The Reduction of Theories

11

It is a commonplace that classical mechanics is no longer regarded as the universal and fundamental science of nature. Its brilliant successes in explaining and bringing into systematic relations a large variety of phenomena were at one time indeed unprecedented. And the belief, once widely held by physicists and philosophers, that all the processes of nature must eventually fall within the scope of its principles was repeatedly confirmed by the absorption of various sectors of physics into it. Nevertheless, the period of the "imperialism" of mechanics was practically over by the latter part of the nineteenth century. The difficulties which faced the extension of mechanics into still unconquered territory, and particularly into the domain of electromagnetic phenomena, came to be acknowledged as insuperable.

However, new candidates for the office of a universal physical science were proposed, sometimes with the backing of a priori arguments analogous to those once used to support the claims of mechanics. To be sure, with some few doubtful exceptions no serious student of the sciences today believes that any physical theory can be warranted on a priori grounds, or that such arguments can establish a theory in that high office. Moreover, many outstanding physicists are frankly skeptical whether it is possible to realize the ideal of a comprehensive theory which will integrate all domains of natural science in terms of a common set of principles and will serve as the foundation for all less inclusive theories. Nevertheless, that ideal continues to leaven current scientific speculation; and, in any case, the phenomenon of a relatively autonomous theory becoming absorbed by, or reduced to, some other more inclusive theory is an undeniable and recurrent feature of the history of modern science. There is every reason to suppose that such reduction will continue to take place in the future.

The present chapter is concerned with this phenomenon, and with some of the broader issues associated with it. Scientists as well as philosophers have exploited both successful and unsuccessful reductions of one theory to another as occasions for developing far-reaching interpretations of science, of the limits of human knowledge, and of the ultimate constitution of things in general. These interpretations have taken various forms, but only a few typical ones need be mentioned here.

Discoveries relating to the physics and physiology of perception are often used to support the claim that the findings of physics are radically incompatible with so-called "common sense"—with customary beliefs that the familiar things of everyday experience possess the traits they manifest even to carefully controlled observation. Again, the successful reduction of thermodynamics to statistical mechanics in the nineteenth century was taken to prove that spatial displacements are the only form of intelligible change, or that the diverse qualities of things and events which men encounter in their daily lives are not "ultimate" traits of the world and are perhaps not even "real." But, conversely, the difficulty in finding consistent visualizable models for the mathematical formalism of quantum mechanics has been taken as evidence for the "mysterious" character of subatomic processes and for the claim that behind the opaque symbolism of the "world of physics" there is a pervasive "spiritual reality" that is not indifferent or alien to human values. On the other hand, the failure to explain electromagnetic phenomena in terms of mechanics, and the general decline of mechanics from its earlier position as the universal science of nature, have been construed as evidence for the "bankruptcy" of classical physics, for the necessity of introducing "organismic" categories of explanation in the study of all natural phenomena, and for a variety of sweeping doctrines concerning levels of being, emergence, and creative novelty.

We shall not examine the detailed arguments that culminate in these and similar contentions. However, one broad comment is relevant to most of the claims. As has been repeatedly noted in previous chapters, expressions associated with certain established habits or rules of usage in one context of inquiry are frequently adopted in the exploration of fresh fields of study because of assumed analogies between the several domains. Nevertheless, its users do not always note that, when the range of application of a given expression is thus extended, the expression often undergoes a critical change in its established meaning. Serious misunderstandings and spurious problems are then bound to be generated unless care is taken to understand the expression in the sense relevant to, and required by, the special context in which the expression has acquired a fresh use. Such alterations are particularly prone to occur when one theory is explained by, or reduced to, another theory; and the shifts in the meanings of familiar expressions that often result as a consequence of the reduction are not always accompanied by a clear awareness of the logical and experimental conditions under which the reduction has been effected. In consequence, both successful and unsuccessful attempts at reduction have been occasions for comprehensive philosophical reinterpretations of the import and nature of physical science, such as those cited in the preceding paragraph. These interpretations are in the main highly dubious because they are commonly undertaken with little appreciation for the conditions that must be fulfilled if a successful reduction is to be achieved. It is therefore of some importance to state with care what these conditions are, both for the light that the discussion of those conditions throws on the structure of scientific explanation and for the help which the discussion can provide toward an adequate appraisal of a number of widely held philosophies of science. An examination of the conditions for reduction and of their bearing on some moot issues in the philosophy of science is the central task of the present chapter.

I. *The Reduction of Thermodynamics to Statistical Mechanics*

Reduction, in the sense in which the word is here employed, is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain. For the sake of brevity, we shall call the set of theories or experimental laws that is reduced to another theory the "secondary science," and the theory to which the reduction is effected or proposed the "primary science." However, many cases of reduction seem to be normal steps in the progressive expansion of a scientific theory and rarely generate serious perplexities or misunderstandings. It will therefore be convenient to distinguish, with the help of some examples, between the two types of reduction, the first of which is commonly regarded as being quite unproblematic and which we shall ignore in consequence, while the second is often felt to be a source of intellectual discomfort.

1. A theory may be formulated initially for a type of phenomenon exhibited by a somewhat restricted class of bodies, though subsequently the theory may be extended to cover that phenomenon even when manifested by a more inclusive class of things. For example, the theory of mechanics was first developed for the motions of point-masses (i.e., for

the motions of bodies whose dimensions are negligibly small when compared with the distances between the bodies) and was eventually extended to the motions of rigid as well as deformable bodies. In such cases, if laws have already been established within the more inclusive domain (perhaps on a purely experimental basis, and before the development of the theory), these laws are then reduced to the theory. However, in these cases there is a marked qualitative similarity between the phenomena occurring in the initial and the enlarged provinces of the theory. Thus, the motions of point-masses are quite like those of rigid bodies, since the motions in both cases involve only changes in spatial position, even though rigid bodies can exhibit a form of motion (rotation) that point-masses do not. Such reductions usually raise no serious questions as to what has been effected by them.

Analogously, the range of application of a macroscopic theory may be extended from one domain to another homogeneous with it in the features under study, so that substantially the same concepts are employed in formulating the laws in both domains. For example, Galileo's *Two New Sciences* was a contribution to the physics of free-falling terrestrial bodies, a discipline which in his day was considered to be distinct from the science of celestial motions. Galileo's laws were eventually absorbed into Newtonian mechanics and gravitational theory, which was formulated to cover both terrestrial and celestial motions. Although the two classes of motions are clearly distinct, no concepts are required for describing motions in one area that are not also employed in the other. Accordingly, the reduction of the laws of terrestrial and celestial motions to a single set of theoretical principles has for its outcome simply the incorporation of two classes of qualitatively similar phenomena into a more inclusive class whose members are likewise qualitatively homogeneous. Because of this circumstance, the reduction again generates no special logical puzzles, although it did in point of fact produce a revolution in men's outlook upon the world.

In reductions of the sort so far mentioned, the laws of the secondary science employ no descriptive terms that are not also used with approximately the same meanings in the primary science. Reductions of this type can therefore be regarded as establishing deductive relations between two sets of statements that employ a homogeneous vocabulary. Since such "homogeneous" reductions are commonly accepted as phases in the normal development of a science and give rise to few misconceptions as to what a scientific theory achieves, we shall pay no further attention to them.

2. The situation is usually different in the case of a second type of reduction. Difficulties are frequently experienced in comprehending the import of a reduction as a consequence of which a set of distinctive

traits of some subject matter is assimilated to what is patently a set of quite dissimilar traits. In such cases, the distinctive traits that are the subject matter of the secondary science fall into the province of a theory that may have been initially designed for handling qualitatively different materials and that does not even include some of the characteristic descriptive terms of the secondary science in its own set of basic theoretical distinctions. The primary science thus seems to wipe out familiar distinctions as spurious, and appears to maintain that what are *prima facie* indisputably different traits of things are really identical. The acute sense of mystification that is thereby engendered is especially frequent when the secondary science deals with macroscopic phenomena, while the primary science postulates a microscopic constitution for those macroscopic processes. An example will show the sort of puzzle that is generated.

Most adults in our society know how to measure temperatures with an ordinary mercury thermometer. If provided with such an instrument, they know how to determine with reasonable accuracy the temperature of various bodies; and, in terms of operations that are performed with the instrument, they understand what is meant by such statements as that the temperature of a glass of milk is 10° C. A good fraction of these adults would doubtless be unable to explicate the meaning of the word 'temperature' to the satisfaction of someone schooled in thermodynamics; and these same adults would probably also be unable to state explicitly the tacit rules governing their use of the word. Nevertheless, most adults do know how to use the word, even if only within certain limited contexts.

Let us now assume that some person has come to understand what is meant by 'temperature' exclusively in terms of manipulating a mercury thermometer. If that individual were told that there is a substance which melts at a temperature of fifteen thousand degrees, he would probably be at a loss to make sense of this statement, and he might even claim that what has been told him is quite meaningless. In support of this claim he might maintain that, since a temperature can be assigned to bodies only on the basis of employing a mercury thermometer, and since such thermometers are vaporized when brought into the proximity of bodies whose temperatures (as specified by a mercury thermometer) are a little above 350° C, the phrase "temperature of fifteen thousand degrees" has no defined sense and is therefore meaningless. However, his puzzlement over the information given him would be quickly removed by a little study of elementary physics. For he would then discover that the word 'temperature' is associated in physics with a more embracing set of rules of usage than the rules that controlled his own use of the word. In particular, he would learn that laboratory scientists employ the word to refer to a certain state of physical bodies, and that variations in this state are often manifested in other ways than by the volume expansion of mercury—for example, in changes in the electrical resistance of a body, or in the generation of electrical currents under specified conditions. Accordingly, once the laws are explained that formulate the relations between the behaviors of instruments such as thermocouples, which are sometimes used to record changes in the physical state of bodies called their 'temperature,' the person understands how the word can be meaningfully employed in situations other than those in which a mercury thermometer can be used. The enlargement of the word's range of application then appears no more puzzling or mysterious than does the extension of the word 'length,' from its primitive meaning as fixed by using the human foot for determining lengths, to contexts in which a standard metal bar replaces the human organism as a measuring instrument.

Suppose, however, that the layman for whom 'temperature' thus acquires a more generalized meaning now pursues his study of physics into the kinetic theory of gases. Here he discovers that the temperature of a gas is the mean kinetic energy of the molecules which by hypothesis constitute the gas. This new information may then generate a fresh perplexity, and indeed in an acute form. On the one hand, the layman has not forgotten his earlier lesson, according to which the temperature of a body is specified in terms of various overtly performed instrumental operations. But on the other hand, he is also assured by some authorities he now consults that the individual molecules of a gas cannot be said to possess a temperature, and that the meaning of the word is identical "by definition" with the meaning of 'the mean kinetic energy of molecules.'[1] Confronted by such apparently conflicting ideas, he may therefore find a host of typically "philosophical" questions both relevant and inescapable.

If the meaning of 'temperature' is indeed the same as that of 'the mean kinetic energy of molecules,' what is the plain man in the street talking about when he says that milk has a temperature of 10° C? Most consumers of milk who might make such statements are surely not asserting anything about the energies of molecules; for even though they understand and know how to use such statements, they are generally uninstructed in physics, and know nothing about the molecular composition of milk. Accordingly, when the man in the street learns about molecules in milk, he may come to believe that he is confronted with a serious issue as to what is genuine "reality" and what is only "appearance." He may then perhaps be persuaded by a traditional philosophical argu-

[1] Cf. Bernhard Bavink, *The Anatomy of Science*, London, 1932, p. 99.

ment that the familiar distinctions between hot and cold (indeed, even the distinctions between various temperatures of bodies as specified in terms of instrumental operations), refer to matters that are "subjective" manifestations of an underlying but mysterious physical reality, a reality which cannot properly be said to possess temperatures in the common-sense meaning of the word. Or he may accept the view, supported by a different mode of reasoning, that it is temperature as defined by procedures involving the use of thermometers and other such instruments which is the genuine reality, and that the molecular energies in terms of which the kinetic theory of matter "defines" temperature are just a fiction. Alternatively, if the layman adopts a somewhat more sophisticated line of thought, he may perhaps come to regard temperature as an "emergent" trait, manifested at certain "higher levels" of the organization of nature but not at the "lower levels" of physical reality; and he may then question whether the kinetic theory, which ostensibly is concerned only with those lower levels, does after all "really explain" the occurrence of emergent traits such as temperature.

Perplexities of this sort are frequently generated by reductions of the type of which the above example is an instance. In such reductions, the subject matter of the primary science appears to be qualitatively discontinuous with the materials studied by the secondary science. Put somewhat more precisely, in reductions of this type the secondary science employs in its formulations of laws and theories a number of distinctive descriptive predicates that are not included in the basic theoretical terms or in the associated rules of correspondence of the primary science. Reductions of the first or "homogeneous" type can be regarded as a special case of reductions of the second or "heterogeneous" type. But it is with reductions of the second type that we shall be concerned in what follows.

3. To fix our ideas, let us consider a definite example of a reduction of this variety. The incorporation of thermodynamics within mechanics—more exactly, within statistical mechanics and the kinetic theory of matter—is a classic and generally familiar instance of such a reduction. We shall therefore outline one small fragment of the argument by which the reduction is effected, on the assumption that this part of the argument is sufficiently representative of reductions of this type to serve as a basis for a generalized discussion of the logic of reduction in theoretical science.

Let us first briefly recall some historical facts. The study of thermal phenomena goes back in modern times to Galileo and his circle. During the subsequent three centuries a large number of laws were established dealing with special phases of the thermal behavior of bodies; and it was eventually shown with the help of a small number of general principles that these laws have certain systematic interrelations. Thermodynamics, as this science came to be called, uses concepts, distinctions, and general laws that are also employed in mechanics; for example, it makes free use of the notions of volume, weight, and pressure and of laws such as Hooke's law and the laws of the lever. In addition, however, thermodynamics employs a number of distinctive notions such as temperature, heat, and entropy, as well as general assumptions that are not corollaries to the fundamental principles of mechanics. Accordingly, though many laws of mechanics are constantly used in the explorations and explanations of thermal phenomena, thermodynamics was regarded for a long time as a special discipline, plainly distinguishable from mechanics and not simply a chapter in the latter. Indeed, thermodynamics is still usually expounded as a relatively autonomous physical theory; and its concepts, principles, and laws can be understood and verified without introducing any reference to some postulated microscopic structure of thermal systems and without assuming that thermodynamics can be reduced to some other theory such as mechanics. However, experimental work early in the nineteenth century on the mechanical equivalent of heat stimulated theoretical inquiry to find a more intimate connection between thermal and mechanical phenomena than the customary formulation of thermal laws seems to assert. Bernoulli's earlier attempts in this direction were revived, and Maxwell and Boltzmann were able to give a more satisfactory derivation of the Boyle-Charles' law from assumptions, statable in terms of the fundamental notions of mechanics, concerning the molecular constitution of ideal gases. Other thermal laws were similarly derived; and Boltzmann was able to interpret the entropy principle—perhaps the most characteristic assumption of thermodynamics and one which appears to differentiate the latter from mechanics most definitely—as an expression of the statistical regularity that characterizes the aggregate mechanical behavior of molecules. In consequence, thermodynamics was held to have lost its autonomy with respect to mechanics, and to have been reduced to the latter branch of physics.

Just how is this reduction effected? By what reasoning is it apparently possible to derive statements containing such terms as 'temperature,' 'heat,' and 'entropy' from a set of theoretical assumptions in which those terms do not appear? It is not possible to exhibit the complete argument without reproducing a treatise on the subject. Let us therefore fix our attention on but a small part of the complicated analysis, the derivation of the Boyle-Charles' law for ideal gases from the assumptions of the kinetic theory of matter.

If we suppress most of the details that do not contribute to the clarification of the main issue, a simplified form of the derivation is in outline as follows. Suppose that an ideal gas occupies a container whose volume is V. The gas is assumed to be composed of a large number of perfectly elastic spherical molecules possessing equal masses and volumes but with dimensions that are negligible when compared with the average distances between them. The molecules are further assumed to be in constant relative motions, subject only to forces of impact between themselves and the perfectly elastic walls of the container. Thus the molecules within their container constitute by postulation an isolated or conservative system, and the molecular motions are analyzable in terms of the principles of Newtonian mechanics. The problem now is to calculate the relation of other features of their motion to the pressure (or force per unit area) exerted by the molecules on the walls of the container because of their constant bombardments.

However, since the instantaneous coordinates of state of the individual molecules are not actually ascertainable, the usual mathematical procedures of classical mechanics cannot be applied. In order to make headway, a further assumption must be introduced—a statistical one concerning the positions and momenta of the molecules. This statistical assumption takes the following form: Let the volume V of the gas be subdivided into a very large number of smaller volumes, whose dimensions are equal among themselves and yet large when compared with the diameters of the molecules; and also divide the maximum range of the velocities that the molecules may possess into a large number of equal intervals. Now associate with each small volume all possible velocity-intervals, and call each complex obtained by associating a volume with a velocity-interval a "phase-cell." The statistical assumption then is that the probability of a molecule's occupying an assigned phase-cell is the same for all molecules and is equal to the probability of a molecule's occupying any other phase-cell, and that (subject to certain qualifications involving among other things the total energy of the system) the probability that one molecule occupies a phase-cell is independent of the occupation of that cell by any other molecule.

If in addition to these various assumptions it is stipulated that the pressure p exerted at any instant by the molecules on the walls of the container is the average of the instantaneous momenta transferred from the molecules to the walls, it is possible to deduce that the pressure p is related in a very definite way to the mean kinetic energy E of the molecules, and that in fact $p = 2E/3V$, or $pV = 2E/3$. But a comparison of this equation with the Boyle-Charles' law (according to which $pV = kT$, where k is a constant for a given mass of gas, and T its absolute temperature) suggests that the law could be deduced from the assumptions mentioned *if* the temperature were in some way related to the mean kinetic energy of the molecular motions. Let us therefore introduce the postulate that $2E/3 = kT$, that is, that the absolute temperature of an ideal gas is proportional to the mean kinetic energy of the molecules assumed to constitute it. Just what the character of this postulate is we shall for the moment not inquire. But our final result is that the Boyle-Charles' law is a logical consequence of the principles of mechanics, when these are supplemented by a hypothesis about the molecular constitution of a gas, a statistical assumption concerning the motions of the molecules, and a postulate connecting the (experimental) notion of temperature with the mean kinetic energy of the molecules.[2]

II. *Formal Conditions for Reduction*

Although the derivation of the Boyle-Charles' law from the kinetic theory of gases has only been sketched, the outline can nevertheless serve as a basis for stating the general conditions that must be satisfied if one science is to be reduced to another. It is convenient to divide the discussion into two parts, the first dealing with matters that are primarily of a formal nature, the second with questions of a factual or empirical character. We consider the formal matters first.

1. It is an obvious requirement that the axioms, special hypotheses, and experimental laws of the sciences involved in a reduction must be available as explicitly formulated statements, whose various constituent terms have meanings unambiguously fixed by codified rules of usage or by established procedures appropriate to each discipline. To the extent that this elementary requirement is not satisfied, it is hardly possible to decide with assurance whether one science (or branch of science) has in fact been reduced to another. It must be acknowledged, moreover, that in few if any of the various scientific disciplines in active development is this requirement of maximum explicitness fully realized, since in the normal practice of science it is rarely necessary to spell out in detail all the assumptions that may be involved in attacking a concrete problem. This requirement of explicitness is thus an ideal demand, rather than a description of the actual state of affairs that obtains at a given time. Nevertheless, the statements within each specialized discipline can be classified into distinct groups, based on the logical roles of the statements in the discipline. The following schematic catalogue of such groups of statements is not intended to be exhaustive, but to list the more important types of statements that are relevant to the present discussion.

[2] For a detailed exposition of the deduction, see, for example, James Rice, *Introduction to Statistical Mechanics*, New York, 1930, Chap. 4, or J. H. Jeans, *The Dynamical Theory of Gases*, Cambridge, England, 1925, Chap. 6.

a. In the highly developed science S (such as mechanics, electrodynamics, or thermodynamics) there is a class *T* of statements consisting of the fundamental theoretical postulates of the discipline. These postulates appear as premises (or partial premises) in all deductions within S. They are not derived from other assumptions in a given codification of the science, although in an alternative exposition of S a different set of logically primitive statements may be employed. Since *T* is adopted in order to account for, and to direct further inquiry into, experimental laws and observable events, there will also be a class *R* of coordinating definitions (or rules of correspondence) for a sufficient number of theoretical notions occurring in *T* or in statements formally derivable from those in *T*. Moreover, *T* will presumably satisfy the usual requirements for an adequate scientific theory. In particular, *T* will be capable of explaining systematically a large class of experimental laws belonging to S; it will not contain any assumptions whose inclusion does not significantly augment the explanatory power of *T* but serves merely to account for perhaps only one or two experimental laws; it will be "compendent" (in the sense that any pair of postulates in *T* will have at least one theoretical term in common); and the postulates in *T* will be "simple" and not too numerous. As noted in Chapter 6, it is sometimes convenient to use the assumptions *T* not as premises but as leading principles or as methodological rules of analysis. However, the issues that arise from stressing the role of theories as guiding principles rather than as premises have already been discussed, and those issues are in any case of no moment in the present context.

It is often possible to establish a hierarchy among the statements of *T* in respect to their generality (in the sense of "generality" examined in Chapter 3). When this can be done it is then useful to distinguish the subclass T_1, containing the most general theoretical assumptions in *T*, from the remaining subclass T_2, of more specialized ones. The most general postulates T_1 normally have a scope of application that is more inclusive than the scope of the theory *T* taken as a whole. Accordingly, the postulates T_1 are comprehensive postulates of which *T* is but a special case, while the assumptions T_2 are hypotheses concerning some *special type* of physical systems. For example, the most general theoretical assumptions in the kinetic theory of gases are the Newtonian axioms of motion, so that they belong to T_1; and their scope is clearly more embracing than is the scope of the kinetic theory. On the other hand, the postulate that every gas is a system of perfectly elastic molecules whose dimensions are negligible, or the postulate that all the molecules have the same probability of occupying a given phase-cell, are less general than the Newtonian axioms, and belong to T_2. The assumptions T_2 can thus be regarded as variable supplements to those in T_1, for they can be varied without altering the content of those in T_1, since the latter are applied to different types of systems. For example, the Newtonian axioms are supplemented by distinctive assumptions concerning the molecular structures of gases, liquids, and solids, when these axioms are used in theories about the properties of different states of aggregation of matter. Again, although the kinetic theory of gases retains the fundamental assumptions of Newtonian mechanics when it deals with various types of gases, the theory does not always postulate that gas molecules have negligible dimensions; moreover, the forces assumed by the theory to be acting between the molecules depend on whether or not the gas is far removed from its point of liquefaction.

Although it may not always be possible to distinguish sharply between the more general postulates T_1 of a theory and the less general variable supplements to them, some such distinction is commonly recognized. Thus, despite the fact that the primary science to which thermodynamics has been reduced contains other postulates than those of classical mechanics, thermodynamics is often said (even if only loosely) to be reducible to *mechanics*, presumably because the Newtonian axioms of motion are the most general assumptions of the kinetic theory of gases, so that they formulate the basic framework of ideas within which the special conclusions of the theory are embedded. Moreover, were the kinetic theory of gases able to account for some of the experimental laws of thermodynamics only by modifying one or more of its less general assumptions T_2, it is unlikely that anyone would therefore dispute the reducibility of thermodynamics to mechanics, provided that the principles of mechanics are retained as the most general explanatory premises of the revised theory.

b. A science S possessing a fundamental theory T will also have a class of theorems that are logical consequences of T. Some of the theorems will be formally derivable exclusively from T (indeed, often from the most general postulates T_1) without any help from the correspondence rules R, while others can be obtained only by using R as well. For example, a familiar theorem of the first kind in planetary theory is that, if a point-mass is moving under the action of a single central force, its orbit is a conic section; a theorem of the second kind is that, if a planet is moving under the action of the sun's gravitational force alone, its areal velocity is constant.

But whether or not S has a comprehensive theory, it will in general contain a class L of experimental laws that are conventionally regarded as falling into the special province of S. Thus, the various laws dealing with the reflection, refraction, and diffraction of light constitute part of the experimental content of the science of optics. Although at any given stage of development of S the class of its experimental laws L is in principle unambiguously determinable, this class is frequently augmented (and sometimes even diminished) with the progress of inquiry. Nor is there a permanently fixed demarcation between the experimental laws L that are grouped together as belonging to one branch of science S and the laws that are considered to fall into a different branch. Thus, it was not always understood that electrical and magnetic phenomena are intimately related; and in older books on physics, though not in most of the recent ones, experimental laws about *prima facie* different phenomena are classified as belonging to distinct departments of experimental inquiry. Indeed, the limits assumed for the domain of a given science, and the rationale operative in classifying experimental laws under different scientific disciplines, are often based on the explanatory scope of currently held theories.

c. Every positive science contains a large class of singular statements that either formulate the outcome of observations on the subject matter regarded as the province of the science or describe the overt procedures instituted in conducting some actual inquiry within that discipline. We shall call such singular statements "observation statements," but with the understanding that in using this label we are not committed to any special psychological or philosophical theory as to what are the "real" data of observation. In particular, observation statements are not to be identified with statements about "sense data" sometimes alleged to be exclusive objects of "direct experience." Thus, 'There was a total eclipse of the sun at Sabral in North Brazil on May 29, 1919,' and 'The switch was turned on yesterday in my office when the temperature of the room dropped to 50° F,' both count as observation statements in the present use of this designation. Observation statements may on occasion formulate initial and boundary conditions for a theory or law; they may also be employed to confirm or refute theories and laws.

d. Many observation statements of a given science S describe the arrangement and behavior of apparatus required for conducting experiments in S or for testing various assumptions adopted in S. Accordingly, the assertion of such observation statements may tacitly involve the use of laws concerning characteristics of different sorts of instruments; some of these laws may not fall into the generally acknowledged province of S and may not be explained by any theory of S. For example, photographic equipment attached to telescopes is commonly employed in testing Newtonian gravitational theory, so that the construction of such apparatus, as well as the interpretation of data obtained with its help, takes for granted theories and experimental laws both of optics and chemistry. However, the general assumptions thus taken for granted do not belong to the science of mechanics; and Newtonian gravitational theory does not pretend to explain or to warrant optical and chemical laws. When cameras and telescopes are used in inquiries into mechanical phenomena, distinctions and laws are therefore "borrowed" from other special disciplines. We shall refer to such laws, which are *used* in a science S but are not established or explained within S itself, as "borrowed laws" of S.

Most sciences will also contain statements that are certifiable as logically true, such as those of logic and mathematics. Even if we ignore these, we have identified four major classes of statements that may occur in a science S, whether or not any degree of autonomy is claimed for it relative to other special disciplines: (a) the theoretical postulates of S, the theorems derivable from them, and the coordinating definitions associated with theoretical notions in the postulates or theorems; (b) the experimental laws of S; (c) the observation statements of S; and (d) the borrowed laws of S.

2. We come to the second formal point. Every statement of a science S can be analyzed as a linguistic structure, compounded out of more elementary expressions in accordance with tacit or explicit rules of construction. It will be assumed that, though these elementary expressions may be vague in varying degrees, they are employed unambiguously in S, with meanings fixed either by habitual usage or explicitly formulated rules. Some of the expressions will be locutions of formal logic, arithmetic, and other branches of mathematical analysis. We shall, however, be primarily concerned with the so-called "descriptive expressions," signifying what are generally regarded as "empirical" objects, traits, relations, or processes, rather than purely formal or logical entities. Although there are difficulties in developing a precise distinction between logical and descriptive expressions, these difficulties do not impinge upon the present discussion. Let us in any case consider the class *D* of descriptive expressions in S that do not occur in borrowed laws of S.

Many of the descriptive expressions of a science are simply taken over from the language of ordinary affairs, and retain their everyday meanings. This is frequently true for expressions occurring in observation statements, since a large fraction of the overt procedures employed even in carefully devised laboratory experiments can be described in the language of gross experience. On the other hand, other descriptive expressions may be specific to a given science; they may have a use restricted to highly specialized technical contexts; and the meanings assigned to them in that science may even preclude their being employed to describe matters identifiable either by direct or indirect observation. Descriptive expressions of this latter sort occur typically in the theoretical assumptions of a science.

It is often possible to explicate the meaning of an expression in D with the aid of other expressions in D supplemented by logical ones. Such explications can sometimes be supplied in the form of conventional *explicit* definitions, though usually more complicated techniques for fixing the meanings of terms are required. But whatever formal techniques of explication may be used, let us call the set of expressions in D which, with the help of purely logical locutions, suffice to explicate the meanings of all other expressions in D, the "primitive expressions" of S. There will always be at least one set P of primitive expressions, since, in the least favorable cases, when no descriptive expression can be explicated in terms of others, the set P will be identical with the class D. On the other hand, there may be more than one such set P, for, as is well known, expressions that are primitive in one context of analysis may lose their primitive status in another context; but this possibility does not affect the present discussion.

