turn in the present section to yet another analysis of the idea of "information." Ever since Shannon began to talk about "information theory" (by which he meant a probabilistic analysis of the deleterious effects of propagating signals through "channels"; cf. Shannon, 1949) this concept has been relentlessly analyzed and reanalyzed. The time and effort expended on these analyses must surely rank as one of the most unprofitable investments in modern scientific history; not only has there been no profit, but the currency itself has been debased to worthlessness. Yet in biology, for example, the terminology of information intrudes itself insistently at every level—code, signal, program, computation, recognition. It may be that these informational terms are simply not scientific at all, that they are an anthropomorphic stopgap, a *façon de parler* that merely reflects the immaturity of biology as a science, to be replaced at the earliest opportunity by the more rigorous terminology of force, energy, and potential that are the province of more mature sciences (i.e., physics) in which "information" is never mentioned. Or it may be that the informational terminology that seems to force itself upon us bespeaks something fundamental, something that is *missing* from physics as we now understand it.

In human terms, information is easy to define; it is anything that is or can be the answer to a question. Therefore we shall preface our more formal considerations with a brief discussion of the status of interrogatives, in logic and in science.

The amazing fact is that interrogation is never a part of formal logic, including mathematics. The symbol "?" is not a logical symbol, as, for instance, are "$\vee$," "$\wedge$," "$\exists$," or "$\forall$," nor is it a mathematical symbol. It belongs entirely to informal discourse, and as far as I know, the purely logical or formal character of interrogation has never been investigated. Thus if "information" is indeed connected in an intimate fashion with interrogation, it is not surprising that it has not been formally characterized in any real sense. There is simply no existing basis on which to do so.

I do not intend to go deeply here into the problem of extending formal logic (always including mathematics in this domain) so as to include interrogatories. What I want to suggest here is a relation between our informal notions of interrogation and the familiar logical operation "$\Rightarrow$"—the conditional, or the implication operation. Colloquially, this operation can be rendered in the form "If A, then B." My argument will involve two steps. First, I will argue that *every* interrogative can be put into a kind of conditional form:

If A, then B?

(where B can be an indefinite pronoun such as *who*, *what*, etc., as well as a definite proposition). Second, and most important, I will argue that every interrogative can be expressed in a more special conditional form, which can be described as follows. Suppose I know that some proposition of the form

If A, then B

is true. Suppose I now change or vary A, that is replace A by a new expression which I will call δA. The result will be an interrogative, which I can express as

If δA, then δB?

Roughly, I am treating the true proposition "If A, then B" as a reference, and I am asking what happens to this proposition if I replace the reference expression A by the new expression δA. I could of course do the same thing which B in the reference proposition, replace it by a new proposition δB and ask what happens to A. I assert that every interrogative can be expressed this way, in what I shall call a *variational form*.

The importance of these notions for us will lie in their relation to the external world, most particularly in their relation to the concept of *measurement* and to the notions of causality to which they become connected when a formal or logical system is employed to represent what is happening in the external world (i.e., to describe some physical or biological system or situation) (cf. Section VI).

Before doing this, I want to motivate the two assertions made earlier regarding the expression of arbitrary interrogatives in a kind of conditional form. I will do this by considering a few typical examples and leaving the rest to the reader for the moment.

Suppose I consider the question

"Did it rain yesterday?"

First, I will write it in the form

"If (yesterday), then (rain)?"

which is the first kind of conditional form described earlier. To find the variational form, I presume I know that some proposition such as

"If (today), then (sunny)"

is true. The general variational form of this proposition is

"If δ(today), then δ(sunny)?"

In particular, then, if I put

$$\delta(\text{today}) = (\text{yesterday})$$

$$\delta(\text{sunny}) = (\text{rain})$$

I have indeed expressed my original question in the variational form. A little experimentation with interrogatives of various kinds taken from informal discourse (of great interest are questions of classification, including existence and universality) should serve to make manifest the generality of the relation between interrogation and the implicative forms described earlier. Of course this cannot be *proved* in any logical sense, since as noted earlier, interrogation sits outside logic.

It is clear that the notions of observation and experiment are closely related to the concept of interrogation. That is why the results of observation and experiment (i.e., data) are so generally regarded as being information. In a formal sense, simple observation can be regarded as a special case of experimentation; intuitively, an observer simply determines what *is*, while an experimenter systematically perturbs what is and then observes the effects of his perturbation. In the conditional form, then, an observer is asking a question that can generally be expressed as:

"If (initial conditions), then (meter reading)?"