However, if S has a comprehensive theory as well as observation statements and experimental laws, the explication of an expression may proceed in either one of two directions that must be noted, since in general each direction involves the use of a distinctive set of primitives.

a. Let us designate as "observation expressions" those expressions in D that refer to things, properties, relations, and processes capable of being observed. The distinction between observation expressions and other descriptive ones is admittedly vague, especially since different degrees of stringency may be used in different contexts in deciding what matters are to count as observable ones. But, despite its vagueness, the distinction is useful and is unavoidable in both scientific inquiry and everyday practice. In any event, many explications aim at specifying the meanings of descriptive expressions in terms of observable ones. The program (advocated by Peirce and Bridgman among others) of fixing the meanings of terms by giving what are currently known as "operational definitions" for them appears to have explications of this sort for its objective. Let us call the set P_1 of observation expressions needed for explicating in this manner the maximum number of expressions in D, the "observation primitives" of S. For example, the meaning of 'temperature' is frequently explained in physics in terms of the volume expansions of liquids and gases or in terms of other observable behaviors of bodies; in such cases the explication of 'temperature' is given by way of observable primitives.

b. Let us suppose that S has a theory capable of explaining all the experimental laws of the science; and let us designate as the "theoretical expressions" of S the descriptive expressions employed in the theoretical postulates (exclusive of the coordinating definitions) and the theorems formally derivable from them. Many explications aim at specifying the meanings of expressions by way of theoretical ones; and we shall call the set P_2 of theoretical expressions needed for explicating in this way the maximum number of expressions in D, the "theoretical primitives" of S. For example, the meaning of 'temperature' is given a theoretical explication in the science of heat with the help of statements describing the Carnot cycle of heat transformations, and therefore in terms of such theoretical primitives as 'perfect nonconductors,' 'infinite heat reservoirs,' and 'infinitely slow volume expansions.'

As we have seen in Chapter 6, the question whether theoretical expressions are explicitly definable in terms of observable ones has been much debated. If theoretical expressions were always so definable, they could be eliminated in favor of observable ones, so that the distinction would have little point. However, although a negative answer to the question has not been demonstrably established, all the available evidence supports that answer. Indeed, there are good reasons for maintaining the stronger claim that theoretical expressions cannot in general be adequately explicated with the help of observation ones alone, even when forms of explication other than explicit definitions are employed. It is not necessary to adopt a position on these questions for the purposes of the present discussion. We must nevertheless not assume as a matter of course that the set of observation primitives P_1 is sufficient to explicate all the descriptive expressions D; and we must allow for the possibility that the class P of primitive expressions of S does not in general coincide with the class P_1. Accordingly, although 'temperature' is explicated in the science of heat both in terms of theoretical and of observation primitives, it does not follow that the word understood in the sense of the first explication is synonymous with 'temperature' construed in the sense of the second.

3. We can now turn to the third formal consideration on reduction. The primary and secondary sciences involved in a reduction generally have in common a large number of expressions (including statements) that are associated with the same meanings in both sciences. Statements certifiable in formal logic and mathematics are obvious illustrations of such common expressions, but there usually are many other descriptive ones as well. For example, many laws belonging to the science of mechanics, such as Hooke's law or the laws of the lever, also appear in the science of heat, if only as borrowed laws; and the latter science employs in its own experimental laws such expressions as 'volume,' 'pressure,' and 'work' in senses that coincide with the meanings of these words in mechanics. On the other hand, before its reduction the secondary sci-

ence generally uses expressions and asserts experimental laws formulated with their help which do not occur in the primary science, except possibly in the latter's classes of observation statements and borrowed laws. For example, the science of mechanics in its classical form does not count the Boyle-Charles' law as one of its experimental laws; nor does the term 'temperature' occur in the theoretical assumptions of mechanics, though the word may sometimes be employed in its experimental inquiries to describe the circumstances under which some law of the science is being used.

It is, however, of utmost importance to note that expressions belonging to a science possess meanings that are fixed by its *own* procedures of explication. In particular, expressions distinctive of a given science (such as the word 'temperature' as employed in the science of heat) are intelligible in terms of the rules or habits of usage of that branch of inquiry; and when those expressions are used in that branch of study, they must be understood in the senses associated with them in that branch, whether or not the science has been reduced to some other discipline. Sometimes, to be sure, the meaning of an expression in a science can be explicated with the help of the primitives (whether theoretical or observational) of some other science. For example, there are firm grounds for the assumption that the word 'pressure' as understood in thermodynamics is synonymous with the term 'pressure' as explicated by way of the theoretical primitives of mechanics. It nevertheless does not follow that in general every expression employed in a given science, in the sense specified by its own distinctive rules or procedures, is explicable in terms of the primitives of some other discipline.

With these preliminaries out of the way, we must now state the formal requirements that must be satisfied for the reduction of one science to another. As has already been indicated in this chapter, a reduction is effected when the experimental laws of the secondary science (and if it has an adequate theory, its theory as well) are shown to be the logical consequences of the theoretical assumptions (inclusive of the coördinating definitions) of the primary science. It should be observed that we are not stipulating that the borrowed laws of the secondary science must also be derivable from the theory of the primary science. However, if the laws of the secondary science contain terms that do not occur in the theoretical assumptions of the primary discipline (and this is the type of reduction to which we agreed earlier to confine the discussion), the logical derivation of the former from the latter is *prima facie* impossible. The claim that the derivation is impossible is based on the familiar logical canon that, save for some essentially irrelevant exceptions, no term can appear in the conclusion of a formal demonstration unless the term also appears in the premises.[3] Accordingly, when the laws of the secondary science do contain some term '*A*' that is absent from the theoretical assumptions of the primary science, there are two necessary formal conditions for the reduction of the former to the latter: (1) Assumptions of some kind must be introduced which postulate suit-

[3] Possible objections to this logical canon are based for the most part on the fact that, in view of some theorems in modern formal logic, a valid deductive argument can have a conclusion containing terms not occurring in the premises.

There are at least two laws in the sentential calculus (or logic of unanalyzed propositions) that permit the deduction of such conclusions. According to one of them, any statement of the form 'If S_1, then S_1 or S_2,' where S_1 and S_2 are any statements, is logically true, so that S_1 or S_2 is derivable from S_1. But since S_2 can be chosen arbitrarily, S_1 or S_2 can be made to contain terms not occurring in S_1. According to a second logical law, any statement of the form 'S_1, if and only if S_1 and (S_2 or not-S_2)' is logically true; hence 'S_1 and (S_2 or not-S_2)' is derivable from S_1, with the same general outcome as in the first case. However, it is clear that neither type of deductive step can yield the Boyle-Charles' law from the kinetic theory of gases. If it could (for example, by way of substituting this law for S_2 in the first of the two logical laws mentioned), then, since S_2 is entirely arbitrary, the deduction would also yield the contradictory of this law; and this cannot happen, unless the kinetic theory itself is self-contradictory. This argument is quite general and applies to other examples of reduction. Accordingly, insofar as reductions make use only of the logical laws of the sentential calculus in deducing statements of the secondary science from the theory of the primary science, it is sufficient to meet the objection to the logical canon mentioned in the text by emending the latter to read: In a valid deduction no term appears in the conclusion that does not occur in the premises, unless a term enters into the conclusion via logical laws of the sentential calculus, which permit the introduction of any *arbitrary* term into the conclusion.

However, there are other logical laws, developed in other parts of formal logic, that also sanction conclusions with terms not in the premises. Substitution for variables expressing universality is a familiar type of such inference. For example, although the premise "For any x, if x is a planet then x shines by reflected light" does not contain the term 'Mars,' the statement "If Mars is a planet, then Mars shines by reflected light" can be validly deduced from it. Another type of such inference is illustrated by the derivation from "All men are mortal" of the conclusion "All hungry men are hungry mortals." Nevertheless, an examination of the derivation of the Boyle-Charles' law reveals that the term 'temperature,' contained in this law but not in the kinetic theory, is not introduced into the derivation by way of any such universally valid deductive steps; and an argument, analogous to the one presented in the preceding paragraph of this note for the case of deductions in the sentential calculus, can be constructed to show that this must also be the case in the deduction of other laws, containing distinctive terms, of a secondary science that is reducible to some primary one. Accordingly, these various exceptions to the logical canon of the text can be ignored as not relevant to the matters under discussion.

A different objection to this canon is that, formal logic aside, we often do recognize arguments as valid even though they ostensibly violate the canon. Thus, 'John is a cousin of Mary' is said to follow from 'The uncle of John is the father of Mary,' and 'Smith's shirt is colored' is said to follow from 'Smith's shirt is red,' despite the fact that a term appears in each of the conclusions that is absent from the corresponding premise. However, these examples and others like them are essentially enthymematic inferences, with a tacit assumption either in the form of an explicit definition or some other kind of a priori statement. When these suppressed assumptions are made explicit, the examples no longer appear to be exceptions to the logical canon under examination.

able relations between whatever is signified by '*A*' and traits represented by theoretical terms already present in the primary science. The nature of such assumptions remains to be examined; but without prejudging the outcome of further discussion, it will be convenient to refer to this condition as the "condition of connectability." (2) With the help of these additional assumptions, all the laws of the secondary science, including those containing the term '*A*,' must be logically derivable from the theoretical premises and their associated coordinating definitions in the primary discipline. Let us call this the "condition of derivability." [4]

There appear to be just three possibilities as to the nature of the linkages postulated by these additional assumptions: (1) The first is that the links are *logical connections* between established meanings of expressions. The assumptions then assert '*A*' to be logically related (presumably by synonymy or by some form of one-way analytical entailment) to a theoretical expression '*B*' in the primary science. On this alternative, the meaning of '*A*' as fixed by the rules or habits of usage of the secondary science must be explicable in terms of the established meanings of theoretical primitives in the primary discipline. (2) The second possibility is that the linkages are *conventions*, created by deliberate fiat. The assumptions are then coordinating definitions, which institute a correspondence between '*A*' and a certain theoretical primitive, or some construct formed out of the theoretical primitives, of the primary science. On this alternative, unlike the preceding one, the meaning of '*A*' is not being explicated or analyzed in terms of the meanings of theoretical primitives. On the contrary, if '*A*' is an observation term of the secondary science, the assumptions in this case *assign* an experimental significance to a certain theoretical expression of the primary science, consistent with other such assignments that may have been previously made. (3) The third possibility is that the linkages are *factual* or *material*. The assumptions then are physical hypotheses, asserting that the occurrence of the state of affairs signified by a certain theoretical expression '*B*' in the primary science is a sufficient (or necessary and sufficient) condition for the state of affairs designated by '*A*.' It will be evident that in this case independent evidence must in principle be

[4] The condition of connectability requires that *theoretical* terms of the primary science appear in the statement of these additional assumptions. It would not suffice, for example, if these assumptions formulated an explication of '*A*' by way of observation primitives of the primary science, even if the theoretical primitives could also be explicated by way of the observation primitives. For it would not thereby follow that '*A*' could be explicated by way of the theoretical primitives. Thus, although 'uncle' and 'grandfather' are each definable in terms of 'male' and 'parent,' 'uncle' is not definable in terms of 'grandfather.' In consequence, the additional assumption would not contribute toward the fulfillment of the condition of derivability.

obtainable for the occurrence of each of the two states of affairs, so that the expressions designating the two states must have identifiably different meanings. On this alternative, therefore, the meaning of '*A*' is not related analytically to the meaning of '*B*.' Accordingly, the additional assumptions cannot be certified as true by logical analysis alone, and the hypothesis they formulate must be supported by empirical evidence.[5]

In the light of this discussion, let us now examine the derivation of the Boyle-Charles' law from the kinetic theory of gases. For the sake of simplicity let us also assume that the word 'temperature' is the only term in this law that does not occur in the postulates of that theory. However, as already noted, the deduction of the law from the theory depends on the additional postulate that the temperature of a gas is proportional to the mean kinetic energy of its molecules. Our problem is to decide on the status of this postulate and to determine which if any of the three types of linkage we have been discussing is asserted by the postulate.

For reasons mentioned in the first section of the present chapter, it is safe to conclude that 'temperature,' in the sense the word is employed in classical thermodynamics, is not synonymous with 'mean kinetic energy of molecules,' nor can its meaning be extracted from the meaning of the latter expression. Certainly no standard exposition of the kinetic theory of gases pretends to establish the postulate by analyzing the meanings of the terms occurring in it. The linkage stipulated by the postulate cannot therefore be plausibly regarded as a logical one.

But it is far more difficult to decide which of the remaining two types

[5] It follows that the condition of connectability is in general not sufficient for reduction and must be supplemented by the condition of derivability. Connectability would indeed assure derivability if, as has been rightly argued by John G. Kemeny and Paul Oppenheim ["On Reduction," *Philosophical Studies*, Vol. 7 (1956), p. 10], for every term '*A*' in the secondary science but not in the primary one there is a theoretical term '*B*' in the primary science such that *A* and *B* are linked by a biconditional: *A* if and only if *B*. If the linkage has this form, '*A*' can be replaced by '*B*' in any law *L* of the secondary science in which '*A*' occurs, and so yield a warranted theoretical postulate L'. If L' is not itself derivable from the available theory of the primary science, the theory need only be augmented by L' to become a modified theory, but nonetheless a theory of the primary science. In any event, *L* will be deducible from a theory of the primary science with the help of the biconditionals. However, the linkage between *A* and *B* is not necessarily biconditional in form, and may for example be only a one-way conditional: If *B*, then *A*. But in this eventuality '*A*' is not replaceable by '*B*,' and hence the secondary science will not in general be deducible from a theory of the primary discipline. Accordingly, even if we waive the question whether a reduction is satisfactory when achieved by augmenting the theory of the primary science by a new postulate L' which is empirically confirmed but may contribute next to nothing to the explanatory power of the initial theory, connectability does not in general suffice to assure derivability. On the other hand, the condition of derivability is both necessary and sufficient for reduction, since derivability obviously entails connectivity. The condition of connectability is nevertheless stated separately, because of its importance in the analysis of reduction.

of linkage is asserted by the postulate, for there are plausible reasons favoring each of these alternatives. The argument in support of the claim that the postulate is simply a coordinating definition is essentially as follows: The kinetic theory of gases cannot be put to experimental test, unless rules of correspondence first associate some of its theoretical notions with experimental control. For, although the temperature of a gas can be determined by familiar laboratory procedures, there is apparently no way of ascertaining the mean kinetic energy of the hypothetical gas molecules—unless, indeed, the temperature is stipulated by fiat to be a measure of this energy. Accordingly, the postulate can be nothing other than one of the correspondence rules which institute an association between theoretical and experimental concepts.[6] On the other hand, the claim that the postulate is a physical hypothesis is also not an unfounded one; and, indeed, it is in this fashion that the postulate is introduced in many technical presentations of the subject. The major reason advanced for this claim is that, although the postulate cannot be tested by direct measurements on the mean kinetic energy of gas molecules, the value of this energy can nevertheless be ascertained indirectly, by calculation from experimental data on gases other than data obtained by measuring temperatures. In consequence, it does seem possible to determine experimentally whether the temperature of a gas is proportional to the mean kinetic energy of its molecules.

Despite appearances to the contrary, these alternative claims and supporting reasons for them are not necessarily incompatible. Indeed, the alternatives illustrate what is by now a familiar point—that the cognitive status of an assumption often depends on the mode adopted for articulating a theory in a particular context. The reduction of thermodynamics to mechanics can undoubtedly be so expounded that the additional postulates about the proportionality of temperature to the mean kinetic energy of gas molecules institutes what is at first the sole link between the theoretical notions of the primary science and experimental concepts of the secondary one. In such a context of exposition, the postulate cannot be subjected to experimental test but functions as a coordinating definition. However, different modes of exposition are also possible, in which coordinating definitions are introduced for other pairs of theoretical and experimental concepts. For example, one theoretical notion can be made to correspond to the experimental idea of viscosity, and another can be associated with the experimental concept of heat flow. In consequence, since the mean kinetic energy of gas molecules is related, by virtue of the assumptions of the kinetic theory, to these other

[6] Cf. Norman R. Campbell, *Physics, the Elements*, Cambridge, England, 1920, pp. 126ff.

theoretical notions, a connection may thus be indirectly established between temperature and kinetic energy. Accordingly, in such a context of exposition, it would make good sense to ask whether the temperature of a gas is proportional to the value of the mean kinetic energy of the gas molecules, where this value is calculated in some indirect fashion from experimental data other than that obtained by measuring the temperature of the gas. In this case the postulate would have the status of a physical hypothesis.

It is therefore not possible to decide in general whether the postulate is a coordinating definition or a factual assumption, except in some given context in which the reduction of thermodynamics to mechanics is being developed. This circumstance does not, however, wipe out the distinction between rules of correspondence and material hypotheses, nor does it destroy the importance of the distinction. But in any event, the present discussion does not require that a decision be made between these alternative interpretations of the postulate. The essential point in this discussion is that in the reduction of thermodynamics to mechanics a postulate connecting temperature and mean kinetic energy of gas molecules must be introduced, and that this postulate cannot be warranted by simply explicating the meanings of the expressions contained in it.

One objection to this central contention must be briefly considered. The redefinition of expressions with the development of inquiry, so the objection notes, is a recurrent feature in this history of science. Accordingly, though it must be admitted that in an earlier use the word 'temperature' had a meaning specified exclusively by the rules and procedures of thermometry and classical thermodynamics, it is *now* so used that temperature is "identical by definition" with molecular energy. The deduction of the Boyle-Charles' law does not therefore require the introduction of a further postulate, whether in the form of a coordinating definition or a special empirical hypothesis, but simply makes use of this definitional identity. This objection illustrates the unwitting double talk into which it is so easy to fall. It is certainly possible to redefine the word 'temperature' so that it becomes synonymous with 'mean kinetic energy of molecules.' But it is equally certain that on this redefined usage the word has a different meaning from the one associated with it in the classical science of heat, and therefore a meaning different from the one associated with the word in the statement of the Boyle-Charles' law. However, if thermodynamics is to be reduced to mechanics, it is temperature in the sense of the term in the classical science of heat which must be asserted to be proportional to the mean kinetic energy of gas molecules. Accordingly, if the word 'temperature' is redefined as suggested by the objection, the hypothesis must be invoked that the state of bodies described as 'temperature' (in the classical thermodynamical sense) is also characterized by 'temperature' in the redefined sense of the term. *This* hypothesis, however, will then be one that does not hold as a matter of definition, and will not be one for which logical necessity can be rightly claimed. Unless the hypothesis is adopted, it is not the Boyle-Charles' law which can be derived from the assumptions of the kinetic theory of gases. What is derivable without the hypothesis is a sentence similar in syntactical structure to the standard formulation of that law, but possessing a sense that is unmistakably different from what the law asserts.

III. *Nonformal Conditions for Reduction*

We must now turn to features of reduction that are not primarily formal, though some of them have already been touched upon in passing.

1. The two formal conditions for reduction discussed in the previous section do not suffice to distinguish trivial from noteworthy scientific achievements. If the sole requirement for reduction were that the secondary science is logically deducible from arbitrarily chosen premises, the requirement could be satisfied with relatively little difficulty. In the history of significant reductions, however, the premises of the primary science are not *ad hoc* assumptions. Accordingly, although it would be a far too strong condition that the premises must be known to be true, it does seem reasonable to impose as a nonformal requirement that the theoretical assumptions of the primary science be supported by empirical evidence possessing some degree of probative force. The problems connected with the logic of weighing evidence are difficult and at many crucial points still unsettled. However, the issues raised by these problems are not exclusively relevant to the analysis of reduction; and, except for some brief comments especially pertinent to the reduction of thermodynamics to mechanics, we shall not at this place examine the notion of adequate evidential support.

The evidence for the several assumptions of the kinetic theory of gases comes from a variety of inquiries, only a fraction of which fall into the domain of thermodynamics. Thus, the hypothesis of the molecular constitution of matter was supported by quantitative relations exhibited in chemical interactions even before thermodynamics was reduced to mechanics; and it was also confirmed by a number of laws in molar physics not primarily about thermal properties of bodies. The adoption of this hypothesis for the new task of accounting for the thermal behavior of gases was therefore in line with the normal strategy of the science to exploit on a new front ideas and analogies found to be fruitful elsewhere. Similarly, the axioms of mechanics, constituting the most general parts of the premises in the primary science to which thermodynamics is reduced, are supported by evidence from many fields quite distinct from the study of gases. The assumption that these axioms also hold for the hypothetical molecular components of gases thus involved the extrapolation of a theory from domains in which it was already well confirmed into another domain postulated to be homogeneous in important respects with the former ones. But the point having greatest weight in this connection is that the combined assumptions of the primary science to which the science of heat was reduced have made it possible to incorporate into a unified system many apparently unrelated laws of the science of heat as well as of other parts of physics. A number of gas laws had of course been established before the reduction. However, some of these laws were only approximately valid for gases not satisfying certain narrowly restrictive conditions; and most of the laws, moreover, could be affirmed only as so many independent facts about gases. The reduction of thermodynamics to mechanics altered this state of affairs in significant ways. It paved the way for a reformulation of gas laws so as to bring them into accord with the behaviors of gases satisfying less restrictive conditions; it provided leads to the discovery of new laws; and it supplied a basis for exhibiting relations of systematic dependence among gas laws themselves, as well as between gas laws and laws about bodies in other states of aggregation.

This last point deserves a brief elaboration. If the Boyle-Charles' law were the sole experimental law deducible from the kinetic theory of gases, it is unlikely that this result would be counted by most physicists as weighty evidence for the theory. They would probably take the view that nothing of significance is achieved by the deduction of only this one law. For prior to its deduction, so they might maintain, this law was known to be in good agreement with the behavior of only "ideal" gases, that is, those at temperatures far above the points at which the gases liquefy; and by hypothesis, nothing further follows from the theory as to the behavior of gases at lower temperatures. Moreover, physicists would doubtless call attention to the telling point that even the deduction of this law can be effected only with the help of a special postulate connecting temperature with the energy of gas molecules—a postulate that, under the circumstances envisaged, has the status of an *ad hoc* assumption, supported by no evidence other than the evidence warranting the Boyle-Charles' law itself. In short, if this law were the sole experimental consequence of the kinetic theory, the latter would be dead wood from which only artificially suspended fruit could be gathered.

In actual fact, however, the reduction of thermodynamics to the kinetic theory of gases achieves much more than the deduction of the Boyle-Charles' law. There is available other evidence that counts heavily with most physicists as support for the theory and that removes from the special postulate connecting temperatures and molecular energy even the appearance of arbitrariness. Indeed, two related sets of considerations make the reduction a significant scientific accomplishment. One set consists of experimental laws, deduced from the theory, which have not been previously established or which are in better agreement with a wider range of facts than are laws previously accepted. For example, the Boyle-Charles' law holds only for ideal gases and is deducible from the kinetic theory when some of the less general assumptions of the kinetic theory have the limiting form corresponding to a gas being an ideal gas. However, these special assumptions can be replaced by others without modifying the fundamental ideas of the theory, and in particular by assumptions less simple than those introduced for ideal gases. Thus, instead of the stipulations with the aid of which the Boyle-Charles' law is derivable from the theory, we can assume that the dimensions of gas molecules are not negligible when compared to the mean distance between them, and that in addition to forces of impact there are also cohesive forces acting upon them. It is then possible to deduce from the theory employing these more complex special assumptions the van der Waals' law for gases, which formulates more adequately than does the Boyle-Charles' law the behavior of both ideal and nonideal gases. In general, therefore, for a reduction to mark a significant intellectual advance, it is not enough that previously established laws of the secondary science be represented within the theory of the primary discipline. The theory must also be fertile in usable suggestions for developing the secondary science, and must yield theorems referring to the latter's subject matter which augment or correct its currently accepted body of laws.

The second set of considerations in virtue of which the reduction of thermodynamics to mechanics is generally regarded as an important achievement consists of the intimate and frequently surprising relations of dependence that can thereby be shown to obtain between various experimental laws. An obvious type of such dependence is illustrated when laws, hitherto asserted on independent evidential grounds, are deducible from an integrated theory as a consequence of the reduction. Thus, both the second law of thermodynamics (according to which the entropy of a closed physical system never diminishes) as well as the Boyle-Charles' law are derivable from statistical mechanics, although in classical thermodynamics these laws are stated as independent primitive assumptions. In some ways a more impressive and subtler type of dependence is illustrated when some numerical constant appearing in different experimental laws of the secondary science is exhibited as a definite function of theoretical parameters in the primary discipline—an outcome that is particularly striking when congruous numerical values can be calculated for those parameters from experimental data obtained in independent lines of inquiry. Thus, one of the postulates of the kinetic theory is that, under standard conditions of temperature and pressure, equal volumes of a gas contain an equal number of molecules, irrespective of the chemical nature of the gas. The number of molecules in a liter of gas under standard conditions is thus the same for all gases, and is known as Avogadro's number. Moreover, a certain constant appearing in several gas laws (among others, in the Boyle-Charles' law and the laws of specific heats) can be shown to be a function of this number and other theoretical parameters. On the other hand, Avogadro's number can be calculated in alternative ways from experimental data gathered in different kinds of inquiry, e.g., from measurements in the study of thermal phenomena, of Brownian movements, or of crystal structure; and the values obtained for the number from each of these diverse sets of data are in good agreement with one another. Accordingly, apparently independent experimental laws (including thermal ones) are shown to involve a common invariant component, represented by a theoretical parameter that in turn becomes firmly tied to several kinds of experimental data. In consequence, the reduction of thermodynamics to kinetic theory not only supplies a unified explanation for the laws of the former discipline; it also integrates these laws so that directly relevant evidence for any one of them can serve as indirect evidence for the others, and so that the available evidence for any of the laws cumulatively supports various theoretical postulates of the primary science.

2. These general comments on the considerations that determine whether a reduction is a significant advance in the organization of knowledge or only a formal exercise, and on the character of the evidence that actually supports the kinetic theory, direct attention to an important feature of sciences in active development. As has already been suggested, different branches of science may sometimes be delimited on the basis of the theories used as explanatory premises and leading principles in their respective domains. Nevertheless, theories do not as a rule remain unaltered with the progress of inquiry; and the history of science provides many examples of special branches of knowledge becoming reorganized around new types of theory. Moreover, even if a discipline continues to retain the most general postulates of some theoretical system, the less general ones are often modified or are augmented by others as fresh problems arise.

Accordingly, the question whether a given science is reducible to another cannot in the abstract be usefully raised without reference to some particular stage of development of the two disciplines. Questions about reducibility can be profitably discussed only if they are made definite by specifying the established content at a given date of the sciences under consideration. Thus, no practicing physicist is likely to take seriously the claim that the contemporary science of nuclear physics is reducible to some variant of classical mechanics—even if the claim should be accompanied by a formal deduction of the laws of nuclear physics from admittedly purely mechanical assumptions—unless these assumptions are supported by adequate evidence available at the time the claim is made, and also possess at that time the heuristic advantages normally expected of the theory belonging to a proposed primary science. Again, it is one thing to say that thermodynamics is reducible to mechanics when the latter counts among its recognized postulates assumptions (including statistical ones) about molecules and their modes of action; it is quite a different thing to claim that thermodynamics is reducible to a science of mechanics that does not countenance such assumptions. In particular, though contemporary thermodynamics is undoubtedly reducible to a statistical mechanics postdating 1866 (the year in which Boltzmann succeeded in giving a statistical interpretation for the second law of thermodynamics with the help of certain statistical hypotheses), that secondary science is not reducible to the mechanics of 1700. Similarly, certain parts of nineteenth-century chemistry (and perhaps the whole of this science) is reducible to post-1925 physics, but not to the physics of a hundred years ago.

Moreover, the possibility should not be ignored that little if any new knowledge or increased power for significant research may actually be gained from reducing one science to another at certain periods of their development, however great may be the potential advantages of such reduction at some later time. Thus, a discipline may be at a stage of active growth in which the imperative task is to survey and classify the extensive and diversified materials of its domain. Attempts to reduce the discipline to another (perhaps theoretically more advanced) science, even if successful, may then divert needed energies from what are the crucial problems at this period of the discipline's expansion, without being compensated by effective guidance from the primary science in the conduct of further research. For example, at a time when the prime need of botany is to establish a systematic typology of existing plant life, the discipline may reap little advantage from adopting a physicochemical theory of living organisms. Again, although one science may be reducible to another, the secondary discipline may be progressively solving its own special class of problems with the help of a theory expressly devised for dealing with the subject matter of that discipline. As a basis for attacking these problems, this less inclusive theory may well be more satisfactory than the more general theory of the primary science—perhaps because the primary science requires the use of techniques too refined and cumbersome for the subjects under study in the secondary science, or because the initial conditions needed for applying it to these subjects are not available, or simply because its structure does not suggest fruitful analogies for handling these problems. For example, even if biology were reducible to the physics of current quantum mechanics, at the present stage of biological science the gene theory of heredity may be a more satisfactory instrument for exploring the problems of biological inheritance than would be the quantum theory. An integrated system of explanation by some inclusive theory of a primary science may be an eventually realizable intellectual ideal. But it does not follow that this ideal is best achieved by reducing one science to another with an admittedly comprehensive and powerful theory, if the secondary science at that stage of its development is not prepared to operate effectively with this theory.