In the variational form, this question may be formulated as follows: Assuming the proposition

"If (initial conditions = 0), then (meter readings = 0)"

is true (this establishes the reference and corresponds to calibrating the meters), our question becomes

"If δ(initial conditions = 0), then δ(meter readings = 0)?"

where simply

$$\delta(\text{initial conditions} = 0) = (\text{initial conditions})$$

and

$$\delta(\text{meter readings} = 0) = (\text{meter readings}).$$

The experimentalist essentially takes the results of observation as his reference and thus basically asks the question that in variational form is just

"If δ(initial conditions), then δ(meter readings)?"

The theoretical scientist, on the other hand, deals with a different class of question, namely, the questions that arise from assuming a δB (which may be B itself) and asking for the corresponding δA. This is a question that an experimentalist cannot approach directly, not even in principle. It is mainly the difference between the two kinds of questions that marks the difference between experiment and theory, as well as the difference between the explanatory and predictive roles of theory itself; clearly, if we give δA and ask for the consequent δB, we are predicting, whereas if we assume a δB and ask for the antecedent δA, we are explaining.

It should be noted that exactly the same duality arises in mathematics and logic themselves; i.e., in purely formal systems. Thus a mathematician can ask (*informally*): If (I make certain assumptions), then (what follows)? Or he can start with a conjecture and ask: If (Fermat's last theorem is true), then (what initial conditions must I assume to construct explicitly a proof)? The former is analogous to prediction, the latter to explanation.

When formal systems (i.e., logic and mathematics) are used to construct images of what is going on in the world, then interrogations and implications become associated with ideas of causality. Indeed, we have seen that the whole concept of natural law depends precisely on the idea that causal processes in natural systems can be made to correspond with implication in some appropriate descriptive inferential system.

But the concept of causality is itself a complicated one; this fact has been largely overlooked in modern scientific discourse, to its cost. That causality is complicated was already noted by Aristotle, when he pointed out that there were four distinct categories of causation, four ways of answering the question *why*. These categories he called *material cause*, *formal cause*, *efficient cause*, and *final cause.* We have already seen that these categories of causation are inequivalent; hence there are correspondingly *different kinds of information,* associated with different causal categories. These different kinds of information have been confused, mainly because we are in the habit of using the same mathematical language to describe all of them; it is from these inherent confusions that much of the ambiguity and murkiness of the concept of information ultimately arises. Indeed, we can repeat in the present context what we have already noted: The very fact that the same mathematical language does not (in fact, cannot) distinguish between essentially distinct categories of causation means that the mathematical language we have been using is in itself somehow fundamentally deficient and that it must be extended by means of supplementary structures to eliminate those deficiencies.

However, there is yet a deeper relation between information, interrogation, causality, and mathematics implicit in the preceding discussion. This relation has important consequences for the structure of mathematics itself. Let us introduce it by noting that there is an exact parallel between the Newtonian paradigm, with its partition of system description into states plus dynamical laws, and the structure of mathematical formalisms, with their corresponding partition into propositions and production rules, or rules of inference. We have already noted that the Newtonian paradigm cannot accommodate the Aristotelian category of final causation (cf. Section IX). It is for precisely the same reason that *logical or mathematical systems cannot accommodate interrogation,* on which we have based the idea of "information"; namely, *an interrogation or question always involves a telic aspect.* It is precisely this telic aspect that eludes capture in a simple system, whether that system be a real material system or a mathematical image of such a system. We can now see intuitively that any attempt to construct a logical or mathematical formalism big enough to accommodate interrogation will lead us again directly into the category of complex systems, this time by a purely formal route. Indeed, seen in this light, the famous Gödel theorems, to which we have already referred, are about the approximation of a complex formalism by simple ones; here a "complex" formalism is, roughly speaking, one big enough to encode *within itself* a question of the form, "if (p), then (provable)?"

It may not be out of place here to mention parenthetically that (1) the telic nature of interrogation and (2) the close relation between observation (experiment) and interrogation are at the root of the conundrums associated with many analyses of the measurement problem in quantum mechanics. Quantum mechanics is entirely a classical theory in its partition of the world into states plus dynamical laws. Thus from our point of view, it is entirely subject to the analysis we have provided. Its only novel feature (and it is, of course, a central one) is in its postulation of what constitutes a state and how such a state is related to what is actually measured.

To conclude this section, let us return briefly to the role of interrogation in the theory of complex natural systems. We recall again that information, for us, is the answer to a question and that questions can be put into what we called the variational form: If δA, then δB? The connecting bridge between these considerations and physics lies in the interpretation of a question in variational form and the general concept of *virtual displacement*. In this guise the abstract considerations we have developed have already played a central role in classical physics—in mechanics, field theory, and thermodynamics.

In mechanics, a virtual displacement is a small, imaginary change imposed on the *configuration* of a mechanical system, with the forces kept fixed. The animating question is: If such a virtual displacement is made, then what happens? The answer, in mechanics, is: If the mechanical system is at equilibrium, then the (virtual) work done by the impressed forces as a result of the virtual displacement must vanish. This principle of virtual work is a static principle (i.e., pertains only to equilibrium), but it can be extended from statics to dynamics, where it is known as *D'Alembert's principle* and leads directly to the equations of motion.