Much controversy over the interrelations of the special sciences, and over the limits of the explanatory power of their theories, neglects these elementary considerations. The irreducibility of one science to another (for example, of biology to physics) is sometimes asserted absolutely, and without temporal qualifications. In any event, arguments for such claims often appear to forget that the sciences have a history, and that the reducibility (or irreducibility) of one science to another is contingent upon the specific theory employed by the latter discipline at some stated time. On the other hand, converse claims maintaining that some particular science is reducible to a favored discipline also do not always give sufficient heed to the fact that the sciences under consideration must be at appropriately mature levels of development if the reduction is to be of scientific importance. Such claims and counterclaims are perhaps most charitably construed as debates over what is the most promising direction systematic research should take at some given stage of a science. Thus biologists who insist on the "autonomy" of their science and who reject *in toto* so-called "mechanistic theories" of biological phenomena sometimes appear to adopt these positions because they believe that in the present state of physical and biological theory biology stands to gain more by carrying on its investigations in terms of distinctively biological categories than by abandoning them in favor of modes of analysis typical of modern physics. Analogously, mechanists in biology can often be understood as recommending the reduction of biology to physics, because in their view biological problems can now be handled more effectively within the framework of current physical theories than with the help of any purely biological ones. As we shall see in the following chapter, however, this is not the way that the issues are usually stated by those taking sides in such debates. On the contrary, largely because of a failure to note that claims concerning the reducibility or irreducibility of a science must be temporally qualified, questions that at bottom relate to the strategy of research, or to the logical relations between sciences as constituted at a certain time, are commonly discussed as if they were about some ultimate and immutable structure of the universe.

3. Throughout the present discussion stress has been placed on conceiving the reduction of one science to another as the deduction of one set of empirically confirmable *statements* from another such set. However, the issues of reduction are frequently discussed on the supposition that reduction is the derivation of the *properties* of one subject matter from the properties of another. Thus a contemporary writer maintains that psychology is demonstrably an autonomous discipline with respect to physics and physiology, because "a headache is not an arrangement or rearrangement of particles in one's cranium," and "our sensation of violet *is* not a change in the optic nerve." Accordingly, though the mind is said to be "connected mysteriously" with the physical processes, "it cannot be reduced to those processes, nor can it be explained by the laws of those processes." [7] Another recent writer, in presenting the case for the occurrence of "genuine novelties" in inorganic nature, declares that "it is an error to assume that *all* the properties of a compound can be deduced solely from the nature of its elements." In a similar vein, a third contemporary author asserts that the characteristic behavior of a chemical compound, such as water, "*could* not, even in theory, be deduced from the most complete knowledge of the behavior of its components, taken separately or in other combinations, and of their properties and arrangements in this whole." [8] We must now briefly indicate that the conception of reduction as the deduction of *properties* from other properties is potentially misleading and generates spurious problems.

The conception is misleading because it suggests that the question of whether one science is reducible to another is to be settled by inspecting the "properties" or alleged "natures" of things rather than by investigating the logical consequences of certain explicitly formulated *theories* (that is, systems of statements). For the conception ignores the crucial point that the "natures" of things, and in particular of the "elementary constituents" of things, are not accessible to direct inspection and that we cannot read off by simple inspection what it is they do or do not imply. Such "natures" must be stated as a theory and are not the objects of observation; and the range of the possible "natures" which

[7] Brand Blanshard, "Fact, Value and Science," in *Science and Man* (ed. by Ruth N. Anshen), New York, 1942, p. 203.
[8] C. D. Broad, *The Mind and Its Place in Nature*, London, 1925, p. 59.

chemical elements may possess is as varied as the different theories about atomic structures that we can devise. Just as the "fundamental nature" of electricity used to be stated by Maxwell's equations, so the fundamental nature of molecules and atoms must be stated as an explicitly articulated theory about them and their structures. Accordingly, the supposition that, in order to reduce one science to another, some properties must be deduced from certain other properties or "natures" converts what is eminently a logical and empirical question into a hopelessly irresolvable speculative one. For how can we discover the "essential natures" of the chemical elements (or of anything else) except by constructing theories which postulate definite characteristics for these elements, and then controlling the theories in the usual way by confronting consequences deduced from the theories with the outcome of appropriate experiments? And how can we know in advance that no such theory can ever be constructed which will permit the various laws of chemistry to be derived systematically from it?

Accordingly, whether a given set of "properties" or "behavioral traits" of macroscopic objects can be explained by, or reduced to, the "properties" or "behavioral traits" of atoms and molecules is a function of whatever theory is adopted for specifying the "natures" of these elements. The deduction of the "properties" studied by one science from the "properties" studied by another may be impossible if the latter science postulates these properties in terms of one theory, but the reduction may be quite feasible if a different set of theoretical postulates is adopted. For example, the deduction of the laws of chemistry (e.g., of the law that under certain conditions hydrogen and oxygen combine to form a stable compound commonly known as water, which in turn exhibits certain definite modes of behavior in the presence of other substances) from the physical theories of the atom accepted fifty years ago was rightly held to be impossible. But what was impossible relative to one theory need not be impossible relative to another physical theory. The reduction of various parts of chemistry to the quantum theory of atomic structure now appears to be making steady if slow headway; and only the stupendous mathematical difficulties involved in making the relevant deductions from the quantum theoretical assumptions seem to stand in the way of carrying the task much further along. Again, to repeat in the present context a point already made in another, if the "nature" of molecules is stipulated in terms of the theoretical primitives of classical statistical mechanics, the reduction of thermodynamics is possible only if an additional postulate is introduced that connects temperature and kinetic energy. However, the impossibility of the reduction without such special hypothesis follows from purely formal considerations, and not from some alleged ontological hiatus between the mechanical and the thermodynamical. Laplace was thus demonstrably in error when he believed that a Divine Intelligence could foretell the future in every detail, given the instantaneous positions and momenta of all material particles as well as the magnitudes and directions of the forces acting between them. At any rate, Laplace was in error if his Divine Intelligence is assumed to draw inferences in accordance with the canons of logic, and is therefore assumed to be incapable of the blunder of asserting a statement as a conclusion of an inference when the statement contains terms not occurring in the premises.

However this may be, the reduction of one science to a second—e.g., thermodynamics to statistical mechanics, or chemistry to contemporary physical theory—does not wipe out or transform into something insubstantial or "merely apparent" the distinctions and types of behavior which the secondary discipline recognizes. Thus, if and when the detailed physical, chemical, and physiological conditions for the occurrence of headaches are ascertained, headaches will not thereby be shown to be illusory. On the contrary, if in consequence of such discoveries a portion of psychology will be reduced to another science or to a combination of other sciences, all that will have happened is that an explanation will have been found for the occurrence of headaches. But the explanation that will thus become available will be of essentially the same sort as those obtainable in other areas of positive science. It will not establish a logically necessary connection between the occurrence of headaches and the occurrence of certain events or processes specified by physics, chemistry, and physiology. Nor will it consist in establishing the synonymy of the term 'headache' with some expression defined by means of the theoretical primitives of these disciplines. It will consist in stating the conditions, formulated by means of these primitives, under which, and as a matter of sheer contingent fact, a determinate psychological phenomenon takes place.

IV. *The Doctrine of Emergence*

The analysis of reduction is intimately relevant to a number of currently debated theses in general philosophy, especially the doctrine known as "emergent evolution" or "holism." Indeed, some results of that analysis have already been applied in the preceding section of this chapter to some of the issues raised by the doctrine of emergence. We shall now examine this doctrine more explicitly, in the light supplied by the discussion of reduction.

The doctrine of emergence is sometimes formulated as a thesis about the hierarchical organization of things and processes, and the consequent occurrence of properties at "higher" levels of organization which are not predictable from properties found at "lower" levels. On the other hand, the doctrine is sometimes stated as part of an evolutionary cosmogony, according to which the simpler properties and forms of organization already in existence make contributions to the "creative advance" of nature by giving birth to more complex and "irreducibly novel" traits and structures. In one of its forms, at any rate, emergent evolution is the thesis that the present variety of things in the universe is the outcome of a progressive development from a primitive stage of the cosmos containing only undifferentiated and isolated elements (such as electrons, protons, and the like), and that the future will continue to bring forth unpredictable novelties. This evolutionary version of the emergence doctrine is not entailed by the conception of emergence as irreducible hierarchical organization, and the two forms of the doctrine must be distinguished. We shall first consider emergence as a thesis about the nonpredictability of certain characteristics of things, and subsequently examine briefly emergence as a temporal, cosmogonic process.

1. Although emergence has been invoked as an explanatory category most frequently in connection with social, psychological, and biological phenomena, the notion can be formulated in a general way so as to apply to the inorganic as well. Thus, let O be some object that is constituted out of certain elements $a_1, \ldots, a_n$ standing to each other in some complex relation R; and suppose that O possesses a definite class of properties P, while the elements of O possess properties belonging to the classes $A_1, \ldots, A_n$ respectively. Although the elements are numerically distinct, they may not all be distinct in kind; moreover, they may enter into relations with one another (or with other elements not parts of O) that are different from R, to form complex wholes different from O. However, the occurrence of the elements $a_1, \ldots, a_n$ in the relation R is by hypothesis the necessary and sufficient condition for the occurrence of O characterized by the properties P.

Let us next assume what proponents of the doctrine of emergence call "complete knowledge" concerning the elements of O: we know *all* the properties the elements possess when they exist "in isolation" from one another; and we also know *all* the properties exhibited by complexes other than O that are formed when some or all of these elements stand to each other (or to additional elements) in relations other than R, as well as *all* the properties of the elements in these complexes. According to the doctrine of emergence, two cases must be distinguished. In the first case, it is possible to predict (that is, deduce) from such complete knowledge that, if the elements $a_1, \ldots, a_n$ occur in the relation R, then the object O will be formed and will possess the properties P. In the second case, there is at least one property P_e in the class P such that, despite complete knowledge of the elements, it is impossible to predict from this knowledge that, if the elements stand to each other in relation R, then an object O possessing P_e will be formed. In the latter case, the object O is an "emergent object" and P_e an "emergent property" of O.

It is this form of the emergence doctrine that underlies the passage from Broad cited in the preceding section of this chapter (page 364). Broad illustrates this version of emergence as follows:

> Oxygen has certain properties and Hydrogen has certain other properties. They combine to form water, and the proportions in which they do this is fixed. Nothing that we know about Oxygen itself or in its combination with anything but Hydrogen could give us the least reason to suppose that it could combine with Hydrogen at all. Nothing that we know about Hydrogen by itself or in its combination with anything but Oxygen, could give us the least reason to expect that it would combine with Oxygen at all. And most of the chemical and physical properties of water have no known connexion, either quantitative or qualitative, with those of Oxygen and Hydrogen. Here we have a clear instance where, so far as we can tell, the properties of a whole composed of two constituents could not have been predicted from a knowledge of those properties taken separately, or from this combined with a knowledge of the properties of other wholes which contain these constituents.[9]

There are several issues raised by the present version of the doctrine of emergence, though most of them have already been touched upon in the preceding discussion of reduction and can be settled on the basis of considerations which were introduced there.

a. The supposition underlying the notion of emergence is that, although it is possible in some cases to deduce the properties of a whole from the properties of its constituents, in other cases it is not possible to do so. We have seen, however, that both the affirmative and negative parts of this claim rest upon incomplete and misleading formulations of the actual facts. It is indeed impossible to deduce the properties of water (such as viscosity or translucency) from the properties of hydrogen alone (such as that it is in a gaseous state under certain conditions of pressure and temperature) or of oxygen alone, or of other compounds containing these elements as constituents (such as that hydrofluoric acid dissolves glass). But frequent claims to the contrary notwithstanding, it is also impossible to deduce the behavior of a clock merely from the properties and organization of its constituent parts. However, the deduction is impossible for the same reasons in both cases. It is not *properties*, but *statements* (or propositions) which can be deduced. Moreover, state-

[9] *Ibid.*, pp. 62-63.

ments about properties of complex wholes can be deduced from statements about their constituents only if the premises contain a suitable *theory* concerning these constituents—one which makes it possible to analyze the behavior of such wholes as "resultants" of the assumed behaviors of the constituents. Accordingly, all descriptive expressions occurring in a statement that is allegedly deducible from the theory must also occur among the expressions used to formulate the theory or the assumptions adjoined to the theory when it is applied to specialized circumstances. Thus a statement like 'Water is translucent' cannot indeed be deduced from any set of statements about hydrogen and oxygen which do not contain the expressions 'water' and 'translucent'; but this impossibility derives entirely from purely formal considerations and is relative to the special set of statements adopted as premises in the case under consideration.

b. It is clear, therefore, that to say of a given property that it is an "emergent" is to attribute to it a character which the property may possess relative to one theory or body of assumptions but may not possess relative to some other theory. Accordingly, the doctrine of emergence (in the sense now under discussion) must be understood as stating certain *logical* facts about formal relations between statements rather than any experimental or even "metaphysical" facts about some allegedly "inherent" traits of *properties* of objects.

It is worth repeating in this connection, and particularly when the constituents of complex wholes are assumed to be submicroscopic particles and processes, that the "properties" of such constituents cannot be ascertained by inspection and their "structure" cannot be learned by any form of "direct perception." What these properties and structures are can be formulated only by way of some *theory,* which postulates the existence of those constituents and assumes various characteristics for them. It is patent, moreover, that the theory is subject to indefinite modifications in the light of macroscopic evidence. Accordingly, the question whether a given property of compounds can be predicted from the properties of their atomic constituents cannot be settled by considerations concerning alleged "inherent natures" that atoms are antecedently known to possess. For while one theory of atomic structure may be unequal to the task of predicting a given property, another theory postulating a different structure for atoms may make it possible to do so.

This view of the question is supported by the history of atomic theory. The ancient atomic theory of matter was revived by Dalton in the first quarter of the nineteenth century in order to account systematically for a limited range of chemical facts—initially, facts about constancies in the ratios of combining weights of substances participating in chemical reactions. Dalton's form of the theory postulated relatively few properties for atoms, and his theory was incapable of explaining many features of chemical transformations; for example, it did not account for chemical valence or for thermal changes manifested in chemical transformations. Eventually, however, Dalton's theory was modified, so that an increasing number and variety of laws, dealing with optical, thermal, and electromagnetic as well as chemical phenomena, could be explained by its later variants. But with this series of modifications of the theory, the conception of the "intrinsic nature" of atoms was also transformed; for each variant of the theory—more precisely, each theory in a certain series of theoretical constructions having a number of broad assumptions in common—postulated (or "defined") distinctive kinds of submicroscopic components for macroscopic objects, with distinctive "natures" for the components in each case. Accordingly, the "atoms" of Democritus, the "atoms" of Dalton, and the "atoms" of modern physicochemical theory are quite different sorts of particles; and they can be subsumed under the common name of "atom" chiefly because there are important analogies between the various theories that define them.

We must therefore not be misled by the convenient habit of thinking of the various atomic theories as representing a progress in our knowledge concerning a fixed set of submicroscopic objects. This way of describing the historical succession of atomic theories easily generates the belief that atoms can be said to exist and to have ascertainable "inherent natures," independent of any particular theory that postulates the existence of atoms and prescribes what properties they possess. In point of fact, however, to maintain that there are atoms having some definite set of characteristics is to claim that a certain theory about the constitution of physical objects is warranted by experimental evidence. The succession of atomic theories propounded in the history of science may indeed represent not only advances in knowledge concerning the order and connection of macroscopic phenomena, but also a progressively more adequate understanding of the atomic constitution of physical things. It nevertheless does not follow that, apart from some particular atomic theory, it is possible to assert just what can or cannot be predicted from the "natures" of atomic particles.

In any event, it is certainly the case that properties of compounds not predictable from certain older theories of atomic structure (e.g., the chemical and optical properties of the stable substance formed when hydrogen and oxygen combine under certain conditions) *can* be predicted from the current electronic theory of the composition of atoms. It therefore follows that an elliptic formulation is being employed when it is claimed that a given property of a compound is an "emergent" one. For, although a property may indeed be an emergent trait relative to some given theory, it need not be emergent relative to some different theory.

c. However, while it is an error to claim that a given property is "inherently" or "absolutely" an emergent trait, it is equally an error to maintain that in characterizing a trait as an emergent we are only baptizing our ignorance. It has been argued, for example, that

> it may be that no physical-chemist could have predicted all the properties of H_2O before having studied it, and yet it seems probable that this incapacity to predict is only an expression of ignorance of the nature of H and O. If, on their combination, H and O yield water, presumably they contain in some sense the potentiality of forming water. In fact it is of the essence of Emergent Evolution that nothing new is added from without, that 'emergence' is the consequence of new kinds of relatedness between existents. The presumption is, then, that with sufficient knowledge of the components, highly probable predictions of the properties of water could have been made. In fact, chemists have successfully predicted the properties of compounds they have never observed and have proceeded to produce these 'emergents.' They have even predicted the existence and the properties of elements which had not been observed.[10]

Objections of this sort miss the force of the doctrine of emergence and appear to deny even what is demonstrably sound in it. In the first place, the doctrine employs the phrase 'to predict' in the sense of 'to deduce with strict logical rigor.' A proponent of emergence could readily admit that an allegedly emergent property might be foretold, whether invariably or only occasionally, by some happy insight or fortunate guess, but he would not thereby be compelled to surrender his claim that the property in question cannot be *predicted*. In the second place, it is possible to show that in some cases a given property cannot be predicted from certain other properties—more strictly, that a given statement about the occurrence of a designated property cannot be deduced from a specified set of other statements. For it may be possible to demonstrate with the help of established logical techniques that the statement about the first property is *not entailed* by the statements about the other properties; and such a demonstration is easily produced, especially when the former statement contains expressions that do not appear in the latter class of statements. Third and finally, our alleged "ignorance" or "incomplete knowledge" concerning the "natures" of atoms is entirely irrelevant to the issue at stake. For that issue is the simple one whether a given statement is deducible from a *given* set of statements, and not whether the statement is deducible from some *other* set of statements. As we have

[10] William McDougall, *Modern Materialism and Emergent Evolution*, New York, 1929, p. 129.

already seen, when we are said to improve or enlarge our knowledge concerning the "nature of H and O," what we are doing in effect is replacing one theory about H and O with another theory; and the fact that H and O combine to form water can be deduced from the second theory does not contradict the fact that the statement cannot be deduced from the initial set of premises. As was noted in discussing the reduction of thermodynamics to mechanics, the Boyle-Charles' law cannot be deduced from the assumptions of statistical mechanics unless a postulate is added relating the term 'temperature' to the expression 'mean kinetic energy of molecules.' This postulate cannot itself be deduced from statistical mechanics in its classical form; and this fact—that a postulate (or something equivalent to it) must be added to statistical mechanics as an independent assumption if the gas law is to be deduced —illustrates what is perhaps the central thesis in the doctrine of emergence as we have been interpreting it.

d. We have thus admitted the essential correctness of the doctrine of emergence when construed as a thesis concerning the *logical relation* between certain statements. It should be noted, however, that the doctrine so understood has a far wider range of application than proponents of emergence usually maintain. The doctrine has been urged for the most part in connection with chemical, biological, and psychological properties because these properties characterize systems at "higher levels" of organization and are allegedly "emergents" relative to properties occurring at "lower levels." Indeed, the doctrine is often advanced in opposition to the supposedly universalistic claims of "mechanical explanations," since, if some properties are in fact emergents, their occurrence is held to be inexplicable in "mechanical" terms. The truth of the doctrine of emergence is therefore sometimes believed to set limits to the science of mechanics, in which the principle of composition of forces is a warranted principle of analysis, and to differentiate from mechanics other systems of explanation in which that principle does not hold.[11] Accordingly, proponents of the doctrine often seem to suggest, if they do not explicitly maintain, that there are no emergent properties within the province usually assigned to mechanics or possibly even within the domain of physics; and the commonly cited example of a nonemergent property is the behavior of a clock, which is supposedly predictable from a knowledge of the properties and organization of its constituent cogs and springs.

But the logical point constituting the core of the doctrine of emer-

[11] Cf. the distinction drawn by Mill between the "mechanical" and the "chemical" modes of the "conjunct action of causes," which is the classical source of the doctrine of emergence. J. S. Mill, *A System of Logic*, London, 1879, Book 3, Chap. 6.

gence is applicable to all areas of inquiry and is as relevant to the analysis of explanations within mechanics and physics generally as it is to discussions of the laws of other sciences. The above discussion of the reduction of thermodynamics to mechanics makes this quite evident. But for the sake of additional clarity and emphasis, consider the clock example. It is well to note that the "behavior" of the clock which is predictable on the basis of mechanics is only that phase of its behavior which can be characterized entirely in terms of the primitive ideas of mechanics —for example, the behavior constituted by the motion of the clock's hands. Any phase of its behavior that cannot be brought within the scope of those ideas—for example, behavior consisting in variation in the clock's temperature or in changes in magnetic forces that may be generated by the relative motions of the parts of the clock—is not explained or predicted by mechanical theory. However, it appears that nothing but arbitrary custom stands in the way of calling these "nonmechanical" features of the clock's behavior "emergent properties" relative to mechanics. On the other hand, such nonmechanical features are certainly explicable with the help of theories of heat and magnetism, so that, relative to a wider class of theoretical assumptions, the clock may display no emergent traits.

Proponents of the doctrine of emergence are sometimes inclined to make a special point of the fact that the occurrence of so-called "secondary qualities" cannot be predicted by physical theories. For example, it has been argued that, from a complete knowledge of the microscopic structure of atoms, a mathematical archangel might be able to predict that nitrogen and hydrogen would combine when an electric discharge is passed through a mixture of the two, and would form water-soluble ammonia gas. However, though the archangel might be able to deduce what the exact microscopic structure of ammonia must be,

> he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be that certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself.[12]

But this claim is at best a truistic one, and can be affirmed with the same warrant for physical (or "primary") qualities of things as it can for secondary qualities. It is undoubtedly the case that a theory of chemistry that in its formulations makes no use of expressions referring to olfactory properties of substances cannot predict the occurrence of

[12] Broad, *op. cit.*, p. 71.

smells. But it cannot do so for the same reason that mechanics cannot account for optical or electrical properties of matter—namely that, when a deduction is made formally explicit, no statement employing a given expression can be logically derived from premises that do not also contain the expression. Accordingly, if a mathematical archangel is indeed incapable of predicting smells from a knowledge of the microscopic structure of atoms, this limitation in his powers is simply a consequence of the fact that the logical conditions for deducibility are the same for archangels as they are for men.

2. Let us now briefly consider the doctrine of emergence as an evolutionary cosmogony, whose primary stress is upon the alleged 'novelty' of emergent qualities. The doctrine of emergent evolution thus maintains that the variety of individuals and their properties that existed in the past or occur in the present is not complete, and that qualities, structures, and modes of behavior come into existence from time to time the like of which has never been previously manifested anywhere in the universe. Thus, according to one formulation of the doctrine, an emergent evolution is said to have taken place if, when the present state of the world (called "Ph.N.") is compared with any prior phase (called "Ph.A."), one or more of the following features lacking in Ph.A. can be shown to be present in Ph.N.:

> (1) Instances of some general type of change . . . common to both phases (e.g., relative motion of particles), of which instances the manner or condition of occurrence could not be described in terms of, nor predicted from, the laws which would have been sufficient for the description and . . . the prediction of all changes of that type occurring in Ph.A. Of this evolutionary emergence of laws one, though not the only conceivable, occasion would be the production, in accordance with one set of laws, of new local integrations in matter, the motions of which, and therefore of their component particles, would therefore conform to vector, i.e., directional, laws emergent in the sense defined. . . . (2) New qualities . . . attachable to entities already present, though without those accidents in Ph.A. (3) Particular entities *not* possessing all the essential attributes characteristic of those found in Ph.A., and having distinctive types of attributes (not merely configurational) of their own. (4) Some type or types of event or process irreducibly different in kind from any occurring in Ph.A. (5) A greater quantity, or number of instances, not explicable by transfer from outside the system, of any one or more types of prime entity common to both phases.[13]

[13] Arthur O. Lovejoy, "The Meanings of 'Emergence' and Its Modes," in *Proceedings of the Sixth International Congress of Philosophy* (ed. by Edgar S. Brightman), New York, 1927, pp. 26-27.

Emergent evolution as a doctrine of unceasing "creative novelty" is therefore commonly placed in opposition to the preformationist view, attributed especially to seventeenth-century science, that all the events of nature are simply the spatial rearrangements of a set of ultimate, simple "entities," whose total number, qualities, and laws of behavior remain invariant throughout the various juxtapositions into which they enter. However, some writers have gone beyond the assertion of such "creative novelty" and have outlined what they believe to be the successive stages of creative evolution; but we shall not concern ourselves with the details of these cosmic speculations.

a. It should be noted in the first place that the doctrine of creative evolution appears neither to entail nor to be entailed by the conception of emergence as the unpredictability of various properties. For it may very well be the case that a property is an emergent relative to a given theory but is not novel in a *temporal* sense. To take an extreme example, the property that bodies possess weight is not deducible from the classical theory of physical geometry; however, there is no reason to believe that bodies came to exhibit gravitational properties *after* they acquired spatial ones. On the other hand, it might be possible to deduce from some theory of atomic structure that nitrogen and oxygen could combine to form a water-soluble ammonia gas, although, because the prevailing physical conditions did not permit the formation of water in liquid state—say, before the time when the earth became sufficiently cool—no actual instance of ammonia dissolving in water had ever occurred. A subsequent formation of water with the dissolution of some ammonia gas in it would then be a temporally novel event. Accordingly, the question whether any properties are "emergents" in the sense of being temporally novel is a problem of a different order from the issue whether any properties are "emergents" in the sense of being unpredictable. The latter is an issue largely though not exclusively concerned with the *logical relations* of statements; the former is primarily a question that can be settled only by empirical *historical* inquiry.

b. Accordingly, the question whether a property, process, or mode of behavior is a case of emergent evolution is a straightforward empirical problem and can be resolved at least in principle by recourse to historical inquiry. Nevertheless, there are some difficulties facing attempts to answer it which deserve brief mention. One of these difficulties is a practical one, and arises from the circumstance that to answer the question conclusively we must possess a detailed knowledge of all the past occurrences in the universe (or in some portion of it), so as to be able to decide whether an alleged emergent trait or process is really such. But our knowledge of the past is seriously incomplete, and we possess fairly reliable evidence only in a limited class of cases to show that certain properties and processes could not have occurred before a given time. Thus we do not possess a sufficient basis for deciding beyond a reasonable doubt whether various processes on the atomic and subatomic levels which are believed to occur at present have always taken place, or whether they are characteristic of the current cosmic epoch. On the other hand, if we take for granted the dependence of living organisms upon favorable temperature conditions, and if we also assume that at one time the temperature of the earth was far too great for the functioning of such organisms, it becomes practically certain that living forms did not appear on the earth (or perhaps anywhere in the universe) before a certain age.

A second difficulty has its source in the vagueness of such words as 'property' and 'process' and in the lack of precise criteria for judging whether two properties or processes are to be counted as "the same" or as "different." Thus, the "mere" spatial rearrangement of a set of objects is apparently not to be regarded as an instance of an emergent property, even when that specific rearrangement has not previously occurred. Nevertheless, it is pertinent to ask whether every spatial redistribution of things is not always associated with some "qualitative" changes, so that spatial changes are *ipso facto* also alterations in the "properties" of the things redistributed. For example, the pattern formed by a square resting on one base certainly "looks different" from the pattern formed when the square is rotated so as to stand on one of its vertices. If the second pattern had not existed before, would its occurrence count as the appearance of a novel property? If it would not, what is the mark of a new trait? But if it would count as something novel, then almost any change must also be regarded as an illustration of emergent evolution. For a given state of affairs may be analyzable into a set of traits, each of which has occurred in the past. On the other hand, in their present manifestation the traits occur in a determinate context of relations; and, although the specific *pattern* of these relations is a repeatable one, those traits may in fact never have been previously exemplified in just that pattern. Accordingly, the given state of affairs would in that eventuality illustrate an emergent property; and, since every situation may very well exhibit such novel patterns, especially if no limits are placed on the spatiotemporal extent of a situation, the doctrine of emergence barely escapes collapse into the trivial thesis that things change.