The reader will observe now that the informational networks of Section X are defined by functions that answer questions of the form posed by virtual displacements. For instance, a function such as

$$u_{ij} = \partial/\partial x_j(dx_i/dt)$$

answers a question such as

$$\text{If } \delta x_j, \text{ then } \delta(dx_i/dt)\,?$$

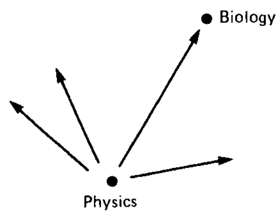
In fact, pursuing these considerations leads precisely back to the same informational nets as the ones we have seen before. But this time we do not have to detour through dynamical systems (i.e., simple systems) at all; we can proceed entirely through informational considerations, in the guise of questions about how a (virtual) displacement of an observable affects another. The interesting thing is that several quite independent approaches lead us back to precisely the same circle of ideas.

XII. CONCLUSION

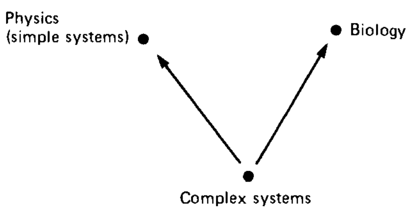
We now turn to our final question: Are there any complex systems in nature? I would argue that biology is filled with them, that the most elementary relational considerations bring us instantly face to face with this fact. Like early man, who could see the rotation of the earth every evening just by watching the sky but could not understand what he was seeing, we have been unable to understand what every organism is telling us. It cannot be stressed strongly enough that the transition from simplicity to complexity is not merely a technical matter to be handled within the Newtonian paradigm; complexity is not just complication, to be described by another number (e.g., the dimension of a state space or the length of a program), but a whole new theoretical world, with a whole new physics associated with it.

If organisms are indeed complex in our sense and if contemporary physics deals exclusively with simple systems, it follows that we cannot in principle do biology within the confines of contemporary physics. This is simply a more precise statement of what we asserted earlier; that the relation of biology to our present physics is not that of particular to general. It is not biology but physics that deals with too limited, too restricted a class of systems. Far from biology being reduced to, and hence disappearing into, contemporary physics, as the reductionists believe, it is physics that will be transformed out of present recognition by being forced to confront, head on, the problems posed by complexity. The shambles that the concept of the "open system" has made of classical thermodynamics (where after 50 years or more there is still no real physics capable of properly coping with even the most elementary open system dynamics) is as nothing compared to the impact of complexity; and thermodynamics has long been regarded with complacency as the repository of the most universal truths of physics.

Let us express the situation outlined earlier in the form of a diagram. The Newtonian paradigm and its corollaries have told us that every science is a logical consequence of physics, i.e., of the science of mechanisms. The science of mechanisms is thus the root of a tree, with biology as one of the specialized branches, as follows:



What we have argued, however, is that it is physics that is a specialized branch of a more general science of complex systems. Biology represents another branch, different from any science of mechanisms:



There are indeed relations between these two collateral branches; some of these we have tried to sketch earlier. But the relations are not reductionistic ones; they are more complicated, and more interesting, than that.

The power of the Newtonian paradigm, from this perspective, rests not in the fact that everything is a machine, but in the fact that complex systems, which by definition are not machines, can often act as if they were. We have much to learn about how this comes about, and even more about how it fails.

REFERENCES

Burks, A. (1966). "Theory of Self-Reproducing Automata." Univ. of Illinois Press, Urbana, Illinois.

Handler, Philip, Ed. (1970). "Biology and the Future of Man." Oxford Univ. Press, London.

Higgins, J. (1967). The theory of oscillating reactions. *J. Ind. & Eng. Chem.* **59**, 18–62.

Monod, Jacques (1971). "Chance and Necessity." English Translation. Knopf, New York.

Rashevsky, N. (1954). Topology and life. *Bull. Math. Biophys.* **16**, 317–348.

Rosen, R. (1964). Abstract biological systems as sequential machines. *Bull. Math. Biophysics* **26**, 103–111.

Rosen, R. (1979). Some comments on activation and inhibition. *Bull. Math. Biology* **41**, 427–445.

Shannon, C. (1949). "The Mathematical Theory of Communication." Univ. of Illinois Press, Urbana, Illinois.

Thompson, D'Arcy. (1917). "On Growth and Form." Cambridge Univ. Press, London and New York.

Turing, A. (1936). On compatible numbers. *Proc. London Math. Soc., Ser. 2* **42**, 230–265.

Index

K

L

M

N

O

P