Furthermore, just what is to be understood by the stipulation contained in the above quotation that a particular entity is to count as an instance of emergent evolution, if it does not possess "all the essential attributes" of entities in previous phases of evolution? In general, whether or not an attribute is to be regarded as an "essential" one depends on the context of the question and on the problem under consideration. But if this is so, then in view of that stipulation the distinction between an emergent trait and a nonemergent one would shift with changes in interest and with the purposes of an inquiry. These difficulties are not cited as being fatal to the doctrine of emergence. They do indicate, however, that, unless the doctrine is formulated with greater care than is customary, it can easily be construed as simply a truism.

c. The claim that there are emergent properties in the sense of emergent evolution is entirely compatible with the belief in the universality of the causal principle, at any rate in the form that there are determinate conditions for the occurrence of all events. Some proponents of emergent evolution do indeed combine the doctrine with versions of radical indeterminism; others invariably associate emergence with so-called "teleological" causation, thus attributing the appearance of novel qualities and processes to the operation of purposive agents. However, neither a belief in indeterminism nor in teleological causation is essential to emergent evolution. There are in fact many emergent evolutionists who maintain that the occurrence of a new chemical compound, for example, is always contingent upon the formation of definite though unique configurations of certain chemical elements; and they hold, furthermore, that, whenever these elements are conjoined in that special manner, whether through the agency of purposive creatures or adventitious circumstances, a compound of the same type is invariably formed.

d. It is also worth noting that, despite widespread opinion to the contrary, the assumptions and procedures of classical physics (and of mechanics in particular) neither imply nor contradict the thesis of emergent evolution. To be sure, there are philosophical interpretations of physics, according to which the properties of things are "ultimately" those distinctive of mechanics, and according to which also the only "real" changes in nature are spatial ones. However, such interpretations are of doubtful validity and cannot be assumed to be adequate accounts of the nature of physical theory. As we have seen, the science of mechanics does indeed operate with a limited and selected set of theoretical notions. However, this fact does not entail the requirement that the science deny either the actual existence or the possible emergence of traits of things other than those with which mechanics is primarily concerned. Such a denial would be unwarranted, even if earlier hopes of physicists had been realized and mechanics had continued to retain its one-time eminence as the universal science of nature. For a mechanical explanation of an event or process consists simply in stating the conditions for its occurrence in *mechanical terms.* But such explanations would clearly be impossible (on pain of making the enterprise of giving explanations for things self-defeating) if the event or process were not first identified by observing its characteristics—whether or not the characteristics are purely mechanical properties, and whether or not they are novel. In short, when the structure of mechanics or of any other theory of classical physics is analyzed, it becomes evident that the operative efficacy of the theory does not depend on acceptance or denial of the *historical* thesis that in the course of time novel traits and individuals appear in the universe.

e. Perhaps the most intriguing suggestion contained in the doctrine of emergent evolution is that the "laws of nature" may themselves change, and that novel patterns of dependence between events are manifested during different cosmic epochs. It will of course be clear that what is intended is not simply that our *knowledge* or our *formulation* of the structures of events and processes may be undergoing development, but that these *structures themselves* are altering with time. Thus, the Boyle-Charles' law is not as adequate a formulation of the behavior of gases as is the van der Waals' equation; but the fact that we have replaced the former with the latter is not taken to signify that the pattern of behavior of gases has undergone a change. Moreover, the suggestion does not consist merely in the supposal that the mode of behavior of some specific physical system is evolving. For example, there is evidence to indicate that the period of the earth's axial rotation is diminishing. However, this special fact is explained not by the assumption that the laws of mechanics are being altered, but in terms of such factors as the "braking" effect of the tides, produced by the sun and the moon in accordance with presumably unchanging laws. Accordingly, what the suggestion contemplates is the possibility that pervasive *types of structure* are changing, or that novel relational patterns are manifested by things; for example, instead of permanently remaining inversely proportional to the square of the distance, the gravitational force between all pairs of particles may be slowly changing so that this latter exponent is increasing with time; or various chemical elements may exhibit progressively new properties and new modes of combination with one another. However, the suggestion is not without serious difficulties, some of which must now be noted.

Perhaps the most obvious and crucial of these stems from the fact that we cannot be sure whether an apparent change in a law is really such, or whether it merely indicates that our knowledge was incomplete concerning the conditions under which some type of structure prevails.

Suppose, for example, that evidence were available which seems to show that some universal constant (such as the velocity of light *in vacuo*) is changing, so that its value during the present century is smaller than it was during prehistoric times. However, other things have also changed in the interim: the relative positions of the galaxies are no longer the same; there have been internal changes in the stars and in the quantity of radiation they emit; and possibly even some hitherto undetected trait of physical bodies has varied (some trait comparable to the electric properties of matter, which have been discovered by men only relatively recently). It is therefore at least conceivable that the hitherto asserted law of the constancy of the velocity of light is simply erroneous, and that this velocity varies with some such factors as have been mentioned. It would certainly not be a simple task to eliminate this alternative interpretation of the evidence; and in fact most scientists would doubtless be more inclined to regard the hitherto accepted law as correct only when certain antecedent conditions are satisfied—and therefore to regard it as simply a limiting case of a more inclusive law—rather than to assume that the pervasive structure of physical occurrences is undergoing evolution. In any event, whether such an assumption will ever be widely accepted will most likely depend on how effective and convenient it proves to be in establishing a thoroughly inclusive and integrated system of knowledge. Accordingly, although the suggestion that some laws may be evolving does not fall outside the bounds of possibility, it is at best a highly speculative one for which it is not easy to supply reasonably conclusive evidence.

There is an additional difficulty of a different order which faces the doctrine that *all* laws are changing with time.[14] For how is evidence obtained for the claim that a law is undergoing change? A pervasive pattern of relations cannot be literally "seen" to evolve, and the basis for such a conclusion must be obtained from comparisons of the present with the past. However, the past is not accessible to direct inspection. It can only be reconstructed from data available in the present, with the help of laws which must be assumed to be unchanging at least during the epoch which includes that past and the present. For example, suppose that the gravitational force between bodies is alleged to be slowly diminishing, on the ground that in the past the tides were generally higher than in the present, even though the number and relative position of celestial bodies were the same as at present. But how can we know that the past was indeed like this, unless we use laws that have not altered in order to infer those past facts from present data? Thus, we might

[14] Cf. Henri Poincaré, "L'Evolution des Lois," in *Dernières Pensées*, Paris, 1926; Pascual Jordan, *Die Herkunft der Sterne*, Stuttgart, 1947.

find deposits of sea salt at altitudes now out of the reach of the tides. However, even if we waive the question whether the land had not been elevated by geological action rather than because of a diminution in the height of tides, the conclusion that the salt was deposited by the ocean takes for granted various laws concerning the motions of tidal water and the evaporation of liquids. Accordingly, the assumption that all laws are simultaneously involved in a process of change is self-annihilating, for, since the past would then be completely inaccessible to knowledge we would be unable to produce any evidence for that assumption.

The form in which the suggestion of emergent laws appears most plausible is that new types of behavior conforming to novel modes of dependence arise when hitherto nonexistent combinations and integrations of matter occur. For example, chemists have produced substances in the laboratory which, as far as we can tell, have never existed before, and which possess properties and ways of interacting with other substances that are distinctive and novel. What has thus occasionally happened in the laboratory of chemists has undoubtedly happened more frequently in the older and vaster laboratory of nature. It might of course be said that such novel types of dependence are not "really novel" but are only the realizations of "potentialities" that have always been present in "the natures of things"; and it might also be said that, with "sufficient knowledge" of these "natures," anyone having the requisite mathematical skills could predict the novelties in advance of their realization. We have already commented sufficiently on the latter part of this rejoinder, and can therefore discount it without further ado as both invalid and irrelevant. As for the first part of the objection, it must be admitted that it is irrefutable; but it will also be clear that what the objection asserts has no factual content, and that its irrefutability is that of a definitional truism.

V. *Wholes, Sums, and Organic Unities*

Before leaving the subjects of reduction and emergence, it will be convenient to discuss a familiar thesis frequently associated with these themes. According to this conception, there is an important type of individual wholes (physical, biological, psychological, as well as social) distinguished from others by the fact that they are "organic unities," and not simply "aggregates" of independent parts or members. Wholes of this type are often characterized by the dictum that they possess an organization in virtue of which "the whole is more than the sum of its parts." Since the existence of organic wholes is sometimes taken to place fixed limits on the possibility of effecting reductions in the sciences, as well as on the scope of the methods of physics, it is desirable to examine such wholes with care.

A preliminary point must first be noted. As commonly employed, the words 'whole,' 'sum,' and their derivatives are unusually ambiguous, metaphorical, and vague. It is therefore frequently impossible to assess the cognitive worth as well as the meaning of statements containing them, so that some of the many senses of those words must be distinguished and clarified. Some examples will make evident the need for such clarification. A quadrilateral encloses an area, and either one of its two diameters divides the figure into two partial areas whose sum is equal to the area of the initial figure. In this geometrical context, and in many analogous ones as well, the statement 'The whole is equal to the sum of its parts' is normally accepted as true. Indeed, the statement in this context is frequently acknowledged to be not only true but necessarily true, so that its denial is regarded as self-contradictory. On the other hand, in discussing the taste of sugar of lead as compared with the tastes of its chemical components, some writers have maintained that in this case the whole is *not* equal to the sum of its parts. This claim is obviously intended to be informative about the matters discussed, and it would be high-handed to reject it outright as simply a logical absurdity. It is clear, nevertheless, that in the context in which this claim is made the words 'whole,' 'part,' 'sum,' and perhaps even 'equal,' are being employed in senses different from those associated with them in the geometrical context. We must therefore assume the task of distinguishing between a number of senses of these words that appear to play a role in various inquiries.

1. The words 'whole' and 'part' are normally used for correlative distinctions, so that x is said to be a whole in relation to something y which is a component or part of x in some sense or other. It will be convenient, therefore, to have before us a brief list of certain familiar "kinds" of wholes and corresponding parts.

a. The word 'whole' is used to refer to something with a spatial extension, and anything is then called a 'part' of such a whole which is spatially included in it. However, several special senses of 'whole' and 'part' fall under this head. In the first place, the terms may refer to specifically spatial properties, so that the whole is then some length, area, or volume containing as parts lengths, areas, or volumes. In this sense, neither wholes nor parts need be spatially continuous; thus, the United States and its territorial possessions are not a spatially continuous whole, and continental United States contains as one of its spatial parts desert regions which are also not spatially continuous. In the second place, 'whole' may refer to a nonspatial property or state of a spatially extended thing, and 'part' designates an identical property of some spatial part of the thing. Thus, the electric charge on a body is said to have for its parts the electric charges on spatial parts of the body. In the third place, though sometimes the only spatial properties counted as parts of a spatial whole are those that have the same spatial dimensions as the whole, at other times the usage of the terms is more liberal. Thus the surface of a sphere is frequently said to be a part of the sphere, but on other occasions only volumes in the sphere's interior are so designated.

b. The word 'whole' refers to some temporal period whose parts are temporal intervals in it. As in the case of spatial wholes and parts, temporal ones need not be continuous.

c. The word 'whole' refers to any class, set, or aggregate of elements, and 'part' may then designate either any proper subclass of the initial set, or any element in the set. Thus, by a part of the whole consisting of all the books printed in the United States during a given year may be understood either all the novels printed that year, or some particular copy of a novel.

d. The word 'whole' sometimes refers to a property of an object or process, and 'part' to some analogous property standing to the first in certain specified relations. Thus, a force in physics is commonly said to have for its parts or components other forces into which the first can be analyzed according to a familiar rule. Similarly, the physical brightness of a surface illuminated by two sources of light is sometimes said to have for one of its parts the brightness associated with one of the sources. In the present sense of the words, a part is not a spatial part of the whole.

e. The word 'whole' may refer to a pattern of relations between certain specified kinds of objects or events, the pattern being capable of embodiment on various occasions and with various modifications. However, 'part' may then designate different things in different contexts. It may refer to any one of the elements which are related in that pattern on some occasion of its embodiment. Thus, if a melody (say "Auld Lang Syne") is such a whole, one of its parts is then the first tone that is sounded when the melody is sung on a particular date. Or it may refer to a class of elements that occupy corresponding positions in the pattern in some specified mode of its embodiment. Thus, one of the parts of the melody will then be the class of first notes when "Auld Lang Syne" is sung in the key of E flat. Or the word 'part' may refer to a subordinate phrase in the total pattern. In this case, a part of the melody may be the pattern of tones that occurs in its first four bars.

f. The word 'whole' may refer to a process, one of its parts being another process that is some discriminated phase of the more inclusive one. Thus, the process of swallowing is part of the process of eating.

g. The word 'whole' may refer to any concrete object, and 'part' to any of its properties. In this sense, the character of being cylindrical in shape or being malleable is a part of a given piece of copper wire.

h. Finally, the word 'whole' is often used to refer to any system whose spatial parts stand to each other in various relations of dynamical dependence. Many of the so-called "organic unities" appear to be systems of this type. However, in the present sense of 'whole' a variety of things are customarily designated as its parts. Thus, a system consisting of a mixture of two gases inside a container is frequently, though not always in the same context, said to have for its parts one or more of the following: its spatially extended constituents, such as the two gases and the container; the properties or states of the system or of its spatial parts, such as the mass of the system or the specific heats of one of the gases; the processes which the system undergoes in reaching or maintaining thermodynamical equilibrium; and the spatial or dynamical organization to which its spatial parts are subject.

This list of senses of 'whole' and 'part,' though by no means complete, will suffice to indicate the ambiguity of these words. But what is more important, it also suggests that, since the word 'sum' is used in a number of contexts in which these words occur, it suffers from an analogous ambiguity. Let us therefore examine several of its typical senses.

2. We shall not inquire whether the word 'sum' actually is employed in connection with each of the senses of 'whole' and 'part' that have been distinguished, and if so just what meaning is to be associated with it. In point of fact, it is not easy to specify a clear sense for the word in many contexts in which people do use it. We shall accordingly confine ourselves to noting only a small number of the well-established uses of 'sum' and to suggesting interpretations for it in a few contexts in which its meaning is unclear and its use misleading.

a. It is hardly surprising that the most carefully defined uses of 'sum' and 'addition' occur in mathematics and formal logic. But even in these contexts the word has a variety of special meanings, depending on what type of mathematical and logical "objects" are being added. Thus, there is a familiar operation of addition for the natural integers; and there are also identically named but really distinct operations for ratios, real numbers, complex numbers, matrices, classes, relations, and other mathematical or logical "entities." It is not altogether evident why all these operations have the common name of 'addition,' though there are at least certain formal analogies between many of them; for example, most of them are commutative and associative. However, there are some important exceptions to the general rule implicit in this example, for the addition of *ordered* sets is not uniformly commutative, though it is associative. On the other hand, the sum of two entities in mathematics is invariably some unique entity which is of the same type as the summands;—thus the sum of two integers is an integer, of two matrices a matrix, and so on. Moreover, though the word 'part' is not always defined or used in connection with mathematical "objects," whenever both it and 'sum' are employed they are so used that the statement 'The whole is equal to the sum of its parts' is an analytic or necessary truth.

However, it is easy to construct an apparent counter-instance to this last claim. Let K^* be the *ordered* set of the integers, ordered in the following manner: first the odd integers in order of increasing magnitude, and then the even integers in that order. K^* may then be represented by the notation: $(1, 3, 5, \ldots, 2, 4, 6, \ldots)$. Next let K_1 be the class of odd integers and K_2 the class of even ones, neither class being an ordered set. Now let K be the class-sum of K_1 and K_2, so that K contains all the integers as members; K also is not an ordered class. But the membership of K is the same as that of K^*, although quite clearly K and K^* are not identical. Accordingly, so it might be argued, in this case the whole (namely K^*) is not equal to the sum (i.e., K) of its parts.

This example is instructive on three counts. It shows the possibility of defining in a precise manner the words 'whole,' 'part,' and 'sum' so that 'The whole is unequal to the sum of its parts' is not only not logically absurd but is in fact logically true. There is thus no a priori reason for dismissing such statements as inevitable nonsense; and the real issue is to determine, when such an assertion is made, in what sense if any the crucial words in it are being used in the given context. But the example also shows that, though such a sentence may be true on one specified usage of 'part' and 'sum,' it may be possible to assign other senses to these words so that the whole *is* equal to the sum of its parts in this redefined sense of the words. Indeed, it is not standard usage in mathematics to call either K_1 or K_2 a part of K^*. On the contrary, it is customary to count as a part of K^* only an *ordered* segment. Thus, let K_1^* be the ordered set of odd integers arranged according to increasing magnitude, and K_2^* the corresponding ordered set of even integers. K_1^* and

K_2^* are then parts of K^*. [K^* has other parts as well, for example, the ordered segments indicated by the following: (1, 3, 5, 7), (9, 11, . . . , 2, 4), and (6, 8, . . .).] Now form the *ordered sum* of K_1^* and K_2^*. But *this* sum yields the ordered set K^*, so that in the specified senses of 'part' and 'sum' the whole *is* equal to the sum of its parts. It is thus clear that, when a given system has a special type of organization or structure, a *useful* definition of 'addition,' if such can be given, must take into account that mode of organization. Any number of operations could be selected for the label 'summation,' but not all of them are relevant or appropriate for advancing a given domain of inquiry.

Finally, the example suggests that, though a system has a distinctive structure, it is not in principle impossible to specify that structure in terms of relations between its elementary constituents, and moreover in such a manner that the structure can be correctly characterized as a 'sum' whose 'parts' are themselves specified in terms of those elements and relations. As we shall see, many students deny, or appear to deny, this possibility in connection with certain kinds of organized systems (such as living things). The present example therefore shows that, though we may not be able *as a matter of fact* to analyze certain highly complex "dynamic" (or "organic") unities in terms of some given theory concerning their ultimate constituents, such inability cannot be established as a matter of *inherent logical necessity*.

b. If we now turn to the positive sciences, we find that here too are a large number of well-defined operations called 'addition.' The major distinction that needs to be drawn is between scalar and vector sums. Let us consider each in turn. Examples of the former are the addition of the numerosity of groups of things, of spatial properties (length, area, and volume), of temporal periods, of weights, of electrical resistance, electric charge, and thermal capacity. They illustrate the first three senses of 'whole' and 'part' which we distinguished above; and in each of them (and in many other cases that could be mentioned) 'sum' is so specified that the whole is the sum of appropriately chosen parts.

On the other hand, there are many magnitudes, such as density or elasticity, for which no operation of addition is defined or seems capable of being defined in any useful manner; most of these cases fall under the last four of the above distinctions concerning 'whole' and 'part.' Moreover, there are some properties for which addition is specified only under highly specialized circumstances; for example, the sum of the brightness of two sources of light is defined only when the light emitted is monochromatic. It makes no sense, therefore, to say that the density (or the shape) of a body is, or is not, the sum of the densities (or shapes) of its parts, simply because there are neither explicitly formulated rules nor ascertainable habits of procedure which associate a usage with the word 'sum' in such a context.

The addition of vector properties, such as forces, velocities, and accelerations, conforms to the familiar rule of parallelogram composition. Thus, if a body is acted on by a force of 3 poundals in a direction due north, and also by a force of 4 poundals in a direction due east, the body will behave as if it were acted on by a single force of 5 poundals in a northeasterly direction. This single force is said to be the 'sum' or 'resultant' of the other two forces, which are called its 'components'; and, conversely, any force can be analyzed as the sum of an arbitrary number of components. This sense of 'sum' is commonly associated with the fourth of the above distinctions concerning 'whole' and 'part'; and it is evident that here the sense of 'sum' is quite different from the sense of the word in such contexts as 'the sum of two lengths.'

It has been argued by Bertrand Russell that a force cannot rightly be said to be the sum of its components. Thus he declared:

> Let there be three particles *A*, *B*, *C*. We may say that *B* and *C* both cause accelerations in *A*, and we compound these accelerations by the parallelogram law. But this composition is not truly addition, for the components are not *parts* of the resultant. The resultant is a new term, as simple as their components, and not by any means their sum. Thus the effects attributed to *B* and *C* are never produced, but a third term different from either is produced. This, we may say, is produced by *B* and *C* together, taken as a whole. But the effect which they produce as a whole can only be discovered by supposing each to produce a separate effect: if this were not supposed, it would be impossible to obtain the two accelerations whose resultant is the actual acceleration. Thus we seem to reach an antinomy: the whole has no effect except what results from the effects of the parts, but the effects of the parts are nonexistent.[15]

However, all this argument shows is that by the component of a force (or of an acceleration) we do not mean anything like what we understand by a component or part of a length—the components of forces are not *spatial parts* of forces. It does not establish the claim that the addition of forces "is not truly addition," unless, indeed, the word 'addition' is being used so restrictively that no operation is to be so designated which does not involve a juxtaposition of spatial (or possibly temporal) parts of the whole said to be their sum. But in this latter event many other operations that are called 'addition' in physics, such as the addition of electrical capacities, would also have to receive different labels. Moreover, no antinomy arises from the supposition that, on the one hand,

[15] Bertrand Russell, *The Principles of Mathematics,* Cambridge, England, 1903, p. 477.

the effect of each component force acting alone does not exist, while on the other hand the actual effect produced by the joint action of the components is the resultant of their partial effects. For the supposition simply expresses what is the case, in a language conforming to the antecedent *definition* of the addition and resolution of forces.

The issue raised by Russell is thus terminological at best. His objection is nevertheless instructive. For it calls needed attention to the fact that, when the matter is viewed abstractly, the 'sum' of a given set of elements is simply an element that is *uniquely determined* by some *function* (in the mathematical sense) of the given set. This function may be assigned a relatively simple and familiar form in certain cases, and a more complex and strange form in others; and in any event, the question whether such a function is to be introduced into a given domain of inquiry, and if so what special form is to be assigned to it, cannot be settled a priori. The heart of the matter is that when such a function is specified, and if a set of elements satisfies whatever conditions are prescribed by the function, it becomes possible to *deduce* from these premises a class of statements about some structural complex of those elements.[16]

c. We must now consider a use of 'sum' associated with the fifth sense of 'whole' and 'part' distinguished above—a use also frequently associated with the dictum that the whole is more than, or at any rate not merely, the sum of its parts. Let us assume that the following statement is typical of such usage: "Although a melody may be produced by sounding a series of individual tones on a piano, the melody is not the sum of its individual notes." The obvious question that needs to be asked is: "In what sense is 'sum' being employed here?" It is evident that the statement can be informative only if there *is* such a thing as the sum of the individual tones of melody. For the statement can be established as true or false only if it is possible to compare such a sum with the whole that is the melody.

However, most people who are inclined to assert such a statement do

[16] An issue similar to the one raised by Russell has been raised in connection with the addition of velocities in relativity theory. Let A, B, C be three bodies, so that the velocity of A with respect to B is v_{AB}, that of B with respect to C is v_{BC} (where the direction of v_{BC} is parallel to the direction of v_{AB}), and of A with respect to C is v_{AC}. Then according to classical mechanics, $v_{AC} = v_{AB} + v_{BC}$. But according to the special relativity theory,

$$v_{AC} = \frac{v_{AB} + v_{BC}}{1 + \frac{v_{AB}v_{BC}}{c^2}}$$

where c is the velocity of light. It has been argued that in the latter we are not "really adding" velocities. However, this objection can be disposed of in essentially the same manner as can Russell's argument.

not specify what that sum is supposed to be; and there is thus a basis for the supposition that they either are not clear about what they mean or do not mean anything whatever. In the latter case the most charitable view that can be taken of such pronouncements is to regard them as simply misleading expressions of the possibly valid claim that the notion of summation is *inapplicable* to the constituent tones of melodies. On the other hand, some writers apparently understand by 'sum' in this context the *unordered class* of individual tones; and what they are therefore asserting is that this class is not the melody. But this is hardly news, though conceivably there may have been some persons who believed otherwise. In any event, there appears to be no meaning, other than this one, which is normally associated with the phrase 'sum of tones' or similar phrases. Accordingly, if the word 'sum' is used in this sense in contexts in which the word 'whole' refers to a pattern or configuration formed by elements standing to each other in certain relations, it is perfectly true though trivial to say that the whole is more than the sum of its parts.

As has already been noted, however, this fact does not preclude the possibility of *analyzing* such wholes into a set of elements related to one another in definite ways; nor does it exclude the possibility of assigning a different sense to 'sum' so that a melody might then be construed as a sum of appropriately selected parts. It is evident that at least a partial analysis of a melody is effected when it is represented in the customary musical notation; and the analysis could obviously be made more complete and explicit, and even expressed with formal precision.[17]

But it is sometimes maintained in this connection that it is a fundamental mistake to regard the constituent tones of a melody as independent parts, out of which the melody can be reconstituted. On the contrary, it has been argued that what we "experience at each place in the melody is a *part* which is itself determined by the character of the whole. . . . The flesh and blood of a tone depends from the start upon its role in the melody: a *b* as leading tone to *c* is something radically different from the *b* as tonic." [18] And as we shall see, similar views have been advanced in connection with other cases and types of Gestalts and "organic" wholes.

Now it may be quite true that the *effect* produced by a given tone depends on its position in a context of other tones, just as the effect produced by a given pressure upon a body is in general contingent upon

[17] For an interesting sketch of a generalized formal analysis of Gestalts such as melodies, cf. Kurt Grelling and Paul Oppenheim, "Der Gestaltbegriff in Lichte der neuen Logik," *Erkenntnis*, Vol. 7 (1938), pp. 211-25.

[18] Max Wertheimer, "Gestalt Theory," in *A Source Book of Gestalt Psychology* (ed. by Willis D. Ellis), New York, 1950, p. 5.

what other pressures are operative. But this supposed fact does not imply that a melody cannot rightly be viewed as a relational complex whose component tones are identifiable independently of their occurrence in that complex. For if the implication did hold, it would be impossible to describe how a melody is constituted out of individual tones, and therefore impossible to prescribe how it is to be played. Indeed, it would then be self-contradictory to say that "a *b* as leading tone to *c* is something radically different from the *b* as tonic." For the name '*b*' in the expression '*b* as leading tone to *c*' could then not refer to the same tone to which the name '*b*' refers in the expression '*b* as tonic'; and the presumable intent of the statement could then not be expressed. In short, the fact that, in connection with wholes that are patterns or Gestalts of occurrences, the word 'sum' is either undefined or defined in such a way that the whole is unequal to the sum of its parts, constitutes no inherently insuperable obstacle to analyzing such wholes into elements standing to each other in specified relations.

d. We must finally examine the use of 'sum' in connection with wholes that are organized systems of dynamically interrelated parts. Let us assume as typical of such usage the statement 'Although the mass of a body is equal to the sum of the masses of its spatial parts, a body also has properties which are not the sums of properties possessed by its parts.' The comments that have just been made about 'sum' in connection with patterns of occurrences such as melodies can be extended to the present context of usage of the word; and we shall not repeat them. In the present instance, however, an additional interpretation of 'sum' can be suggested.

When the behavior of a machine like a clock is sometimes said to be the sum of the behavior of its spatial parts, what is the presumptive content of the assertion? It is reasonable to assume that the word 'sum' does not here signify an unordered class of elements, for neither the clock nor its behavior is such a class. It is therefore plausible to construe the assertion as maintaining that, from the theory of mechanics, coupled with suitable information about the actual arrangements of the parts of the machine, it is possible to deduce statements about the consequent properties and behaviors of the entire system. Accordingly, it seems also plausible to construe in a similar fashion statements such as that of J. S. Mill: "The different actions of a chemical compound will never be found to be the sums of actions of its separate parts." [19] More explicitly, this statement can be understood to assert that from some assumed theory concerning the constituents of chemical compounds, even when it

[19] J. S. Mill, *A System of Logic*, London, 1879, Book 3, Chap. 6, § 2 (Vol. 1, p. 432).

is conjoined with appropriate data on the organization of these constituents within the compounds, it is not in fact possible to deduce statements about many of the properties of these compounds.

If we adopt this suggestion, we obtain an interpretation for 'sum' that is particularly appropriate for the use of the word in contexts in which the wholes under discussion are organized systems of interdependent parts. Let T be a theory that is in general able to explain the occurrence and modes of interdependence of a set of properties $P_1, P_2, \ldots, P_k$. More specifically, suppose it is known that, when one or more individuals belonging to a set K of individuals occur in an environment E_1 and stand to each other in some relation belonging to a class of relations R_1, the theory T can explain the behavior of such a system with respect to its manifesting some or all of the properties P. Now assume that some or all of the individuals belonging to K form a relational complex R_2 not belonging to R_1 in an environment E_2, which may be different from E_1, and that the system exhibits certain modes of behavior that are formulated in a set of laws L. Two cases may then be distinguished: from T, together with statements concerning the organization of the individuals in R_2, it is possible to deduce the laws L; or secondly, not all the laws L can be so deduced. In the first case, the behavior of the system R_2 may be said to be the 'sum' of the behaviors of its component individuals; in the second case, the behavior of R_2 is *not* such a sum. It is evident that in the terminology and distinctions of the present chapter, both conditions for the reducibility of L to T are satisfied in the first case; in the second case, however, although the condition of connectability may be satisfied, the condition of derivability is not.

If this interpretation of 'sum' is adopted for the indicated contexts of its usage (let us call this the "reducibility sense" of the word), it follows that the distinction between wholes that are sums of their parts and those that are not is *relative to some assumed theory* T in terms of which the analysis of a system is undertaken. Thus, as we have seen, the kinetic theory of matter as developed during the nineteenth century was able to explain certain thermal properties of gases, including certain relations between the specific heats of gases. However, that theory was unable to account for these relations between specific heats when the state of aggregation of molecules is that of a solid rather than a gas. On the other hand, modern quantum theory is capable of explaining the facts concerning the specific heats of solids, and presumably also all other thermal properties of solids. Accordingly, although relative to classical kinetic theory the thermal properties of solids are not sums of the properties of their parts, relative to quantum theory those properties are such sums.

3. We must now briefly consider the distinctive feature of those systems that are commonly said to be "organic unities" and that exhibit a mode of organization often claimed to be incapable of analysis in terms of an "additive point of view." However, although living bodies are the most frequently cited examples of organic wholes, we shall not be now concerned specifically with such systems. For it is generally admitted that living bodies constitute only a special class of systems possessing a structure of internally related parts; and it will be an advantage to ignore for the present special issues connected with the analysis of vital phenomena.

Organic or "functional" wholes have been defined as systems "the behavior of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole." [20] What is distinctive of such systems, therefore, is that their parts do not act, and do not possess characteristics, *independently* of one another. On the contrary, their parts are supposed to be so related that any alteration in one of them causes a change in *all* the other parts.[21] In consequence, functional wholes are also said to be systems which cannot be built up out of elements by combining these latter seriatim without producing changes in all those elements. Moreover, such wholes cannot have any part removed without altering both that part and the remaining parts of the system.[22] Accordingly, it is often claimed that a functional whole cannot be properly analyzed from an "additive point of view"; that is, the characteristic modes of functioning of its constituents must be studied *in situ,* and the structure of activities of the whole cannot be inferred from properties displayed by its constituents in isolation from the whole.

A purely physical example of such functional wholes has been made familiar by Köhler. Consider a well-insulated electric conductor of arbitrary shape, for example, one having the form of an ellipsoid; and assume that electric charges are brought to it successively. The charges will immediately distribute themselves over the surface of the conductor

[20] Max Wertheimer, *op. cit.,* p. 2. Cf. also Koffka's statement: "Analysis if it wants to reveal the universe in its completeness has to stop at the wholes, whatever their size, which possess functional reality. . . . Instead of starting with the elements and deriving the properties of the wholes from them a reverse process is necessary, i.e., to try to understand the properties of parts from the properties of wholes. The chief content of Gestalt as a category is this view of the relation of parts and wholes involving the recognition of intrinsic real dynamic whole-properties."—K. Koffka, "Gestalt," in *Encyclopedia of the Social Sciences,* New York, 1931, Vol. 6, p. 645, quoted by kind permission of the publishers, The Macmillan Company.

[21] Cf. Kurt Lewin, *Principles of Topological Psychology,* New York, 1936, p. 218.

[22] W. Köhler, *Die physischen Gestalten im Ruhe und in stationären Zustand,* Braunschweig, 1924, p. 42; also Ellis, *op. cit.,* p. 25.

in such a way that the electric potential will be the same throughout the surface. However, the density of the charge (i.e., the quantity of charge per unit surface) will not in general be uniform at all points of the surface. Thus, in the ellipsoidal conductor, the density of the charge will be greatest at the points of greatest curvature and will be smallest at the points of least curvature.[23] In brief, the distribution of the charges will exhibit a characteristic pattern or organization—a pattern which depends on the shape of the conductor but is independent of the special materials of its construction or of the total quantity of charge placed upon it.

It is, however, not possible to build up this pattern of distribution bit by bit, for example, by bringing charges first to one part of the conductor and then to another so as to have the pattern emerge only after all the charges are placed on the conductor. For when a charge is placed on one portion of the surface, the charge will not remain there but will distribute itself in the manner indicated; and in consequence, the charge density at one point is not independent of the densities at all other points. Similarly, it is not possible to remove some part of the charge from one portion of the surface without altering the charge densities at all other points. Accordingly, although the total charge on a conductor is the sum of separable partial charges, the configuration of charge densities cannot be regarded as composed from independent parts. Köhler thus declares:

> The natural structure assumed by the total charge is not described if one says: at this point the charge-density is this much 'and' at that point the density is that much, etc.; but one might attempt a description by saying: the density is so much at this point, so much at that point, all mutually interdependent, and such that the occurrence of a certain density at one point determines the densities at all other points.[24]

Many other examples—physical, chemical, biological, and psychological—could be cited which have the same intent as this one. Thus

[23] More generally, the charge density on the ellipsoid is proportional to the fourth root of the curvature at a point.

[24] Köhler, *op. cit.,* p. 58, and cf. also p. 166. Many other physical examples of such "functional" wholes could be cited. The surfaces assumed by soap films provide an intuitively evident illustration. The general principle underlying the analysis of such surfaces is that, subject to the boundary conditions imposed on the surface, its area is a minimum. Thus, neglecting gravity, a soap film bounded by a plane loop of wire will assume a plane surface; a soap bubble will assume the shape of a sphere, a figure which has the minimum surface for a given volume. Now consider a part of the surface of a soap bubble bounded by a circle. If this part were removable from the spherical surface, it would no longer retain its convex shape, but would become a plane. Thus, the shape assumed by a part of the film depends on the whole of which it is a part. Cf. the accounts of soap film experiments in Richard Courant and Herbert Robbins, *What Is Mathematics?* New York, 1941, pp. 386ff.

there is no doubt that in many systems the constituent parts and processes are "internally" related, in the sense that these constituents stand to each other in relations of mutual causal interdependence. Indeed, some writers have found it difficult to distinguish sharply between systems which are of this sort and systems which allegedly are not; and they have argued that all systems whatever ought to be characterized as wholes which are "organic" or "functional" in some degree or other.[25] In point of fact, many who claim that there is a fundamental difference between functional and nonfunctional (or "summative") wholes tacitly admit that the distinction is based on *practical decisions* concerning what causal influences may be ignored for certain purposes. Thus, Köhler cites as an example of a "summative" whole a system of three stones, one each in Africa, Australia, and the United States. The system is held to be a summative grouping of its parts, because displacement of one stone has no effect on the others or on their mutual relations.[26] However, if current theories of physics are accepted, such a displacement is not without *some* effects on the other stones, even if the effects are so minute that they cannot be detected with present experimental techniques and can therefore be practically ignored. Again, Köhler regards the total charge on a conductor as a summative whole of independent parts, though it is not at all evident that the electronic constituents of the charge undergo no alterations when parts of the charge are removed from it. Accordingly, although the occurrence of systems possessing distinctive structures of interdependent parts is undeniable, no general criterion has yet been proposed which makes it possible to identify in an absolute way systems that are "genuinely functional" as distinct from systems that are "merely summative."[27]

[25] This is the contention of A. N. Whitehead's philosophy of organism. Cf. his *Process and Reality*, New York, 1929, esp. Part 2, Chaps. 3 and 4.

[26] Köhler, *op. cit.*, p. 47.

[27] This suggestion that the distinction between functional and nonfunctional wholes is not a sharp one, is borne out by an attempt to state more formally the character of an "organic" whole. Let S be some system and K a class of properties $P_1, \ldots, P_n$ which S may exhibit. Assume, for the sake of simplicity of exposition, that these properties are measurable in some sense, so that the specific forms of these properties can be associated with the values of numerical variables; and assume, also for the sake of simplicity, that statements about these properties have the form 'At time t the property P_i of S has the value x,' or, more compactly, '$P_i(S,t) = x$.' We now define a property in K, say P_1, to be "dependent" on the remaining properties in K when P_1 has the same value at different times if the remaining properties have equal values at those times; that is, when for every property P_i in K, if $P_i(S,t_1) = P_i(S,t_2)$ then $P_1(S,t_1) = P_1(S,t_2)$. Moreover we shall say that the class K of properties is "interdependent" if *each* property in the class is dependent on the remaining properties in K, that is, when for every P_i and P_j in K, if $P_i(S,t_1) = P_i(S,t_2)$ then $P_j(S,t_1) = P_j(S,t_2)$. On the other hand, we can define the class K to be an "independent" class if no property in K is dependent on the remaining properties of K. To fix our ideas, let S be a gas, V its volume, p its pressure, and T its absolute temperature. Then according to the Boyle-Charles' law, V is depend-

Moreover, it is essential to distinguish in this connection between the question whether a given system can be *overtly constructed* in a piecemeal fashion by a seriatim juxtaposition of parts, and the question whether the system can be *analyzed in terms of a theory* concerning its assumed constituents and their interrelations. There undoubtedly are wholes for which the answer to the first question is affirmative—for example, a clock, a salt crystal, or a molecule of water; and there are wholes for which the answer is negative—for example, the solar system, a carbon atom, or a living body. However, this difference between systems does not correspond to the intended distinction between functional and summative wholes; and our inability to construct effectively a system out of its parts, which in some cases may only be a consequence of temporary technological limitations, cannot be taken as evidence for deciding negatively the second of the above two questions.

But let us turn to this second question, for it raises what appears to be the fundamental issue in the present context. That issue is whether the analysis of "organic unities" necessarily involves the adoption of irreducible laws for such systems, and whether their mode of organization precludes the possibility of analyzing them from the so-called "additive point of view." The main difficulty in this connection is that of ascertaining in what way an "additive" analysis differs from one which is not. The contrast seems to hinge on the claim that the parts of a functional whole do not act independently of one another, so that any laws which may hold for such parts when they are not members of a functional whole cannot be assumed to hold for them when they actually are members. An "additive" analysis therefore appears to be one which

ent on p and T; and also this class of properties is an interdependent class of properties. Again, if S is an insulated conductor possessing a definite shape, R the curvature at any point, s the charge density at any region, and p the pressure at any region, then p is not dependent on R and s, and the properties p, R, and s do not form an interdependent class, though they do not form an independent class either. For this analysis, and further details involved in its elaboration, see the papers by Kurt Grelling, "A Logical Theory of Dependence," and Kurt Grelling and Paul Oppenheim, "Logical Analysis of 'Gestalt' and 'Functional Whole,'" reprinted for members of the Fifth International Congress for the Unity of Science held in Cambridge, Mass., 1939, from the *Journal of Unified Science*, Vol. 9. This volume of the Journal was a casualty of World War II and has never been published.

However, if now we define a system S to be a "functional whole" with respect to a class K of properties if K is an interdependent class, and also define S to be a "summative whole" if K is an independent class, two points should be noted. In the first place, whether a property will be said to depend on certain others will be affected in part by the degree of experimental precision with which values of the properties in question can be established. This is the point already made in the text. In the second place, though S may not be a functional whole in the sense defined, it need not therefore be a summative whole; for some properties in K may be dependent on the remaining ones, though not all are. Accordingly, there may be various "degrees" of interdependence of parts of a system.

accounts for the properties of a system in terms of assumptions about its constituents, where these assumptions are not formulated with specific reference to the characteristics of the constituents as elements in the system. A "nonadditive" analysis, on the other hand, seems to be one which formulates the characteristics of a system in terms of relations between certain of its parts as functioning elements in the system.

However, if this is indeed the distinction between these allegedly different modes of analysis, the difference is not one of fundamental principle. We have already noted that it does not seem possible to distinguish sharply between systems that are said to be "organic unities" and those which are not. Accordingly, since even the parts of summative wholes stand in relations of causal interdependence, an additive analysis of such wholes must include special assumptions about the actual organization of parts in those wholes when it attempts to apply some fundamental theory to them. There are certainly many physical systems, such as the solar system, a carbon atom, or a calcium fluoride crystal, which despite their complex form of organization lend themselves to an "additive" analysis; but it is equally certain that current explanations of such systems in terms of theories about their constituent parts cannot avoid supplementing these theories with statements about the special circumstances under which the constituents occur as elements in the systems. In any event, the mere fact that the parts of a system stand in relations of causal interdependence does not exclude the possibility of an additive analysis of the system.

The distinction between additive and nonadditive analysis is sometimes supported by the contrast commonly drawn between the particle physics of classical mechanics and the field approach of electrodynamics. It will therefore be instructive to dwell for a moment on this contrast. According to Newtonian mechanics, the acceleration induced in a particle by the action of other bodies is the vector sum of the accelerations which would be produced by each of these bodies were they acting singly; and the assumption underlying this principle is that the force exerted by one such body is independent of the force exerted by any other. In consequence, a mechanical system such as the solar system can be analyzed additively. In order to account for the characteristic behavior of the solar system as a whole, we need to know only the force (as a function of the distance) that each body in the system exerts separately on the other bodies.

But in electrodynamics the situation is different. For the action of an electrically charged body on another depends not only on their distances but also on their relative motions. Moreover, the effect of a change in motion is not propagated instantaneously, but with a finite velocity. Accordingly, the force on a charged body due to the presence of other such bodies is not determined by the positions and velocities of the latter but by the conditions of the electromagnetic "field" in the vicinity of the former. In consequence, since such a field cannot be regarded as a 'sum' of 'partial' fields, each due to a distinct charged particle, an electromagnetic system is commonly said to be incapable of an additive analysis. "The field can be treated adequately only as a unit," so it is claimed, "not as the sum total of the contributions of individual point charges." [28]

Two brief comments must be made on this contrast. In the first place, the notion of 'field' (as used in electromagnetic theory) undoubtedly represents a mathematical technique for analyzing phenomena that is different in many important respects from the mathematics employed in particle mechanics. The latter operates with discrete sets of state variables, so that the state of a system is specified by a finite number of coordinates; the field approach requires that the values of each of its state variables be specified for each point of a mathematically continuous space. And there are further corresponding differences in the kinds of differential equations, the variables that enter into them, and the limits between which mathematical integrations are performed.

But in the second place, though it is true that the electromagnetic field associated with a set of charged particles is not a 'sum' of partial fields associated with each particle separately, it is also true that the field is uniquely determined (i.e., the values of each state variable for each point of space are unequivocally fixed) by the set of charges, their velocities, and the initial and boundary conditions under which they occur. Indeed, in one technique employed within field theory, the electromagnetic field is simply an intermediary device for formulating the effects of electrically charged particles upon other such particles.[29] Accordingly, though it may be convenient to treat an electromagnetic field as a "unit," this convenience does not signify that the properties of the field cannot be analyzed in terms of assumptions concerning its constituents. And though the field may not be a 'sum' of partial fields in any customary sense, an electromagnetic system is a 'sum' in the special sense of the word proposed previously, namely, there is a theory about the constituents of these systems such that the relevant laws of the system can be deduced from the theory. In point of fact, if we take a final glance at the functional whole

[28] Peter G. Bergmann, *Introduction to the Theory of Relativity,* New York, 1942, p. 223. It would be pointless to ask in the present context whether any "physical reality" is to be assigned to electromagnetic fields or whether, as some writers maintain, electromagnetic fields are only a "mathematical fiction." It is sufficient to note that, whatever its "ultimate status," the field concept in physics represents a mode of analysis which can be distinguished from the particle approach.

[29] The technique to which reference is made is the device of retarded potentials. Cf. the remarks in Max Mason and Warren Weaver, *The Electromagnetic Field,* Chicago, 1929, Introduction.

illustrated by the charges on the insulated conductor, the law that formulates the distribution of charge densities can be deduced from assumptions concerning the behavior of charged particles.[30]

The upshot of this discussion of organic unities is that the question whether they can be analyzed from the additive point of view does not possess a general answer. Some functional wholes certainly can be analyzed in that manner, while for others (for example, living organisms) no fully satisfactory analysis of that type has yet been achieved. Accordingly, the mere fact that a system is a structure of dynamically interrelated parts does not suffice, by itself, to prove that the laws of such a system cannot be reduced to some theory developed initially for certain assumed constituents of the system. This conclusion may be meager; but it does show that the issue under discussion cannot be settled, as so much of extant literature on it assumes, in a wholesale and a priori fashion.

[30] Cf., for example, O. D. Kellogg, *Foundations of Potential Theory*, Berlin, 1929, Chap. 7.

Mechanistic Explanation and Organismic Biology

12

The analytic methods of the modern natural sciences are universally admitted to be competent for the study of all nonliving phenomena, even those which, like cosmic rays and the weather, are still not completely understood. Moreover, attempts at unifying special branches of physical science, by reducing their several systems of explanation to an inclusive theory, are generally encouraged and welcomed. During the past four centuries these methods have also been fruitfully employed in the study of living organisms; and many features of vital processes have been successfully explained in physicochemical terms. Outstanding biologists as well as physical scientists have therefore concluded that the methods of the physical sciences are fully adequate to the materials of biology, and many of these scientists have been confident that eventually the whole of biology would become simply a chapter of physics and chemistry.

But, despite the undeniable successes of physicochemical explanations in the study of living things, biologists of unquestioned competence continue to regard such explanations as not entirely adequate for the subject matter of biology. Most biologists are in general agreement that vital processes, like nonliving ones, occur only under determinate physicochemical conditions and form no exceptions to physicochemical laws. Some of them nevertheless maintain that the mode of analysis required for understanding living phenomena is fundamentally different from that which obtains in the physical sciences. Opposition to the systematic absorption of biology into physics and chemistry is sometimes based on the practical ground that it does not conform to the correct strategy of biological research. However, such opposition is often also supported by theoretical arguments which aim to show that the reduction of biology to physicochemistry is inherently impossible. Biology has long been an area in which crucial issues in the logic of explanation have been the subject of vigorous debate. In any event, it is instructive to examine some of the reasons biologists commonly advance for the claim that the logic of explanatory concepts in biology is distinctive of the science and that biology is an inherently autonomous discipline.

What are the chief supports for this claim?

1. Let us first dispose of two less weighty ones. Although it is difficult to formulate in precise terms the generic differences between the living and the nonliving, no one seriously doubts the obvious fact that there are such differences. Accordingly, the various "life sciences" are concerned with special questions that are patently unlike those with which physics and chemistry deal. In particular, biology studies the anatomy and physiology of living things, and investigates the modes and conditions of their reproduction, development, and decay. It classifies vital organisms into types or species; and it inquires into their geographic distribution, their lines of descent, and the modes and conditions of their evolutionary changes. Biology also analyzes organisms as structures of interrelated parts and seeks to discover what each part contributes to the maintenance of the organism as a whole. Physics and chemistry, on the other hand, are not specifically concerned with such problems, although the subject matter of biology also falls within the province of these sciences. Thus a stone and a cat when dropped from a height exhibit behaviors which receive a common formulation in the laws of mechanics; and cats as well as stones therefore belong to the subject matter of physics. Nevertheless, cats possess structural features and engage in processes in which physics and chemistry, at any rate in their current form, are not interested. Stated more formally, biology employs expressions referring to identifiable characteristics of living phenomena (such as 'sex,' 'cellular division,' 'heredity,' or 'adaptation') and asserts laws containing them (such as 'Hemophilia among humans is a sex-linked hereditary trait') that do not occur in the physical sciences and are not at present definable or derivable within these sciences. Accordingly, while the subject matter of biology and the physical sciences is not disparate, and though biology makes use of distinctions and laws borrowed from the physical sciences, the two sciences do not at present coincide.

It is no less evident that the techniques of observation and experimentation in biology are in general different from those current in the physical sciences. To be sure, some tools and techniques of observations, measurement, and calculation (such as lenses, balances, and algebra) are used in both groups of disciplines. But biology also requires special skills (such as those involved in the dissection of organic tissues) which serve no purpose in physics; and physics employs techniques (such as those needed for handling high-voltage currents) that are irrelevant in present-day biology. A physical scientist untrained in the special techniques of biological research is no more likely to perform a biological experiment successfully than is a pianist untutored in playing wind instruments likely to perform well on an oboe.

These differences between the special problems and techniques of the physical and biological sciences are sometimes cited as evidence for the inherent autonomy of biology, and for the claim that the analytical methods of physics are not fully adequate to the objectives of biological inquiry. However, though the differences are genuine, they certainly do not warrant such conclusions. Mechanics, electromagnetism, and chemistry, for example, are *prima facie* distinct branches of physical science, in each of which different special problems are pursued and different techniques are employed. But as we have seen, these are not sufficient reasons for maintaining that each of those divisions of physical science is an autonomous discipline. If there is a sound basis for the alleged absolute autonomy of biology, it must be sought elsewhere than in the differences between biology and the physical sciences which have been noted thus far.

2. What, then, are the weightier reasons which support that allegation? The main ones appear to be as follows. Vital processes have a *prima facie* purposive character; organisms are capable of self-regulation, self-maintenance, and self-reproduction, and their activities seem to be directed toward the attainment of goals that lie in the future. It is usually admitted that one can study and formulate the morphological characteristics of plants and animals in a manner comparable with the way physical sciences investigate the structural traits of nonliving things. Thus, the categories of analysis and explanation in physics are generally held to be adequate for studying the gross and minute anatomy of the human kidney, or the serial order of its development. But morphological studies are only one part of the biologist's task, since it also includes inquiry into the *functions* of structures in sustaining the activities of the organism as a whole. Thus, biology studies the role played by the kidney and its microscopic structure in preserving the chemical composition of the blood, and thereby in maintaining the whole body and its other parts in their characteristic activities. It is such manifestly "goal-directed" be-

havior of living things that is often regarded as requiring a distinctive category of explanation in biology.

Moreover, living things are organic wholes, not "additive systems" of independent parts, and the behavior of these parts cannot be properly understood if they are regarded as so many isolable mechanisms. The parts of an organism must be viewed as internally related members of an integrated whole. They mutually influence one another, and their behavior regulates and is regulated by the activities of the organism as a whole. Some biologists have argued that the coordinated, adaptive behavior of living organisms can be explained only by assuming a special vitalistic agent; others believe that an explanation is possible in terms of the hierarchical organization of internally related parts of the organism. But in either case, so it is frequently claimed, biology cannot dispense with the notion of organic unity; and in consequence it must use modes of analysis and formulation that are unmistakably *sui generis*.

Accordingly, two main features are commonly alleged to differentiate biology from the physical sciences in an essential way. One is the dominant place occupied by *teleological* explanations in biological inquiry. The other is the use of conceptual tools uniquely appropriate to the study of systems whose total behavior is not the resultant of the activities of independent components. We must now examine these claims in some detail.

I. *The Structure of Teleological Explanations*

Almost any biological treatise or monograph yields conclusive evidence that biologists are concerned with the functions of vital processes and organs in maintaining characteristic activities of living things. In consequence, if "teleological analysis" is understood to be an inquiry into such functions, and into processes directed toward attaining certain end-products, then undoubtedly teleological explanations are pervasive in biology. In this respect, certainly, there appears to be a marked difference between biology and the physical sciences. It would surely be an oddity on the part of a modern physicist were he to declare, for example, that atoms have outer shells of electrons in order to make chemical unions between themselves and other atoms possible. In ancient Aristotelian science, categories of explanations suggested by the study of living things and their activities (and in particular by human art) were made canonical for all inquiry. Since nonliving as well as living phenomena were thus analyzed in teleological terms—an analysis which made the notion of final cause focal—Greek science did not assume a fundamental cleavage between biology and other natural science. Modern science, on the other hand, regards final causes to be vestal virgins which bear no fruit in the study of physical and chemical phenomena; and, because of the association of teleological explanations with the doctrine that goals or ends of activity are dynamic agents in their own realizations, it tends to view such explanations as a species of obscurantism. But does the presence of teleological explanations in biology and their apparent absence from the physical sciences entail the absolute autonomy of the former? We shall try to show that it does not.

1. Quite apart from their association with the doctrine of final causes, teleological explanations are sometimes suspect in modern natural science because they are assumed to invoke purposes or ends-in-view as causal factors in natural processes. Purposes and deliberate goals admittedly play important roles in human activities, but there is no basis whatever for assuming them in the study of physicochemical and most biological phenomena. However, as has already been noted, a great many explanations counted as teleological do not postulate any purposes or ends-in-view; for explanations are often said to be "teleological" only in the sense that they specify the *functions* which things or processes possess. Most contemporary biologists certainly do not impute purposes to the organic parts of living things whose functions are investigated; most of them would probably also deny that the means-ends relationships discovered in the organization of living creatures are the products of some deliberate plan on the part of a purposeful agent, whether divine or in some other manner supranatural. To be sure, there are biologists who postulate psychic states as concomitants and even as directive forces of all organic behavior. But such biologists are in a minority; and they usually support their views by special considerations that can be distinguished from the facts of functional or teleological dependencies which most biologists do not hesitate to accept. Since the word 'teleology' is ambiguous, confusion and misunderstandings would doubtless be prevented if the word were eliminated from the vocabulary of biology. But biologists do use it, and say they are giving a teleological explanation when, for example, they explain that the function of the alimentary canal in vertebrates is to prepare ingested materials for absorption into the bloodstream. The crucial point is that when biologists do employ teleological language they are not necessarily committing the pathetic fallacy or lapsing into anthropomorphism.

We shall therefore assume that teleological (or functional) statements in biology normally neither assert nor presuppose in the materials under discussion either manifest or latent purposes, aims, objectives, or goals. Indeed, it seems safe to suppose that biologists would generally deny they are postulating any conscious or implicit ends-in-view even when they employ such words as 'purpose' in their functional analyses—as when the 'purpose' (i.e., the function) of kidneys in the pig is said to be that of eliminating various waste products from the bloodstream of the organism. On the other hand, we shall adopt as the mark of a teleological statement in biology, and as the feature that distinguishes such statements from nonteleological ones, the occurrence in the former but not in the latter of such typical locutions as 'the function of,' 'the purpose of,' 'for the sake of,' 'in order that,' and the like—more generally, the occurrence of expressions signifying a means-ends nexus.

Nevertheless, despite the *prima facie* distinctive character of teleological (or functional) explanations, we shall first argue that they can be reformulated, without loss of asserted content, to take the form of nonteleological ones, so that in an important sense teleological and nonteleological explanations are equivalent. To this end, let us consider a typical teleological statement in biology, for example, 'The function of chlorophyll in plants is to enable plants to perform photosynthesis (i.e., to form starch from carbon dioxide and water in the presence of sunlight).' This statement accounts for the presence of chlorophyll (a certain substance A) in plants (in every member S of a class of systems, each of which has a certain organization C of component parts and processes). It does so by declaring that, when a plant is provided with water, carbon dioxide, and sunlight (when S is placed in a certain "internal" and "external" environment E), it manufactures starch (a certain process P takes place yielding a definite product or outcome) only if the plant contains chlorophyll. The statement usually carries with it the additional tacit assumption that without starch the plant cannot continue its characteristic activities, such as growth and reproduction (it cannot maintain itself in a certain state G); but for the present we shall ignore this further claim.

Accordingly, the teleological statement is a telescoped argument, so that when the content is unpacked it can be rendered approximately as follows: When supplied with water, carbon dioxide, and sunlight, plants produce starch; if plants have no chlorophyll, even though they have water, carbon dioxide, and sunlight, they do not manufacture starch; hence, plants contain chlorophyll. More generally, a teleological statement of the form 'The function of A in a system S with organization C is to enable S in environment E to engage in process P' can be formulated more explicitly by: Every system S with organization C and in environment E engages in process P; if S with organization C and in environment E does not have A, then S does not engage in P; hence, S with organization C must have A.

It is clearly not relevant in the present context to inquire whether the premises in this argument are adequately supported by competent evidence. However, because the issue is sometimes raised in discussions of teleological explanations, at least passing notice deserves to be given to the question of whether chlorophyll is really necessary to plants and whether they could not manufacture starch (or other substances essential for their maintenance) by some alternative process not requiring chlorophyll. For, if the presence of chlorophyll is not actually necessary for the production of starch (or if plants can maintain themselves without the mechanism of photosynthesis), so it has been urged, the second premise in the above argument is untenable. The premise would then have to be modified; and in its emended form it would assert that chlorophyll is an element in a set of conditions that is *sufficient* (but not necessary) for the production of starch. In that case, however, the new argument with the emended premise would be invalid, so that the proposed teleological explanation of the presence of chlorophyll in plants would apparently be unsatisfactory.

This objection is in part well-taken. It is certainly *logically* possible that plants might maintain themselves without manufacturing starch, or that processes in living organisms might produce starch without requiring chlorophyll. Indeed, there are plants (the funguses) that can flourish without chlorophyll; and in general, there is more than one way of skinning a cat. On the other hand, the above teleological explanation of the occurrence of chlorophyll in plants is presumably concerned with living organisms having certain determinate forms of organization and definite modes of behavior—in short, with the so-called "green plants." Accordingly, although living organisms (plants as well as animals) capable of maintaining themselves without processes involving the operation of chlorophyll are both abstractly and physically possible, there appears to be no evidence whatever that in view of the limited capacities green plants possess as a consequence of their *actual* mode of organization, these organisms can live without chlorophyll.

Two important complementary points thus emerge from these considerations. In the first place, teleological analyses in biology (or in other sciences in which such analyses are pursued) are not explorations of merely logical possibilities, but deal with the actual functions of definite components in concretely given living systems. In the second place, on pain of failure to recognize the possibility of alternative mechanisms for achieving some end-product, and of unwittingly (and perhaps mistakenly) assuming that a process known to be indispensable in a given class of systems is also indispensable in a more inclusive class, a teleological explanation must articulate with exactitude both the character of the end-product and the defining traits of the systems manifesting them, relative to which the indicated processes are supposedly indispensable.

In any event, however, the above teleological account of chlorophyll, in its expanded form, is simply an illustration of an explanation that conforms to the deductive model, and contains no locution distinctive of teleological statements. Accordingly, the initial, unexpanded statement about chlorophyll appears to assert nothing that is not asserted by 'Plants perform photosynthesis only if they contain chlorophyll,' or alternatively by 'A necessary condition for the occurrence of photosynthesis in plants is the presence of chlorophyll.' These latter statements do not explicitly ascribe a function to chlorophyll, and in that sense are therefore not teleological formulations. If this example is taken as a paradigm, it seems that, when a function is ascribed to a constituent element in an organism, the content of the teleological statement is fully conveyed by another statement that is not explicitly teleological and that simply asserts a necessary (or possibly a necessary and sufficient) condition for the occurrence of a certain trait or activity of the organism. In the light of this analysis, therefore, a teleological explanation in biology indicates the *consequences* for a given biological system of a constituent part or process; the equivalent nonteleological formulation of this explanation, on the other hand, states some of the *conditions* (sometimes, but not invariably, in physicochemical terms) under which the system persists in its characteristic organization and activities. The difference between a teleological explanation and its equivalent nonteleological formulation is thus comparable to the difference between saying that *Y* is an effect of *X*, and saying that *X* is a cause or condition of *Y*. In brief, the difference is one of selective attention, rather than of asserted content.

This point can be reinforced by another consideration. If a teleological explanation had an asserted content different from the content of every conceivable nonteleological statement, it would be possible to cite procedures and evidence employed for establishing the former that differ from the procedures and evidence required for warranting the latter. But in point of fact there appear to be no such procedures and evidence. Consider, for example, the teleological statement 'The function of the leucocytes in human blood is to defend the body against foreign microorganisms.' Now whatever may be the evidence that warrants this statement, that evidence also confirms the nonteleological statement 'Unless human blood contains a sufficient number of leucocytes, certain normal activities of the body are impaired,' and conversely. If this is so, however, there is a strong presumption that the two statements do not differ in factual content. More generally, if, as seems to be the case, the conceivable evidence for any given teleological explanation is identical with the conceivable evidence for a certain nonteleological one, the conclusion appears inescapable that those statements cannot be distinguished with respect to what they *assert*, even though they are distinguishable in other ways.

2. This proposed equation of teleological and nonteleological explanations must nevertheless face a fundamental objection. Many biologists would perhaps admit that a teleological statement *implies* a certain nonteleological one; but some of them, at any rate, are prepared to maintain that the latter statement generally does not in turn imply the former one, and that in consequence the alleged equivalence between the statements does not in fact hold.

The claim that there is indeed no such equivalence can be forcefully presented as follows. If there were such an equivalence, not only could a teleological explanation be replaced by a nonteleological one, but conversely a nonteleological explanation could also be replaced by a teleological one. In consequence, the customary statements of laws and theories in the physical sciences would be translatable without change in asserted content into teleological formulations. In point of fact, however, modern physical science does not appear to sanction such reformulations. Indeed, most physical scientists would doubtless resist the introduction of teleological statements into their disciplines as a misguided attempt to reinstate the point of view of Greek and medieval science. For example, the statement 'The volume of a gas at constant temperature varies inversely with its pressure' is a typical physical law, which is entirely free of teleological connotations. If it were equivalent to a teleological statement, its equivalent (constructed on the model of the example adopted above as paradigmatic) would presumably be 'The function of a varying pressure in a gas at constant temperature is to produce an inversely varying volume of the gas,' or perhaps 'Every gas at constant temperature under a variable pressure alters its volume in order to keep the product of the pressure and the volume constant.' But most physicists would undoubtedly regard these formulations as preposterous, and at best as misleading. Accordingly, if no teleological statement can correctly translate a law of physics, the contention that for every teleological statement a logically equivalent nonteleological one can be constructed seems hardly tenable. There must therefore be some important difference between teleological and nonteleological statements, so the objection concludes, that the discussion has thus far failed to make explicit.

The difficulty just expounded cannot be disposed of easily. To assess it adequately, we must consider the type of subject matter in which teleological analyses are currently undertaken, and in which teleological explanations are not rejected ostensibly as a matter of general principle.

a. The attitude of physical scientists toward teleological formulations in their disciplines is doubtless as alleged in this objection. Nevertheless, this fact is not completely decisive on the point at issue. Two comments are in order which tend to weaken its critical force.

In the first place, it is not entirely accurate to maintain that the physical sciences never employ formulations that have at least the appearance of teleological statements. As is well known, some physical laws and theories are often expressed in so-called "isoperimetric" or "variational" form, rather than in the more familiar manner of numerical or differential equations. When laws and theories are expressed in this fashion, they strongly resemble teleological formulations, and have in fact been frequently assumed to express a teleological ordering of events and processes. For example, an elementary law of optics states that the angle of incidence of a light ray reflected by a surface is equal to the angle of reflection. However, this law can also be rendered by the statement that a light ray travels in such a manner that the length of its actual path (from its source to reflecting surface to its terminus) is the minimum of all possible paths. More generally, a considerable part of classical as well as contemporary physical theory can be stated in the form of "extremal" principles. These principles assert that the actual development of a system proceeds in such a manner as to minimize or maximize some magnitude which represents the possible configurations of the system.[1]

The discovery that the principles of mechanics can be given such extremal formulations was once considered as evidence for the operation of a divine plan throughout nature. This view was made prominent by Maupertuis, an eighteenth-century thinker who was perhaps the first to state mechanics in variational form; and it was widely accepted in the eighteenth and nineteenth centuries. Such theological interpretations of extremal principles are now almost universally recognized to be entirely gratuitous; and with rare exceptions, physicists today do not accept the earlier claim that extremal principles entail the assumption of a plan or purpose animating physical processes. The use of such principles in physical science nevertheless does show that the dynamical structure of physical systems can be formulated so as to make focal the effect of constituent elements and subsidiary processes upon certain global properties of the system taken as a whole. If physical scientists dislike teleo-

[1] Cf. A. D'Abro, *The Decline of Mechanism in Modern Physics*, New York, 1939, Chap. 18; Adolf Kneser, *Das Prinzip der kleinsten Wirkung*, Leipzig, 1928; Wolfgang Yourgrau and Stanley Mandelstam, *Variational Principles in Dynamics and Quantum Theory*, London, 1955.

It can in fact be shown that, when certain very general conditions are satisfied, all quantitative laws can be given an "extremal" formulation.

logical language in their own disciplines, it is not because they regard teleological notions in this sense as foreign to their task. Their dislike stems in some measure from the fear that, except when such teleological language is made rigorously precise through the use of quantitative formulations, it is apt to be misunderstood as connoting the operation of purposes.

In the second place, the physical sciences, unlike biology, are in general not concerned with a relatively special class of organized bodies, and they do not investigate the conditions making for the persistence of some selected physical system rather than of others. When a biologist ascribes a function to the kidney, he tacitly assumes that it is the kidney's contribution to the maintenance of the living animal which is under discussion; and he ignores as irrelevant to his primary interest the kidney's contribution to the maintenance of any other system of which it may also be a constituent. On the other hand, a physicist generally attempts to discuss the effects of solar radiation upon a wide variety of things; and he is reluctant to ascribe a "function" to the sun's radiation, because no one physical system of which the sun is a part is of greater interest to him than is any other such system. And similarly for the law relating the pressure and volume of a gas: if a physicist views with suspicion the formulation of this law in functional or teleological language, it is because (in addition to the reasons which have been or will be discussed) he does not regard it as his business to assign special importance, even if only by vague suggestion, to one rather than another consequence of varying pressures in a gas.

b. However, the discussion thus far can be accused, with some justice, of naïveté if not of irrelevance, on the ground that it has ignored completely the fundamental point, namely, the "goal-directed" character of organic systems. It is because living things exhibit in varying degrees adaptive and regulative structures and activities, while the systems studied in the physical sciences do not—so it is frequently claimed—that teleological explanations are peculiarly appropriate for biological systems but not for physical systems. Thus, because the solar system, or any other system of which the sun is a part, does not tend to persist in some integrated pattern of activities in the face of environmental changes, and because the constituents of the system do not undergo mutual adjustments so as to maintain this pattern in relative independence from the environment, it is preposterous to ascribe any function to the sun or to solar radiation. Nor does the fact that physics can state some of its theories in the form of extremal principles, so the objection continues, minimize the differences between biological and purely physical systems. It is true that a physical system develops in such a way as to minimize or maximize a certain magnitude which represents a property of the system as a whole. But physical systems are not organized to maintain, in the face of considerable alterations in their environment, some *particular* extremal values of such magnitudes, or to develop under widely varying conditions in the direction of realizing some particular values of such magnitudes.

Biological systems, on the other hand, do possess such organization, as a single example (which could be matched by an indefinite number of others) makes quite clear. The human body maintains many of its characteristics in a relatively steady state (or homeostasis) by complicated but coordinated physiological processes. Thus, the internal temperature of the body must remain fairly constant if it is not to be fatally injured. In point of fact, the temperature of the normal human being varies during a day only from about 97.3° F to 99.1° F, and cannot fall much below 75° F or rise much above 110° F without permanent injury to the body. However, the temperature of the external environment can fluctuate much more widely than this; and it is clear from elementary physical considerations that the body's characteristic activities would be profoundly impaired or curtailed unless it were capable of compensating for such environmental changes. But the body is indeed capable of doing just this; and in consequence its normal activities can continue, in relative independence of the temperature of the environment—provided, of course, that the environmental temperature does not fall outside a certain interval of magnitudes. The body achieves this homeostasis by means of a number of mechanisms, which serve as a series of defenses against shifts in the internal temperature. Thus, the thyroid gland is one of several that control the body's basal metabolic rate (which is the measure of the heat produced by combustion in various cells and organs); the heat radiated or conducted through the skin depends on the quantity of blood flowing through peripheral vessels, a quantity which is regulated by dilation or contraction of these vessels; sweating and the respiration rate determine the quantity of moisture that is evaporated, and so affect the internal temperature; adrenaline in the blood also stimulates internal combustion, and its secretion is affected by changes in the external temperature; and automatic muscular contractions involved in shivering are an additional source of internal heat. There are thus physiological mechanisms in the body that automatically preserve its internal temperature, despite disturbing conditions in the body's internal and external environment.[2]

Three separate questions, frequently confounded, are raised by such facts of biological organization. (1) Is it possible to formulate in gen-

[2] Cf. Walter B. Cannon, *The Wisdom of the Body*, New York, 1932, Chap. 12.

eral but fairly precise terms the distinguishing structure of "goal-directed" systems, but in such a way that the analysis is neutral with respect to assumptions concerning the existence of purposes or the dynamic operation of goals as instruments in their own realization? (2) Does the fact, if it is a fact, that teleological explanations are customarily employed only in connection with "goal-directed" systems constitute relevant evidence for deciding the issue of whether a teleological explanation is equivalent to some nonteleological one? (3) Is it possible to explain in purely physicochemical terms—that is, exclusively in terms of the laws and theories of current physics and chemistry—the operations of biological systems? This third question will not concern us for the present, though we shall return to it later; but the other two require our immediate attention.

i. Since antiquity there have been many attempts at constructing machines and physical systems simulating the behavior of living organisms in one respect or another. None of these attempts has been entirely successful, for it has not been possible thus far to manufacture in the workshop and out of inorganic materials any device that acts fully like a living body. Nevertheless, it has been possible to construct physical systems that are self-maintaining and self-regulating in respect to certain of their features, and which therefore resemble living organisms in at least this one important characteristic. In an age in which servomechanisms (governors on engines, thermostats, automatic airplane pilots, electronic calculators, radar-controlled anti-aircraft firing devices, and the like) no longer excite wonder, and in which the language of cybernetics and "negative feedbacks" has become widely fashionable, the imputation of "goal-directed" behavior to purely physical systems certainly cannot be rejected as an absurdity. Whether "purposes" can also be imputed to such physical systems, as some expounders of cybernetics claim,[3] is perhaps doubtful, though the question is in large measure a semantic one; and in any event, this further issue is not relevant to the present context of discussion. Moreover, it is worth noting that the possibility of constructing self-regulating physical systems does not, by itself, constitute a proof that the activities of living organisms can be explained in exclusively physicochemical terms. Nevertheless, the fact that such systems have been constructed does suggest that there is no sharp demarcation setting off the teleological organizations, often assumed to be

[3] Cf. Arturo Rosenblueth, Norbert Wiener, Julian Bigelow, "Behavior, Purpose and Teleology," *Philosophy of Science*, Vol. 10 (1943); Norbert Wiener, *Cybernetics*, New York, 1948; A. M. Turing, "Computing Machines and Intelligence," *Mind*, Vol. 59 (1950); Richard Taylor, "Comments on a Mechanistic Conception of Purposefulness," *Philosophy of Science*, Vol. 17 (1950), and the reply by Rosenblueth and Wiener with a rejoinder by Taylor in the same volume.

distinctive of living things, from the goal-directed organizations of many physical systems. At the minimum, that fact offers strong support for the presumption that the teleologically organized activities of living organisms and of their parts can be analyzed without requiring the postulation of purposes or goals as dynamic agents.

With the homeostasis of the temperature of the human body as an exemplar, let us now state in general terms the formal structure of systems possessing a goal-directed organization.[4] The characteristic feature of such systems is that they continue to manifest a certain state or property G (or that they exhibit a persistence of development "in the direction" of attaining G) in the face of a relatively extensive class of changes in their external environments or in some of their internal parts—changes which, if not compensated for by internal modification in the system, would result in the disappearance of G (or in an altered direction of development of the systems). The abstract pattern of organization of such systems can be formulated with considerable precision, although only a schematic statement of that pattern can be presented in what follows.

Let S be some system, E its external environment, and G some state, property, or mode of behavior that S possesses or is capable of possessing under suitable conditions. Assume for the moment (this assumption will eventually be relaxed) that E remains constant in all relevant respects, so that its influence upon the occurrence of G in S may be ignored. Suppose also that S is analyzable into a structure of parts or processes, such that the activities of a certain number (possibly all) of them are causally relevant for the occurrence of G. For the sake of simplicity, assume that there are just three such parts, each capable of being in one of several distinct conditions or states. The state of each part at any given time will be represented by the predicates 'A_x,' 'B_y,' and 'C_z,' respectively, with numerical values of the subscripts to indicate the different particular states of the corresponding parts. Accordingly, 'A_x,' 'B_y,' and 'C_z' are state variables, though they are not necessarily numerical variables, since numerical measures may not be available for representing the states of the parts; and the state of S that is causally relevant to G at any given time will thus be expressed by a specialization of the matrix '$(A_xB_yC_z)$.' The state variables may, however, be quite complex in form—for example, 'A_x' may represent the state of the peripheral blood vessels in a human body at a given time—and they may be either individual or

[4] The following discussion is heavily indebted to G. Sommerhoff, *Analytical Biology*, London, 1950. Cf. also Alfred J. Lotka, *Elements of Physical Biology*, New York, 1926, Chap. 25; W. Ross Ashby, *Design for a Brain*, London, 1953, and *An Introduction to Cybernetics*, London, 1956; and R. B. Braithwaite, *Scientific Explanation*, Cambridge, 1954, Chap. 10.

statistical coordinates. But in order to avoid inessential complications in exposition, we shall suppose that, whatever the nature of the state variables, in respect to the states they represent S is a deterministic system: the states of S change in such a way that, if S is in the same state at any two different moments, the corresponding states of S after equal lapses of time from those moments will also be the same.

One further important general assumption must also be made explicit. Each of the state variables can be assigned any particular "value" to characterize a state, provided the value is compatible with the known character of the part of S whose state the variable represents. In effect, therefore, the values of 'A_x' must fall into a certain restricted class K_A; and there are similar classes K_B and K_C for the permissible values of the other two state variables. The reason for these restrictions will be clear from an example. If S is the human body, and 'A_x' states the degree of dilation of peripheral blood vessels, it is obvious that this degree cannot exceed some maximum value; for it would be absurd to suppose that the blood vessels might have a mean diameter of, say, five feet. On the other hand, the possible values of one state variable *at a given time* will be assumed to be independent of the possible values of the other state variables *at that time.* This assumption must not be misunderstood. It does not assert that the value of a variable at one time is independent of the values of the other variables at some *other* time; it merely stipulates that the value of a variable at some specified instant is not a function of the values of the other variables *at that very same instant.* The assumption is the one normally made for state variables, and is introduced in part to avoid redundant coordinates of state. For example, the state variables in classical mechanics are the position and the momentum coordinates of a particle at an instant. Although the position of a particle at one instant will in general depend on its momentum (and position) at some *previous* time, the position at a given instant is not a function of the momentum *at that given instant.* If the position were such a function of the momentum, it is clear that the state of a particle in classical mechanics could be specified by just one state variable (the momentum), so that mention of the position would be redundant. In our present discussion we are similarly assuming that none of the state variables is dispensable, so that any combination of simultaneous values of the state variables yields a permissible specialization of the matrix '$(A_xB_yC_z)$,' provided that the values of the variables belong to the classes K_A, K_B, and K_C, respectively. This is tantamount to saying that, apart from the proviso, the state of S stipulated to be causally relevant to G must be so analyzed that the state variables employed for describing the state at a given time are mutually independent of one another.

Suppose now that if S is in the state $(A_0B_0C_0)$ at some initial time, then either S has the property G, or else a sequence of changes occurs in S as a consequence of which S will possess G at some subsequent time. Let us call such an initial state of S a "causally effective state with respect to G," or a "G-state" for short. Not every possible state of S need be a G-state, for one of the causally relevant parts of S may be in a certain state at a given time, such that *no* combination of possible states of the other parts will yield a G-state for S. Thus, suppose that S is the human body, G the property of having an internal temperature lying in the range 97° to 99° F, A_x again the state of the peripheral blood vessels, B_y the state of the thyroid glands, and C_z the state of the adrenal glands. It may happen that B_y assumes a value (e.g., corresponding to acute hyperactivity) such that for no possible values of A_x and C_z, respectively, will G be realized. It is of course also conceivable that no possible state of S is a G-state, so that in fact G is never realized in S. For example, if S is the human body and G the property of having an internal temperature lying in the range from 150° to 160°, then there is no G-state for S. On the other hand, more than one possible state of S may be a G-state. But if there is more than one possible G-state, then (since S has been assumed to be a deterministic system) the one that is realized at a given time is uniquely determined by the actual state of S at some previous time. The case in which there is more than one such possible G-state for S is of particular relevance to the present discussion, and we must now consider it more closely.

Assume once more that at some initial time t_0, the system S is in the G-state $(A_0B_0C_0)$. Suppose, however, that a change occurs in S so that in consequence A_0 is caused to vary, with the result that at time t_1 subsequent to t_0 the state variable 'A_x' has some other value. What value it will have at t_1 will in general depend on the particular changes that have taken place in S. We shall assume, however, that S will continue to be in a G-state at time t_1, provided that the values of 'A_x' at t_1 fall into a certain class K_A' (a subclass of K_A) containing more than one member, and provided also that certain further changes take place in the other state variables. To fix our ideas, suppose that A_1 and A_2 are the only possible members of K_A'; and assume also that neither $(A_1B_0C_0)$ nor $(A_2B_0C_0)$ is a G-state. In other words, if A_0 were changed into A_3 (a member of K_A but not of K_A'), S would no longer be in a G-state; but even though the new value of 'A_x' falls into K_A', if this were the only change in S the system would also no longer be in a G-state at time t_1. Let us assume, however, S to be so constituted that if A_0 is caused to vary so that the value of 'A_x' at time t_1 falls into K_A', there will be further compensatory changes in the values of some or all of the other state variables such that S continues to be in a G-state.

These further changes are stipulated to be of the following kind. If, as a concomitant of the change in A_0, the values of 'B_y' and 'C_z' at time t_1 fall into certain classes K_B' and K_C', respectively (where of course K_B' is a subclass, though not necessarily a proper subclass, of K_B, and K_C' is a subclass of K_C), then for each value in K_A' there is a unique pair of values, one member of the pair belonging to K_B' and the other to K_C', such that for those values S continues to be in a G-state at time t_1. These pairs of values can be taken to be elements in a certain class K_{BC}'. On the other hand, were the altered values of 'B_y' and 'C_z' not accompanied by the indicated changes in the value of 'A_x,' the system S would no longer be in a G-state at time t_1. In terms of the notation just introduced, accordingly, if at time t_1 the state variables of S have values such that two of them are members of a pair belonging to the class K_{BC}' while the value of the third variable is not the corresponding element in K_A', then S is not in a G-state. For example, suppose that, when A_0 changes into A_1, the initial G-state $(A_0B_0C_0)$ is changed into the G-state $(A_1B_1C_1)$, but that $(A_0B_1C_1)$ is not a G-state; and suppose also that when A_0 changes into A_2, the initial G-state is changed into the G-state $(A_2B_1C_2)$, with $(A_0B_1C_2)$ not a G-state. In this example, K_A' is the class (A_1, A_2); K_B' is the class (B_1); K_C' is the class (C_1, C_2); and K_{BC}' is the class of pairs $[(B_1, C_1), (B_1, C_2)]$, with A_1 corresponding to the pair (B_1, C_1) and A_2 to the pair (B_1, C_2).

Let us now bring together these various points, and introduce some definitions. Assume S to be a system satisfying the following conditions: (1) S can be analyzed into a set of related parts or processes, a certain number of which (say three, namely A, B, and C) are causally relevant to the occurrence in S of some property or mode of behavior G. At any time the state of S causally relevant to G can be specified by assigning values to a set of state variables 'A_x,' 'B_y,' and 'C_z.' The values of the state variables for any given time can be assigned independently of one another; but the possible values of each variable are restricted, in virtue of the nature of S, to certain classes of values K_A, K_B, and K_C, respectively. (2) If S is in a G-state at a given initial instant t_0 falling into some interval of time T, a change in any of the state variables will in general take S out of the G-state. Assume that a change is initiated in one of the state variables (say the parameter 'A'); and suppose that in fact the possible values of the parameter at time t_1 within the interval T but later than t_0 fall into a certain class K_A', with the proviso that if this were the sole change in the state of S the system would be taken out of its G-state. Let us call this initiating change a "primary variation" in S. (3) However, the parts A, B, and C of S are so related that, when the primary variation in S occurs, the remaining parameters also vary, and in point of fact their values at time t_1 fall into certain classes K_B' and K_C', respectively. These changes induced in B and C thus yield unique pairs of values for their parameters at time t_1, the pairs being elements of a class K_{BC}'. Were these latter changes the only ones in the initial G-state of S, unaccompanied by the indicated primary variation in S, the system would not be in a G-state at time t_1. (4) As a matter of fact, however, the elements of K_A' and K_{BC}' correspond to each other in a uniquely reciprocal manner, such that, when S is in a state specified by these corresponding values of the state variables, the system is in a G-state at time t_1. Let us call the changes in the state of S induced by the primary variation and represented by the pairs of values in K_{BC}' the "adaptive variations" of S with respect to the primary variation of S (i.e., with respect to possible values of the parameter 'A' in K_A'). Finally, when a system S satisfies all these assumptions for every pair of initial and subsequent instants in the interval T, the parts of S causally relevant to G will be said to be "directively organized during the interval of time T with respect to G"—or, more shortly, to be "directively organized," if the reference to G and T can be taken for granted.

This discussion of directively organized systems has been based on several simplifying assumptions. However, the analysis can be readily generalized for a system requiring the use of any number of state variables (including numerical ones), for changes in the state of a system that are initiated in more than one of the causally relevant parts of the system, and for continuous as well as discrete series of transitions from one G-state of a system to another.[5] Indeed, it is not difficult to develop

[5] When the state coordinates are assumed to be numerical, it is possible to formulate the conditions for a directively organized system as follows:

Let S be a system, G a trait of S, and 'x_1,' 'x_2,' . . . , 'x_n' the state variables for G. The variables are stipulated to be independent and continuous functions of the time; and superscripts will indicate their values at any given time t.

a) If S is a deterministic system with respect to G, the state of S at time t is uniquely determined by the state of S at some preceding time t_0. Hence

$$\begin{aligned} x_1^t &= f_1(x_1^{t_0}, \ldots, x_n^{t_0}, t - t_0) \\ &\ldots\ldots\ldots\ldots \\ x_i^t &= f_i(x_1^{t_0}, \ldots, x_n^{t_0}, t - t_0) \\ &\ldots\ldots\ldots\ldots \\ x_n^t &= f_n(x_1^{t_0}, \ldots, x_n^{t_0}, t - t_0) \end{aligned}$$

where the f_i's are single-valued functions of their arguments. Their first derivatives with respect to the time are also single-valued functions of their arguments and of no other functions of the time.

b) Since the special character of S imposes restrictions on the values of the state variables, the values of each variable 'x_i' will fall within an interval determined by a pair of numbers a_i and b_i. That is,

$$a_i \leq x_i \leq b_i$$

with $i \leq 1, 2, \ldots, n$, or alternately

$$x_i \in \Delta x_i$$

where Δx_i is some definite interval and 'ϵ' is the usual sign for class membership.

c) If S is in a G-state at a given time t falling into a given period of time T, the within this framework of analysis the notion of a system exhibiting self-regulatory behaviors with respect to several G's at the same time, alternative (and even incompatible) G's at different times, a set of G's constituting a hierarchy on the basis of some postulated scale of "relative importance," or more generally a set of G's whose membership changes with time and circumstance. But apart from complexity, nothing immediately relevant would be gained by such extensions of the analysis; and the schematic and incompletely general definitions that have been presented will suffice for our purposes.

It will in any case be clear from the above account that if S is directively organized, the persistence of G is in an important sense independent of the variations in any one of the causally relevant parts of S, provided that these variations do not exceed certain limits. For although by hypothesis the occurrence of G in S depends upon S being in a G-state, and therefore upon the state of the causally relevant parts of S, an alteration in the state of one of those parts may be compensated by induced changes in one or more of the other causally relevant parts,

state variable must satisfy a set of conditions or equations. That S is in a G-state at time t can be expressed by requirement:

$$\begin{aligned} g_1(x_1^t, \ldots, x_n^t) &= 0 \\ \cdots\cdots & \\ g_r(x_1^t, \ldots, x_n^t) &= 0 \end{aligned}$$

where each g_j ($j = 1, 2, \ldots, r$) is a function differentiable with respect to each of the state variables, and $r < n$.

d) The values of each state variable 'x_i^t' satisfying these equations defining a G-state of S fall into certain restricted intervals:

$$a_i \leq a_i^G \leq x_i^t \leq b_i^G \leq b_i$$

or alternately:

$$x_i^t \in \Delta x_i^G$$

where Δx_i^G falls into the interval Δx_i.

e) Assume that S is in a G-state at the initial time t_0 during the period T, and that a change takes place in the value of some state variable 'x_k' so that at time t later than t_0 in T its value is x_k^t. The condition that this change is a G-preserving change (so that $x_k^t \in \Delta x_k^G$), is that for each function g_j:

$$\frac{\partial g_j}{\partial x_k^{t_0}} = \frac{\partial g_j}{\partial x_1^t}\frac{\partial x_1^t}{\partial x_k^{t_0}} + \frac{\partial g_j}{\partial x_2^t}\frac{\partial x_2^t}{\partial x_k^{t_0}} + \cdots + \frac{\partial g_j}{\partial x_n^t}\frac{\partial x_n^t}{\partial x_k^{t_0}} = 0$$

f) The system S is directively organized with respect to G during T if, when such G-preserving changes occur in any given state variable 'x_k,' there are compensating variations in one or more of the other state variables. Accordingly, there must be at least one function g_j such that in the partial differential equations just mentioned there are at least two nonvanishing summands. That is, there are at least two summands in one or more of these equations such that

$$\frac{\partial g_j}{\partial x_i^t}\frac{\partial x_i^t}{\partial x_k^{t_0}} \neq 0$$

so as to preserve S in its assumed G-state. The *prima facie* distinctive character of so-called "goal-directed" or teleological systems is thus formulated by the stated conditions for a directively organized system. The above analysis has therefore shown that the notion of a teleological system can be explicated in a manner not requiring the adoption of teleology as a fundamental or unanalyzable category. What may be called the "degree of directive organization" of a system, or perhaps the "degree of persistence" of some trait of a system, can also be made explicit in terms of the above analysis. For the property G is maintained in S (or S persists in its development, which eventuates in G) to the extent that the range K_A' of the possible primary variations is associated with the range of induced compensatory changes K_{BC}' (i.e., the adaptive variations) such that S is preserved in its G-state. The more inclusive the range K_A' that is associated with such compensatory changes, the more is the persistence of G independent of variations in the state of S. Accordingly, on the assumption that it is possible to specify a measure for the range K_A', the "degree of directive organization" of S with respect to variations in the state parameter 'A' could be defined as the measure of this range.

We may now relax the assumption that the external environment E has no influence upon S. But in dropping this assumption, we merely complicate the analysis, without introducing anything novel into it. For suppose that there is some factor in E which is causally relevant to the occurrence of G in S, and whose state at any time can be specified by some determinate form of the state variable 'F_w.' Then the state of the enlarged system S' (consisting of S together with E) which is causally relevant to the occurrence of G in S is specified by some determinate form of the matrix '$(A_xB_yC_zF_w)$,' and the discussion proceeds as before. However, it is generally not the case that a variation in any of the internal parts of S produces any significant variation in the environmental factors. What usually is the case is that the environmental factors vary quite independently of the internal parts; they do not undergo changes which compensate for changes in the state of S; and, while a limited range of changes in them may be compensated by changes in S so as to preserve S in some G-state, most of the states which environmental factors are capable of assuming cannot be so compensated by changes in S. It is customary, therefore, to talk of the "degree of plasticity" or the "degree of adaptability" of organic systems in relation to their environments, and not conversely. However, it is possible to define these notions without special reference to organic systems, in a manner analogous to the definition of the "degree of directive organization" of a system already suggested. Thus, suppose that the variations in the environmental state variable 'F,' assumed to be compensated by further changes in S

so as to preserve S in some G-state, all fall into the class K_F'. If an appropriate measure for the magnitude of this class could be devised, the "degree of plasticity" of S with respect to the maintenance of some G in relation to F could then be defined as equal to the measure of K_F'.

This must suffice as an outline of the abstract structure of goal-directed or teleological systems. The account given deliberately leaves undiscussed the detailed mechanisms involved in the operation of particular teleological systems; and it simply assumes that all such systems can in principle be analyzed into parts which are causally relevant to the maintenance of some feature in those systems, and which stand to each other and to environmental factors in determinate relations capable of being formulated as general laws. The discovery and analysis of such detailed mechanisms is the task of specialized scientific inquiry. Accordingly, since the above account deals only with what is assumed to be the common distinctive structure of teleological systems, it is also entirely neutral on such substantive issues as to whether the operations of all teleological systems can be explained in exclusively physicochemical terms. On the other hand, if the account is at least approximately adequate, it requires a positive answer to the question whether the distinguishing features of goal-directed systems can be formulated without invoking purposes and goals as dynamic agents.

There is, however, one further matter that must be briefly discussed. The definition of directively organized systems has been so stated that it can be used to characterize both biological and nonvital systems. It is in fact easy to cite illustrations for the definition from either domain. The human body with respect to homeostasis of its internal temperature is an example from biology; a building equipped with a furnace and thermostat is an example from physicochemistry. Nevertheless, although the definition is not intended to distinguish between vital and nonvital teleological systems—for the differences between such systems must be stated in terms of the specific material composition, characteristics, and activities they manifest—it *is* intended to set off systems having a *prima facie* "goal-directed" character from systems usually not so characterized. The question therefore remains whether the definition does achieve this aim, or whether on the contrary it is so inclusive that almost *any* system (whether it is ordinarily judged to be goal-directed or not) satisfies it.

Now there are certainly many physicochemical systems that are not ordinarily regarded as being "goal-directed" but that nevertheless appear to conform to the definition of directively organized systems proposed above. Thus, a pendulum at rest, an elastic solid, a steady electric current flowing through a conductor, a chemical system in thermodynamic equilibrium, are obvious examples of such systems. It seems therefore that the definition of directive organization—and in consequence the proposed analysis of "goal-directed" or "teleological" systems—fails to attain its intended objective. However, two comments are in order on the point at issue. In the first place, though we admittedly do distinguish between systems that are goal-directed and those which are not, the distinction is highly vague, and there are many systems which cannot be classified definitely as of one kind rather than another. Thus, is the child's toy sometimes known as the "walking beetle"—which turns aside when it reaches the edge of a table and fails to fall off, because an idle wheel is then brought into play through the action of an "antenna"—a goal-directed system or not? Is a virus such a system? Is the system consisting of members of some biological species that has undergone evolutionary development in a steady direction (e.g., the development of gigantic antlers in the male Irish elk), a goal-directed one? Moreover, some systems have been classified as "teleological" at one time and in relation to one body of knowledge, only to be reclassified as "nonteleological" at a later time, as knowledge concerning the physics of mechanisms improved. "Nature does nothing in vain" was a maxim commonly accepted in pre-Newtonian physics, and on the basis of the doctrine of "natural places" even the descent of bodies and the ascent of smoke were regarded as goal-directed. Accordingly, it is at least an open question whether the current distinction between systems that are goal-directed and those that are not invariably has an identifiable objective basis (i.e., in terms of differences between the actual organizations of such systems), and whether the *same* system may not often be classified in alternative ways depending on the perspective from which it is viewed and on the antecedent assumptions adopted for analyzing its structure.

In the second place, it is by no means certain that physical systems such as the pendulum at rest, which is not usually regarded as goal-directed, really do conform to the definition of "directively organized" systems proposed above. Consider a simple pendulum that is initially at rest and is then given a small impulse (say by a sudden gust of wind); and assume that apart from the constraints of the system and the force of gravitation the only force that acts on the bob is the friction of the air. Then on the usual physical assumptions, the pendulum will perform harmonic oscillations with decreasing amplitudes, and finally assume its initial position of rest. The system here consists of the pendulum and the various forces acting on it, while the property G is the state of the pendulum when it is at rest at the lowest point of its path of oscillation. By hypothesis, its length and the mass of the bob are fixed, and so is the force of gravitation acting on it, as well as the coefficient of damping; the variables are the impulsive force of the gust of wind, and the restoring force which operates on the bob as a consequence of the constraints of the system and of the presence of the gravitational field. However—and this is the crucial point—these two forces are *not* independent of one another. Thus, if the effective component of the former has a certain magnitude, the restoring force will have an equal magnitude with an opposite direction. Accordingly, if the state of the system at a given time were specified in terms of state variables which take these forces as values, these state variables would not satisfy one of the stipulated conditions for state variables of directively organized systems; for the value of one of them at a given time is uniquely determined by the value of the other at that same time. In short, the values of these proposed state variables at any given instant are not independent.[6] It therefore follows that the simple pendulum is not a directively organized system in the sense of the definition presented. Moreover, it is also possible to show in a similar manner that a number of other systems, generally regarded as nonteleological ones, fail to satisfy that definition. Whether one could show this for all systems currently so regarded is admittedly an open question. However, since there are at least some systems not usually characterized as teleological which must also be so characterized on the basis of the definition, the label of 'directively organized system' whose meaning the definition explicates does not apply to everything whatsoever, and it does not baptize a distinction without a difference. There are therefore some grounds for claiming that the definition achieves what it was designed to achieve, and that it formulates

[6] This can be shown in greater detail by considering the usual mathematical discussion of the simple pendulum. If l is the length of the pendulum, m the mass of its bob, g the constant force of gravity, k the coefficient of damping due to air resistance, t the time as measured from some fixed instant, and s the distance of the bob along its path of oscillation from the point of initial rest, the differential equation of motion of the pendulum (on the assumption that the amplitude of vibration is small) is

$$m\frac{d^2s}{dt^2} + k\frac{ds}{dt} + \frac{mg}{l}s = 0$$

If at time t_0 the pendulum is at rest, both s_0 and v_0 $\left[= \left(\frac{ds}{dt}\right)_0\right]$ are zero, so that

$$\left(m\frac{d^2s}{dt^2}\right)_0 = 0;$$

i.e., no unbalanced forces are acting on the bob. Suppose now that at time t_1 the bob is at s_1 with a velocity v_1; the restoring force will then be

$$\left(m\frac{d^2t}{dt^2}\right)_1 = -kv_1 - \frac{mg}{l}s_1$$

But an impulsive force F_1 communicated to the bob at time t_1 uniquely determines the velocity v_1 and the position s_1 of the bob at that time. Hence the restoring force can be calculated, so that it is uniquely determined by the impulsive force.

the abstract structure commonly held to be distinctive of "goal-directed" systems.

ii. We can now settle quite briefly the second question, on page 410, we undertook to examine, namely, whether the fact that teleological explanations are usually advanced only in connection with "goal-directed" systems affects the claim that, in respect to its asserted content, every teleological explanation is translatable into an equivalent nonteleological one. The answer is clearly in the negative, if such systems are analyzable as directively organized ones in the sense of the above definition. For on the supposition that the notion of a goal-directed system can be explicated in the proposed manner, the characteristics that ostensibly distinguish such systems from those not goal-directed can be formulated entirely in nonteleological language. In consequence, every statement about the subject matter of a teleological explanation can in principle be rendered in nonteleological language, so that such explanations together with all assertions about the contexts of their use are translatable into logically equivalent nonteleological formulations.

Why, then, does it seem odd to render physical statements such as Boyle's law in teleological form? The answer is plain, if indeed teleological statements (and in particular, teleological explanations) are normally advanced only in connection with subject matters that are assumed to be directively organized. The oddity does not stem from any difference between the explicitly asserted content of a physical law and of its purported teleologically formulated equivalent. A teleological version of Boyle's law appears strange and unacceptable because such a formulation would usually be construed as resting on the assumption that a gas enclosed in a volume is a directively organized system, in contradiction to the normally accepted assumption that a volume of gas is not such a system. In a sense, therefore, a teleological explanation does connote more than does its *prima facie* equivalent nonteleological translation. For the former presupposes, while the latter normally does not, that the system under consideration in the explanation is directively organized. Nevertheless, if the above analysis is generally sound, this "surplus meaning" of teleological statements can always be expressed in nonteleological language.

3. On the hypothesis that a teleological explanation can always be translated, with respect to what it explicitly asserts, into an equivalent nonteleological one, let us now make more explicit in what way two such explanations nevertheless do differ. The difference appears to be as follows: Teleological explanations focus attention on the culminations and products of specific processes, and in particular upon the contribu-

tions of various parts of a system to the maintenance of its global properties or modes of behavior. They view the operations of things from the perspective of certain selected "wholes" or integrated systems to which the things belong; and they are therefore concerned with characteristics of the parts of such wholes, only insofar as those traits of the parts are relevant to the various complex features or activities assumed to be distinctive of those wholes. Nonteleological explanations, on the other hand, direct attention primarily to the conditions under which specified processes are initiated or persist, and to the factors upon which the continued manifestations of certain inclusive traits of a system are contingent. They seek to exhibit the integrated behaviors of complex systems as the resultants of more elementary factors, frequently identified as constituent parts of those systems; and they are therefore concerned with traits of complex wholes almost exclusively to the extent that these traits are dependent on assumed characteristics of the elementary factors. In brief, the difference between teleological and nonteleological explanations, as has already been suggested, is one of emphasis and perspective in formulation.

If this account is sound, the use of teleological explanations in the study of directively organized systems is as congruent with the spirit of modern science as is the use of nonteleological ones. This conclusion is confirmed by an examination of two currently held assessments of teleological explanations, one suggesting a limit to the value of such explanations, the other objecting in principle to their use.

a. The claim has been advanced that, although teleological explanations are in general legitimate, they are useful only when the knowledge we happen to possess of directively organized systems is of a certain kind.[7] Our available information about the range of environmental changes to which such a system can make adaptive responses (i.e., about what we have called the "plasticity" of goal-directed systems) may have two sources. It may have the status simply of an extrapolation to a given system of inductive generalizations obtained from a direct experimental study of quite similar systems. For example, the knowledge we have at present concerning the plasticity of a particular human organism in maintaining its internal temperature in the face of changes in temperature of the environment is based on our familiarity with the adaptive responses of other human bodies. On the view under consideration, teleological explanations in such cases are valuable, since they enable us to predict certain future behaviors of a given system from our knowledge concerning the past behaviors of similar systems—future behaviors that

[7] R. B. Braithwaite, *Scientific Explanation*, pp. 333ff.

would otherwise not be predictable in the assumed state of our knowledge On the other hand, our information about the plasticity of a given system may have the status of a body of deductions from previously established causal laws concerning the mechanisms embodied in the system. In such cases, adaptive responses of a given system to environmental changes can be calculated with the help of general assumptions, and can be predicted without any familiarity with the past behaviors of similar systems. In consequence, teleological explanations in such cases are said to have little if any value.

Although the distinction between these two types of sources of available knowledge concerning the plasticity of directively organized systems is patently sound, it is nevertheless not evident why the line between valuable and useless teleological explanations should be drawn in the indicated manner. Questions about the value of an explanation are not decided by reference to the logical source of the explanatory premises, and can be answered only by examining the effective role an explanation plays in inquiry and in the communication of ideas. It is in any event far from certain that teleological explanations for goal-directed systems concerning which we possess theoretically based knowledge are invariably or normally regarded as otiose. For there are in fact many artificial self-regulating systems (such as engines with governors regulating their speed) whose plasticity can be deduced from general theoretical assumptions. Teleological explanations for various features of such systems nevertheless continue to fill many pages in technical treatises about those systems, and there is no good reason to suppose that the explanations are commonly regarded as so much worthless lumber.

b. It is sometimes objected, however, that teleological explanations are inexcusably parochial. They are based, so it is argued, on a tacit assumption that a special set of complex systems have a privileged status; and in consequence such explanations make focal the role of things and processes in maintaining just those systems and no others. Processes have no inherent termini, the objection continues, and cannot rightly be supposed to contribute exclusively to the maintenance of some unique set of wholes. It is therefore misleading to say, for example, that *the* function of the white cells in the human blood is to defend the human body against foreign microorganisms. This is undoubtedly *a* function of the leucocytes; and this particular activity may even be said to be *the* function of these cells from the perspective of the human body. But leucocytes are elements in other systems as well; for example, they are parts of the blood stream considered in isolation from the rest of the body, of the system composed of some virus colony together with these white cells, or of the more inclusive and complex solar system. These other systems are also capable of persisting in their "normal" organization and activities only under definite conditions; and, from the perspective of the maintenance of these numerous other systems, the leucocytes possess other functions.

One obvious reply to this objection is in the form of a *tu quoque*. It is as legitimate to focus attention on consequences, culminations, and uses as it is on antecedents, starting points, and conditions. Processes do not have inherent termini, but neither do they have absolute beginnings. Things and processes are in general not elements engaged in maintaining some exclusively unique whole, but neither are wholes analyzable into an exclusively unique set of constituents. It is nevertheless intellectually profitable in causal inquiries to focus attention on certain earlier stages in the development of a process rather than on later ones, and on one set of constituents of a system rather than on another set. Similarly, it is illuminating to select as the point of departure for the investigation of some problems certain complex wholes rather than others. Moreover, as we have seen, some things are parts of directively organized systems, but do not appear to be parts of more than one such system. The study of the unique functions of parts in such unique directively organized systems is therefore not a preoccupation that assigns without warrant a special importance to certain particular systems. On the contrary, it is an inquiry that is sensitive to fundamental and objectively identifiable differences in subject matter.

There is nevertheless a point to the objection. For the refractive influence of provincial human interests on the construction of teleological explanations is perhaps more often overlooked than it is in the case of nonteleological analyses. In consequence, certain end-products of processes and certain directions of change are frequently assumed to be inherently "natural," "essential," or "proper," while all others are then labeled as "unnatural," "accidental," or even "monstrous." Thus, the development of corn seeds into corn plants is sometimes said to be natural, while their transformation into the flesh of birds or men is asserted to be merely accidental. In a given context of inquiry, and in the light of the problem which initiates it, there may be ample justification for ignoring all but one direction of possible changes and all but one system of activities to whose maintenance things and processes contribute. But such disregard of other functions that things may have, and of other wholes of which things may be parts, does not warrant the conclusion that what is ignored is less genuine or natural than what receives selective attention.

4. One final point in connection with teleological explanations in biology must be briefly noted. As has already been mentioned, some biologists maintain that the distinctive character of biological explana-

tions appears in physiological inquiries, in which the functions of organs and vital processes are under investigation, even though most biologists are quite prepared to admit that no special categories of explanation are required in morphology or the study of structural traits. Accordingly, great stress has been placed by some writers on the contrast between structure and function, and on the difficulties in assessing the relative importance of each as a determinant of living phenomena. It is generally conceded that "the development of functions goes hand in hand with the development of structure," that living activity does not occur apart from a material structure, nor does vital structure exist save as a product of protoplasmic activity. In this sense, structure and function are commonly regarded as "inseparable aspects" of biological organization. Nevertheless, eminent biologists believe it is still an unresolved and perhaps insoluble problem "to what extent structures may modify functions or functions structures"; they regard the contrast between structure and function as presenting a "dilemma." [8]

But what is this contrast, why do its terms raise an apparently irresolvable issue, and what does one of its terms cover which allegedly requires a mode of analysis and explanation specific to biology? Let us first remind ourselves in what way a morphological study of some biological organ, say the human eye, differs from the corresponding physiological investigation. A structural account of the eye usually consists in a description of its gross and minute anatomy. Such an account therefore specifies the various parts of the organ, their shapes and relative spatial arrangements with respect to each other and other parts of the body, and their cellular and physicochemical compositions. The phrase "structure of the eye" therefore ordinarily signifies the spatial organization of its parts, together with the physicochemical properties of each part. On the other hand, a physiological account of the organ specifies the activities in which its various parts can or do participate, and the role these parts play in vision. For example, the ciliary muscles are shown to be capable of contracting and slackening, so that because of their connection with the suspensory ligament the curvature of the lens can be accommodated to near and far vision; or the lachrymal glands are identified as the sources of fluids which lubricate and cleanse the conjunctival membranes. In general, therefore, physiology is concerned with the char-

[8] Cf. Edwin G. Conklin, *Heredity and Environment,* Princeton, 1922, p. 32, and Edmund B. Wilson, *The Cell,* New York, 1925, p. 670. In a later volume Conklin declared that "the relation of mechanism to finalism is not unlike that of structure to function—they are two aspects of organization. The mechanistic conception of life is in the main a structural aspect, the teleological view looks chiefly to ultimate function. These two aspects of life are not antagonistic, but complementary."—*Man: Real and Ideal,* New York, 1943, p. 117.

acter, the order, and the consequences of the activities in which the parts of the eye may be engaged.

If this example is typical of the way biologists employ the terms, the contrast between structure and function is evidently a contrast between the *spatial* organization of anatomically distinguishable parts of an organ and the *temporal* (or spatiotemporal) organization of changes in those parts. What is investigated under each term of the contrasting pair is a mode of organization or a type of order. In the one case the organization is primarily if not exclusively a spatial one, and the object of the investigation is to ascertain the spatial distribution of organic parts and the modes of their linkage. In the other case the organization has a temporal dimension, and the aim of the inquiry is to discover sequential and simultaneous orders of change in the spatially ordered and linked parts of organic bodies. It is evident, therefore, that structure and function (in the sense in which biologists appear to use these words) are indeed "inseparable." For it is difficult to make sense of any supposition that a system of activities having a temporal organization is not also a system of spatially structured parts manifesting these activities. In any event, there is obviously no antithesis between an inquiry directed to the discovery of the spatial organization of organic parts and an inquiry addressed to ascertaining the spatiotemporal structures that characterize the activities of those parts.

A comparable distinction between types of inquiries can also be introduced in the physical sciences. Descriptive physical geography, for example, is concerned primarily with the spatial distribution and spatial relations of mountains, plains, rivers, and oceans; historical geology and geophysics, on the other hand, investigate the temporal and dynamic orders of change in which such geographic features are involved. Accordingly, if inquiries into structure and function were antithetical in biology, a comparable antithesis would also occur within the nonbiological sciences. Every inquiry involves discriminating selection from the great variety of patterns of relations embodied in a subject matter; and it is both convenient and unavoidable to direct some inquiries to one kind of pattern and other inquiries to different kinds. There seems to be no reason for generating a fundamental puzzle from the fact that living organisms exhibit both a spatial and a spatiotemporal structure of their parts.

What, then, is the unsolved or irresolvable issue raised by the biological distinction between structure and function? Two questions can be distinguished in this connection. It may be asked, in the first place, what spatial structures are required for the exercise of specified functions, and whether a change in the pattern of activities of an organism or of its parts is associated with any change in the distribution and spatial organization of the constituents of that system. This is obviously a matter to be settled by detailed empirical inquiry, and, though there are innumerable unsettled problems in this connection, they do not raise issues of fundamental principle. One school of philosophers and biological theorists, for example, maintains that the development of certain comparable organs in markedly different species can be explained only on the assumption of a "vital impulse" that directs evolution toward the attainment of some future function. Thus the fact that the eyes of the octopus and of man are anatomically similar, though the evolution of each species from eyeless ancestors has followed different lines of development, has been offered as evidence for the claim that no explanation of this convergence is possible in terms of the mechanisms of chance variation and adaptations. That fact has in consequence been used to support the view that there is an "undivided original vital impulse" which so acts on inert matter as to create appropriate organs for the function of vision.[9] But even this hypothesis, however vague and otherwise unsatisfactory it may be, involves in part factual issues; and if most biologists reject it, they do so largely because the available factual evidence supports more adequately a different theory of evolutionary development.

In the second place, one may ask just why it is that a given structure is associated with a certain set of functions, or conversely. Now this question may be understood as a demand for an explanation, perhaps in physicochemical terms, for the fact that when a living body has a given spatial organization of its parts it exhibits certain patterns of activities. When the question is so understood, it is far from being a preposterous one. Although we may not possess answers to it in most cases, we do have reasonably adequate answers in at least a few others, so that we have some ground for the presumption that our ignorance is not necessarily permanent. However, such explanations must contain as premises not only statements about the physicochemical constitution of the parts of a living thing and about the spatial organization of these parts, but also statements of physicochemical laws or theories. Moreover, at least some of these latter premises must assert connections between the spatial organization of physicochemical systems and the temporal patterns of activities. But if the question continues to be pressed, and an explanation is demanded for these latter connections as well, an impasse is finally reached. For the demand then in effect assumes that the temporal or causal structure of physical processes is deducible simply from the spatial organization of physical systems, or conversely; and neither assumption is in fact tenable.

[9] Cf. H. Bergson, *Creative Evolution,* New York, 1911, Chap. 1, and the brief but incisive critique of views similar to those of Bergson in George G. Simpson, *The Meaning of Evolution,* New Haven, 1949, Chap. 12. See also Theodosius Dobzhansky, *Evolution, Genetics and Man,* New York and London, 1955, Chap. 14.

It is possible, analogously, to give a quite accurate account of the spatial relations in which the various parts of a clock stand to one another. We can specify the sizes of its cogwheels, the location of the mainspring and the escapement wheel, and so on. But although such knowledge of the clock's spatial structure is indispensable, it is not sufficient for understanding how the clock will operate. We must also know the laws of mechanics, which formulate the temporal structure of the clock's behavior by indicating how the spatial distribution of its parts at one time is related to the distribution at a later time. However, this temporal structure cannot be deduced simply from the clock's spatial structure (or its "anatomy"), any more than its spatial structure at any given time can be derived from the general laws of mechanics. Accordingly, the question why a given anatomical structure is associated with specific functions may be irresolvable, not because it is beyond our capacities to answer it, but simply because the question in the sense in which it is intended asks for what is *logically* impossible. In short, anatomical structure does not *logically* determine function, though *as a matter of contingent fact* the specific anatomical structure possessed by an organism does set bounds to the kinds of activities in which the organism can engage. And conversely, the pattern of behavior exhibited by an organism does not *logically* imply a unique anatomical structure, though *in point of fact* an organism manifests specific modes of activity only when its parts possess a determinate anatomical structure of a definite kind.

It follows from these various considerations that the distinction between structure and function covers nothing that distinguishes biology from the physical sciences, or that necessitates the use in biology of a distinctive logic of explanation. It has not been the aim of the present discussion to deny the patent differences between biology and other natural sciences with respect to the role played by functional analyses. Nor has it been its aim to cast doubt on the legitimacy of such explanations in any domain in which they are appropriate because of the special character of the systems investigated. The objective of the discussion has been to show only that the prevalence of teleological explanations in biology does not constitute a pattern of explanation incomparably different from those current in the physical sciences, and that the use of such explanations in biology is not a sufficient reason for maintaining that this discipline requires a radically distinctive logic of inquiry.

II. *The Standpoint of Organismic Biology*

Vitalism of the substantive type advocated by Driesch and other biologists during the preceding century and the earlier decades of the present one is now almost entirely a dead issue in the philosophy of biology.

The issue has ceased to be focal, perhaps less as a consequence of the methodological and philosophical criticisms to which vitalism has been subjected, than because of the sterility of vitalism as a guide in biological research and the superior heuristic value of other approaches to the study of vital phenomena. Nevertheless, the historically influential Cartesian conception of biology as simply a chapter of physics continues to meet resistance. Many outstanding biologists who find no merit in vitalism are equally dubious about the validity of the Cartesian program; and they sometimes advance what they believe are conclusive reasons for affirming the irreducibility of biology to physics and the intrinsic autonomy of biological method. The standpoint from which this antivitalistic and yet antimechanistic thesis is currently advanced commonly carries the label of "organismic biology." The label covers a variety of special biological doctrines that are not always mutually compatible. Nonetheless, the doctrines falling under it generally share the common premise that explanations of the "mechanistic" type are not appropriate for vital phenomena. We shall now examine the main contentions of organismic biology.

1. Although organismic biologists deny the suitability if not always the possibility of "mechanistic theories" for vital processes, it is frequently not clear what it is they are protesting against. But such unclarity can undoubtedly be matched by the ambiguity that often marks the statements of aims and programs by professed "mechanists" in biology. As we had occasion to note in an earlier chapter, the word "mechanism" has a variety of meanings, and "mechanists" in biology as well as their opponents take few pains to make explicit the sense in which they employ it. There are biologists who profess themselves to be mechanists simply in the broad sense that they believe that vital phenomena occur in determinate orders and that the conditions for their occurrence are spatiotemporal structures of bodies. But such a view is compatible with the outlook of all schools in biology, with the exception of the vitalists and radical indeterminists; and in any case, when mechanism in biology is so understood, no issue divides those who profess it from most organismic biologists. There have also been biologists who proclaimed themselves to be mechanists in the sense that they maintained that all vital phenomena were explicable exclusively in terms of the science of mechanics (more specifically, in terms of either pure or unitary mechanical theories in the sense of Chapter 7), and who therefore believed living things to be "machines" in the original meaning of this word. It is doubtful, however, whether any biologists today are mechanists in this sense. Physicists themselves have long since abandoned the seventeenth-century hope that a universal science of nature could be devel-

oped within the framework of the fundamental ideas of classical mechanics. And it is safe to say that no contemporary biologist subscribes literally to the Cartesian program of reducing biology to the science of mechanics, and especially to the mechanics of contact action.

In any event, most biologists today who call themselves mechanists profess a view that is at once much more specific than the general thesis of causal determinism, and much less restrictive than the one which identifies a mechanistic explanation with an explanation in terms of the science of mechanics. A mechanist in biology, we shall assume, is one who believes, as did Jacques Loeb, that all living processes "can be unequivocally explained in physicochemical terms,"[10] that is, in terms of theories and laws which by common consent are classified as belonging to physics and chemistry. However, biological mechanism so understood must not be taken to deny that living bodies have highly complex organizations. On the contrary, most biologists who adopt such a standpoint usually note quite emphatically that the activities of living bodies are not explicable by analyzing "merely" their physical and chemical compositions without taking into account their "ordered structures or organization." Thus, Loeb's characterization of a living body as a "chemical machine" is an obvious recognition of such organization. It is recognized even more explicitly by E. B. Wilson, who declares, after defining the "development" of germ plasm as the totality of operations by which the germ gives rise to its typical product, that the particular course of this development

> is determined (given the normal conditions) by the specific 'organization' of the germ-cells which form its starting-point. As yet we have no adequate conception of this organization, though we know that a very important part of it is represented by the nucleus. . . . Its nature constitutes one of the major unsolved problems of nature. . . . Nevertheless the only available path toward its exploration lies in the mechanistic conception that somehow the organization of the germ-cell must be traceable to the physico-chemical properties of its component substances and the specific configurations which they may assume.[11]

If such is the content of current biological mechanism, and if organismic biologists, like mechanists, reject the postulation of nonmaterial "vitalistic" agents whose operations are to explain vital processes, in what way do the approach and content of organismic biology differ from those of mechanism? The main points of difference, as noted by organismic biologists themselves, appear to be the following:

[10] Jacques Loeb, *The Mechanistic Conception of Life,* Chicago, 1912.

[11] E. B. Wilson, *op. cit.,* p. 1037, quoted by kind permission of The Macmillan Company, New York.

a. It is a mistake to suppose that the sole alternative to vitalism is mechanism. There are sectors of biological inquiry in which physicochemical explanations play little or no role at present, and a number of biological theories have been successfully exploited which are not physicochemical in character. For example, there is available an impressive body of experimental knowledge concerning embryological processes, though few of the regularities that have been discovered can be explained at present in exclusively physicochemical terms; and neither the theory of evolution even in its current forms, nor the gene theory of heredity, is based on any definite physicochemical assumptions concerning vital processes. It is certainly not inevitable that mechanistic explanations will eventually prevail in these domains; and, since in any event these domains are now being fruitfully explored without any necessary commitment to the mechanistic thesis, organismic biologists possess at least some ground for their doubts concerning the ultimate triumph of that thesis in all sectors of biology. For just as physicists may be warranted in holding that some branch of physics (e.g., electromagnetic theory) is not reducible to some other branch of the science (e.g., to mechanics), so an organismic biologist may be warranted in espousing an analogous view with respect to the relation of biology to the physical sciences. Thus there is a genuine alternative in biology to both vitalism and mechanism—namely, the development of systems of explanation that employ concepts and assert relations neither defined in nor derived from the physical sciences.

b. However, organismic biologists generally claim far more than this. Many of them also maintain that the analytic methods of the physicochemical sciences are intrinsically unsuited to the study of living organisms; that the central problems connected with vital processes require a distinctive mode of approach; and that, since biology is inherently irreducible to the physical sciences, mechanistic explanations must be rejected as the ultimate goal of biological research. One reason commonly advanced for this more radical thesis is the "organic" nature of biological systems. Indeed, perhaps the dominant theme upon which the writings of organismic biologists play so many variations is the "integrated," "holistic," and "unified" character of a living thing and its activities. Living creatures, in contrast to nonliving systems, are not loosely jointed structures of independent and separable parts, are not assemblages of tissues and organs standing in merely external relations to one another. Living creatures are "wholes" and must be studied as "wholes"; they are not mere "sums" of isolable parts, and their activities cannot be understood or explained if they are assumed to be such "sums." But mechanistic explanations construe living organisms as "machines" possessing independent parts, and thereby adopt an "additive" point of view in analyzing vital phenomena. Accordingly, since the action of the whole organism "has a certain unifiedness and completeness" which is left out of account in the course of analyzing it into its elementary processes, E. S. Russell concludes that "the activities of the organism as a whole are to be regarded as of a different order from physico-chemical relations, both in themselves and for the purposes of our understanding." [12] Therefore biology must observe two "cardinal laws of method": "The activity of the whole cannot be fully explained in terms of the activities of the parts isolated by analysis"; and "No part of any living entity and no single process of any complex organic unity can be fully understood in isolation from the structure and activities of the organism as a whole." [13]

c. An additional though closely related point which organismic biologists stress is the "hierarchical organization" of living bodies and processes. Thus, a cell is known to be a structure of various constituents, such as the nucleus, the Golgi bodies, and the membranes, each of which may be analyzable into other parts and these in turn into still others, so that the analysis presumably terminates in molecules, atoms, and their "ultimate" parts. But in multicellular organisms the cell is also only an element in the organization of a tissue, the tissue is a part of some organ, the organ a member of an organ system, and the organ system a constituent in the integrated organism. It is patent that these various "parts" do not occur at the same "level" of organization. In consequence, organismic biologists place great stress on the fact that an animate body is not a system of parts homogeneous in complexity of organization, but that on the contrary the "parts" into which an organism is analyzed must be distinguished according to the different levels of some particular type of hierarchical structure (there may be several such types) to which the parts belong. Now organismic biologists do not deny that physicochemical explanations are possible for the activities of parts on the "lower" levels of a hierarchy. Nor do they deny that the physicochemical properties of the parts on lower levels "condition" or "limit" in various ways the occurrence and modes of action of higher levels of organization. They do deny, on the other hand, that the processes found

[12] E. S. Russell, *The Interpretation of Development and Heredity,* Oxford, 1930, pp. 171-72.

[13] *Ibid.,* pp. 146-47. Similar statements of the central tenet of organismic biology will be found in Russell's *Directiveness of Organic Activities,* Cambridge, England, 1945, esp. Chaps. 1 and 7; Ludwig von Bertalanffy, *Theoretische Biologie,* Berlin, 1932, Chap. 2; his *Modern Theories of Development,* Oxford, 1933, Chap. 2; and his *Problems of Life,* New York and London, 1952, Chaps. 1 and 2; and W. E. Agar, *The Theory of the Living Organism,* Melbourne and London, 1943.

at higher levels of a hierarchy are "caused" by, or are fully explicable in terms of, lower-level properties. Biochemistry is acknowledged to be the study of the "conditions" under which cells and organisms act the way they do. Organismic biology, on the other hand, investigates the activities of the whole organism "regarded as conditioned by, but irreducible to, the modes of action of lower unities." [14]

We must now examine these alleged differences between the organismic and the mechanistic approaches to biology, and attempt to assess the claim that the mechanistic approach is generally inadequate to biological subject matter.

2. At first blush, the sole issues raised by organismic biology are those we have already discussed in connection with the doctrine of emergence and the reduction of one science to another. In point of fact, other questions are also involved. But to the extent that the issues are those of reduction, we can dispose of them quite rapidly.

Let us first remind ourselves of the two formal conditions, examined at some length in the preceding chapter, that are necessary and sufficient for the reduction of one science to another. When stated with special reference to biology and physicochemistry, they are as follows:

a. *The condition of connectability.* All terms in a biological law that do not belong to the primary science (such as 'cell,' 'mitosis,' or 'heredity') must be "connected" with expressions constructed out of the theoretical vocabulary of physics and chemistry (out of terms such as 'length,' 'electric charge,' 'free energy,' and the like). These connections may be of several kinds. The meanings of the biological expressions may be analyzable, and perhaps even explicitly definable, in terms of physicochemical ones, so that in the limiting case the biological expressions are eliminable in favor of the physicochemical terms. An alternative mode of connection is that biological expressions are associated with physicochemical ones by some type of coordinating definition, so that the connections have the logical status of conventions. Finally, and this is the more frequent case, the biological terms may be connected with physicochemical ones on the strength of empirical assumptions, so that the suf-

[14] Russell, *The Interpretation of Development and Heredity*, p. 187. For an analogous view, cf. Ludwig von Bertalanffy and Alex B. Novikoff, "The Conception of Integrative Levels and Biology," *Science*, Vol. 101 (1945), pp. 209-15, and the discussion of this article in the same volume, pp. 582-85 and in Vol. 102 (1945), pp. 405-06. A careful and sober analysis of the nature of hierarchical organization in biology and of its import for the possibility of mechanistic explanation is given in J. H. Woodger, *Biological Principles*, New York, 1929, Chap. 6, and in Woodger's "The 'Concept of Organism' and the Relation between Embryology and Genetics," *Quarterly Review of Biology*, Vol. 5 (1930), and Vol. 6 (1931).

ficient conditions (and possibly the necessary ones as well) for the occurrence of whatever is designated by the biological terms can be stated by means of the physicochemical expressions. Thus, if the term 'chromosome' can be associated in neither of the first two ways with some expression constructed out of the theoretical vocabulary of physics and chemistry, then it must be possible to state in the light of an assumed law the truth-conditions for a sentence of the form 'x is a chromosome' entirely by means of a sentence constructed out of that vocabulary.

b. *The condition of derivability.* Every biological law, whether theoretical or experimental, must be logically derivable from a class of statements belonging to physics and chemistry. The premises in these deductions will contain an appropriate selection from the theoretical assumptions of the primary discipline, as well as statements formulating the associations between biological and physicochemical terms required by the condition of connectability. In general, some of the premises will state in the vocabulary of the primary science the boundary conditions or specialized spatiotemporal configurations under which the theoretical assumptions are being applied.

As was shown in the preceding chapter, the condition of derivability cannot be fulfilled unless the condition of connectability is satisfied. It is beyond dispute, however, that the task of satisfying the first of these conditions for biology is still far from completed. We do not know at present, for example, the detailed chemical composition of chromosomes in living cells. We are therefore unable to state in exclusively physicochemical terms the conditions for the occurrence of those organic parts, and hence to state in such terms the truth-conditions for the application of the word 'chromosome.' And a fortiori we are not able at present to formulate in physicochemical language the structure of any of the systems, such as cell nucleus, cell, or tissue, of which chromosomes are themselves parts. Accordingly, in the current state of biological knowledge it is logically impossible to deduce the totality of biological laws and theories from purely physicochemical assumptions. In short, biology is not at present simply a chapter of physics and chemistry.

Organismic biologists are therefore on firm ground in maintaining that mechanistic explanations of all biological phenomena are currently impossible, and will remain impossible until the descriptive and theoretical terms of biology can be shown to satisfy the first condition for the reduction of that science to physics and chemistry—that is, until the composition of every part or process of living things, and the distribution and arrangement of their parts at any time, can be exhaustively specified in physicochemical terms. Moreover, even if this condition were realized, the triumph of the mechanistic standpoint would not thereby be assured. For as we have already shown, the satisfaction of the condition of connectability is a necessary but in general not a sufficient requirement for the absorption of biology into physics and chemistry. Although the connectability condition might be fulfilled, there would still remain the question whether all biological laws are deducible from the current theoretical assumptions of these physical sciences. The answer to this question is conceivably in the negative, since physicochemical theory in its present form may be insufficiently powerful to permit the derivation of various biological laws, even if these laws were to contain only terms properly linked with expressions belonging to those primary disciplines. It should also be noted that, even if both formal conditions for the reducibility of biology were satisfied, the reduction might nevertheless have little if any scientific importance, for the reason that some of the conditions previously labeled "nonformal" might not be adequately realized.

On the other hand, the facts cited and the argument thus far examined do not warrant the conclusion that biology is *in principle* irreducible to the physical sciences. The task facing such a proposed reduction is admittedly a most difficult one; and it undoubtedly impresses many students as one which, if not utterly hopeless, is at present not worth pursuing. However, no *logical* contradiction has yet been exhibited to the supposition that both the formal and nonformal conditions for the reduction of biology may some day be fulfilled. We can therefore terminate this part of the discussion with the conclusion that the question whether biology is reducible to physicochemistry is an open one, that it cannot be settled by a priori argument, and that an answer to it can be provided only by further experimental and logical inquiry.

3. Let us next turn to the argument for the inherent "autonomy" of biology based on the fact that living systems are hierarchically organized. The burden of the argument, as we have seen, is that properties and modes of behavior occurring on a higher level of such a hierarchy cannot in general be explained as the resultants of properties and behaviors exhibited by isolable parts belonging to lower levels of an organism's structure.

There is no serious dispute among biologists over the thesis that the parts and processes into which living organisms are analyzable can be classified in terms of their respective loci into hierarchies of various types, such as the essentially spatial hierarchy mentioned earlier. Nor is there disagreement over the contention that the parts of an organism belonging to one level of a hierarchy frequently exhibit forms of relatedness and of activity not manifested by organic parts belonging to another level. Thus, a cat can stalk and catch mice; but though the continued beating of its heart is a necessary condition for these activities, the cat's heart cannot perform these feats. Again, the heart can pump blood by contracting and expanding its muscular tissues, although no single tissue can keep the blood in circulation; and no tissue is able to divide by fission, even though its constituent cells may have this property. Such examples suffice to establish the claim that modes of behavior appearing at higher levels of a hierarchically organized system are not explained by merely listing each of the various lower-level parts and processes of the system as an aggregate of isolated and unrelated elements. Organismic biologists do not deny that the occurrence of higher-level traits in hierarchically structured living organisms is contingent upon the occurrence, at different levels of the hierarchy, of various component parts related in definite ways. But they do deny, and with apparent good reason, that statements formulating the traits exhibited by components of an organism, when the components are not parts of an actually living organism, can adequately explain the behavior of the living system that contains those components as parts related in complex ways to other elements in a hierarchically structured whole.

But do these admitted facts establish the contention that mechanistic explanations are either impossible or unsuitable for biological subject matter? It should be noted that various forms of hierarchical organization are exhibited by the materials of physics and chemistry, and not only by those of biology Our current theories of matter assume atoms to be structures of electric charges, molecules to be organizations of atoms, solids and liquids to be complex systems of molecules. Moreover, the available evidence indicates that elements at different levels of this hierarchy exhibit traits which their component parts do not invariably possess. However, these facts have not stood in the way of establishing comprehensive theories for the more elementary physical particles and processes, in terms of which it has been possible to account for some if not all of the physicochemical properties exhibited by objects having a more complex organization. To be sure, we do not possess at present a comprehensive and unified theory competent to explain the full range even of purely physicochemical phenomena occurring on various levels of organization. Whether such a theory will ever be achieved is certainly an open question. It is also relevant to note in this connection that biological organisms are "open systems," never in a state of "true equilibrium" but at best only in a steady state of "dynamic equilibrium" with their environment, because they continually exchange material components and not only energies with the latter.[15] In this respect, living organisms are unlike the "closed systems" usually studied in current phys-

[15] L. von Bertalanffy, *Problems of Life,* Chap. 4.

ical science. Indeed, an adequate theory for physicochemical processes in open systems—for example, a thermodynamics competent to deal with systems in nonequilibrium as well as equilibrium states—is at present only in an early stage of development. Nevertheless, the circumstance remains that we can now account for some characteristics of fairly complex systems with the help of theories formulated in terms of relations between relatively more simple ones, for example, the specific heats of solids in terms of quantum theory, or the changes in phase of compounds in terms of the thermodynamics of mixtures. This circumstance must make us pause in accepting the conclusion that the hierarchical organization of living systems by itself precludes a mechanistic explanation for their traits.

Let us, however, examine in greater detail some of the organismic arguments on this issue. One of them has been persuasively stated by J. H. Woodger, whose careful but sympathetic analyses of organismic notions are important contributions to the philosophy of biology. Woodger maintains that it is essential to distinguish between chemical *entities* and chemical *concepts;* he believes that, if the distinction is kept in mind, it no longer appears plausible to assume that a thing can be satisfactorily described in terms of chemical concepts exclusively, merely because the thing is held to be composed of chemical entities. "A lump of iron," Woodger declares, "is a chemical entity, and the word 'iron' stands for a chemical concept. But suppose that the iron has the form of a poker or a padlock, then although the iron is still chemically analyzable in the same way as before it cannot still be fully described in terms of chemical concepts. It now has an organization above the chemical level." [16]

Now there is no doubt that many of the uses to which iron pokers or padlocks may be put are not, and may never be, described in purely physicochemical terms. But does the fact that a piece of iron has the form of a poker or of a padlock stand in the way of explaining an extensive class of its properties and modes of behavior in exclusively physicochemical terms? The rigidity, tensile strength, and thermal properties of the poker, or the mechanism and the qualities of endurance of the padlock, are certainly explicable in such terms, even if it may not be necessary or convenient to invoke a microscopic physical theory to account for all these traits. Accordingly, the mere fact that a piece of iron has a certain organization does not preclude the possibility of a physico-

[16] J. H. Woodger, *Biological Principles,* p. 263. Woodger continues, "In the same way an organism is a physical entity in the sense that it is one of the things we become aware of by means of the senses, and is a chemical entity in the sense that it is capable of chemical analysis just as is the case with any other physical entity, but it does not follow from this that it can be fully and satisfactorily described in chemical terms."

chemical explanation for some of the characteristics it exhibits as an organized object.

Some organismic biologists maintain that, even if we were able to describe in minute detail the physicochemical composition of a fertilized egg, we would nevertheless still be unable to explain mechanistically the fact that such an egg normally segments. In the view of E. S. Russell, for example, we might be able on the stated supposition to formulate the physicochemical conditions for segmentation, but we would be unable to "explain the course which development takes." [17]

This claim raises some of the previously discussed issues associated with the distinction between structure and function. But quite apart from these issues, the claim appears to rest on a misunderstanding if not on a confusion. It is cogent to maintain that a knowledge of the physicochemical composition of a biological organism does not suffice to explain mechanistically its modes of action—any more than an enumeration of the parts of a clock together with a description of their spatial distribution and arrangement suffices to explain or to predict the behavior of the timepiece. To make such an explanation, we must also assume some theory or set of laws (in the case of the clock, the theory of mechanics) which formulates the way certain elements act when they occur in some initial distribution and arrangement, and which permits the calculation (and hence the prediction) of the subsequent development of that organized system of elements. Moreover, it is conceivable that, despite our assumed ability at some given stage of scientific knowledge to describe in full detail the physicochemical composition of a living thing, we might nevertheless be unable to deduce from the established physicochemical theories of the day the course of the organism's development. In short, it is conceivable that the first but not the second formal condition for the reducibility of one science to another is satisfied at a given time. It is a misunderstanding, however, to suppose that a fully codified explanation in the natural sciences can consist only of instantial premises formulating initial and boundary conditions but containing no statements of law or theory. It is an elementary blunder to claim that, because some one physicochemical theory (or some class of such theories) is not competent to explain certain vital phenomena, it is *in principle* impossible to construct and establish a mechanistic theory that can do so.

On the other hand, it would be foolish to underestimate the enormity of the task facing the mechanistic program in biology because of the intricate hierarchical organization of living things. Nor should we dismiss as pointless the protests of organismic biologists against versions of the mechanistic thesis that appear to ignore the fact of such organization.

[17] E. S. Russell, *The Interpretation of Development and Heredity*, p. 186.

As biologists of all schools have often observed, there is no such thing as a homogeneous and structurally undifferentiated "living substance," analogous to "copper substance." There have nevertheless been mechanists who in their statements on biological method, if not in their actual practice as biological investigators, have in effect asserted the contrary. It is therefore worth stressing that the subject matters of their inquiry have compelled biologists to recognize not just a single type of hierarchical organization in living things but several types, and that a central problem in the analysis of organic developmental processes is the discovery of the precise interrelations between such hierarchies.

The hierarchy most frequently cited is generated by the relation of spatial inclusion, as in the case of cell parts, cells, organs, and organisms. However, on any reasonable criterion for distinguishing between various "levels" of such a hierarchy, it turns out that there are bodily parts in most organisms (such as the blood plasma) which cannot be fitted into it. Furthermore, there are types of hierarchy that are not primarily spatial. Thus, there is a "division hierarchy," with cells as elements, which is generated by the division of a zygote and of its cell descendants. Biologists also recognize a "hierarchy of processes": the hierarchy of physicochemical processes in a muscle, the contraction of the muscle, the reaction of a system of muscles, the reaction of the animal organism as a whole; and other types which could be added to this brief list. In any event, it should be noted that in embryological development the spatial hierarchy changes, since in this process new spatial parts are elaborated. This fact can be expressed by saying that, when the division hierarchy of an embryo is compared at different times, its spatial hierarchy at a later time contains elements that did not exist at earlier times. Accordingly, organismic biologists are obviously correct in claiming that to a large extent biological research is concerned with establishing relations of interdependence between various hierarchical structures in living bodies.[18]

Let us now, however, state briefly the schematic form of a hierarchical organization (not necessarily a spatial hierarchy), with a view to assessing in general terms one component in the organismic critique of biological mechanism. Suppose S is some biological system which is analyzable into three major constituents A, B, and C, so that S can be conceived as the relational complex $R(A,B,C)$, where R is some relation. Assume further that each major constituent is in turn analyzable into subordinate constituents $(a_1, a_2, \ldots, a_i)$, $(b_1, b_2, \ldots, b_j)$, and $(c_1, c_2, \ldots, c_k)$, respectively, so that the major constituents of S can be

[18] Cf. the writings of Woodger cited above, as well as his *Axiomatic Method in Biology*, Cambridge, Eng., 1937; also L. von Bertalanffy, *Problems of Life*, Chap. 2.

represented as the relational complexes $R_A(a_1, \ldots, a_i)$, $R_B(b_1, \ldots, b_j)$, and $R_C(c_1, \ldots, c_k)$. The *a*'s, *b*'s, and *c*'s may be analyzable still further, but for the sake of simplicity we shall assume only two levels for the hierarchical organization of *S*. We also stipulate that some of the *a*'s (and similarly some of the *b*'s and *c*'s) stand to each other in various special relations, subject to the condition that all of them are related by R_A to constitute *A* (with analogous conditions for the *b*'s and *c*'s). Moreover, we assume that some of the *a*'s may stand in certain other special relations to some of the *b*'s and *c*'s, subject to the condition that the complexes *A*, *B*, and *C* are related by *R* to constitute *S*. If *S* is such a hierarchy, one aim of research on *S* will be to discover its various constituents, and to ascertain the regularities in the relations connecting them with *S* and with constituents on the same or on different levels.

The pursuit of this aim will in general require the resolution of many serious difficulties. To discover just what the presence of *A*, for example, contributes to the traits manifested by *S* taken as a whole, it may be necessary to establish what *S* would be like in the absence of *A*, as well as how *A* behaves when it is not a constitutive part of *S*. There may be grave experimental problems in attempting to isolate and identify such causal influences. But quite apart from these, the fundamental question must at some point be faced whether the study of *A*, when it is placed in an environment differing in various ways from the environment provided by *S* itself, can yield pertinent information about the behavior of *A* when it occurs as an actual constituent of *S*. Suppose, however, that we possess a theory *T* about the components *a* of *A*, such that if the *a*'s are assumed to be in the relation R_A when they occur in an environment *E*, it is possible to show with the help of *T* just what traits characterize *A* in that environment. On this supposition it may not be necessary to experiment upon *A* in isolation from *S*. The above crucial question will nevertheless continue to be unresolved unless the theory *T* permits conclusions to be drawn not only when the *a*'s are in the relation R_A in some artificial environment *E*, but also when they are in that relation in the particular environment that contains the *b*'s and *c*'s all jointly organized by the relations R_B, R_C, and *R*. Without such a theory, it will generally be the case that the only way of ascertaining just what role *A* plays in *S* is to study *A* as an actual component in the relational complex $R(A,B,C)$.

Accordingly, organismic biologists are right in insisting on the general principle that "an entity having the hierarchical type of organization such as we find in the organism requires investigation at all levels, and investigation of one level cannot replace the necessity for investigation of levels higher up in the hierarchy." [19] On the other hand, this prin-

[19] J. H. Woodger, *Biological Principles*, p. 316.

ciple does not entail the impossibility of mechanistic explanations for vital phenomena, though organismic biologists sometimes appear to believe that it does. In particular, if the *a*'s, *b*'s, and *c*'s in the above schematism are the submicroscopic entities of physics and chemistry, S is a biological organism, and *T* is a physicochemical theory, it is not impossible that the conditions for the occurrence of the relational complexes *A*, *B*, *C*, and S can be specified in terms of the fundamental concepts of *T*, and that furthermore the laws concerning the behaviors of *A*, *B*, *C*, and S can be deduced from *T*. But, as has been argued in the preceding chapter, whether in point of fact one science (such as biology) is reducible to some primary science (such as physicochemistry), is contingent on the character of the particular theory employed in the primary discipline at the time the question is put.

4. We must finally turn to what appears to be the main reason for the negative attitude of organismic biologists toward mechanistic explanations of vital phenomena, namely, the alleged "organic unity" of living things and the consequent impossibility of analyzing biological wholes as "sums" of independent parts. Whether there is merit in this reason obviously depends on what senses are attached to the crucial expressions 'organic unity' and 'sum.' Organismic biologists have done little to clarify the meanings of these terms, but at least a partial clarification has been attempted in the preceding and present chapters of this book. In the light of these earlier discussions the issue now under examination can be disposed of with relative brevity.

Let us assume, as do organismic biologists, that a living thing possesses an "organic unity," in the sense that it is a teleological system exhibiting a hierarchical organization of parts and processes, so that the various parts stand to each other in complex relations of causal interdependence. Suppose also that the particles and processes of physics and chemistry constitute the elements at the lowest level of this hierarchical system, and that *T* is the current body of physicochemical theory. Finally, let us associate with the word 'sum' in the statement 'A living organism is not the sum of its physicochemical parts,' the "reducibility" sense of the word distinguished in the preceding chapter. The statement will then be understood to assert that, even when suitable physicochemical initial and boundary conditions are supplied, it is not possible to deduce from *T* the class of laws and other statements about living things commonly regarded as belonging to the province of biology.

Subject to an important reservation, the statement construed in this manner may very well be true, and probably represents the opinion of most students of vital phenomena, whether or not they are organismic biologists. The statement is widely held, despite the fact that in many cases physicochemical conditions for biological processes have been ascertained. Thus, an unfertilized egg of the sea urchin does not normally develop into an embryo. However, experiments have shown that, if such eggs are first placed for about two minutes in sea water to which a certain quantity of acetic acid has been added and are then transferred to normal sea water, the eggs presently begin to segment and to develop into larvae. But, although this fact certainly counts as impressive evidence for the physicochemical character of biological processes, the fact has thus far not been fully explained, in the strict sense of 'explain,' in physicochemical terms. For no one has yet shown that the statement that sea urchin eggs are capable of artificial parthenogenesis under the indicated conditions is *deducible* from the purely physicochemical assumptions T. Accordingly, if organismic biologists are making only the *de facto* claim that no systems possessing the organic unity of living things have thus far been proved to be sums (in the reducibility sense) of their physicochemical constituents, the claim is undoubtedly well founded.

On the other hand, in the prevailing circumstances of our knowledge there should be no cause for surprise that the fact about the artificial parthenogenesis of sea urchin eggs is not deducible from T. The deduction is not possible, if only because the elementary logical requirements for performing it are currently not satisfied. No theory can explain the operations of any concretely given system unless a complete set of initial and boundary conditions for the application of the theory is stated in a manner consonant with the specific notions employed in the theory. For example, it is not possible to deduce the distribution of electric charges on a given insulated conductor merely from the fundamental equations of electrostatic theory. Additional instantial information must be supplied in a form prescribed by the character of the theory—in this instance, information about the shape and size of the conductor, the magnitudes and distribution of electric charges in the neighborhood of the conductor, and the value of the dielectric constant of the medium in which the conductor is embedded. In the case of the sea urchin eggs, however, although the physicochemical composition of the environment in which the unfertilized eggs develop into embryos is presumably known, the physicochemical composition of the eggs themselves is still unknown, and cannot be formulated for inclusion in the indispensable instantial conditions for the application of T. More generally, we do not know at present the detailed physicochemical composition of any living organism, nor the forces that may be acting between the elements on the lowest level of its hierarchical organization. We are therefore currently unable to state in exclusively physicochemical terms the initial and boundary conditions requisite for the application of T to vital systems. Until we can do this, we are in principle precluded from deducing biological laws from mechanistic theory. Accordingly, although it may indeed be true that a living organism is not the sum of its physicochemical parts, the available evidence does not warrant the assertion either of the truth or of the falsity of this dictum.

Although the point just stressed is elementary, organismic biologists often appear to neglect it. They sometimes argue that, while mechanistic explanations may be possible for traits of organic parts when these parts are studied in "abstraction" (or isolation) from the organism as a whole, such explanations are not possible when the parts function conjointly in mutual dependence as actual constituents of a living thing. But this claim ignores the crucial fact that the initial conditions required for a mechanistic explanation of the traits of organic parts manifested when the parts exist *in vitro* are generally insufficient for accounting mechanistically for the conjoint functioning of the parts in a biological organism. For it is evident that when a part is isolated from the rest of the organism it is placed in an environment which is usually different from its normal environment, where it stands in relations of mutual dependence with other parts of the organism. It therefore follows that the initial conditions for using a given theory to explain the behavior of a part in isolation will also be different from the initial conditions for using that theory to explain behavior in the normal environment. Accordingly, although it may indeed be beyond our actual competence at present or in the foreseeable future to specify the instantial conditions requisite for a mechanistic explanation of the functioning of organic parts *in situ*, there is nothing in the logic of the situation that limits such explanations in principle to the behavior of organic parts *in vitro*.

One final comment must be added. It is important to distinguish the question of whether mechanistic explanations of vital phenomena are possible from the quite different though related problem of whether living organisms can be effectively synthesized in a laboratory out of nonliving materials. Many biologists seem to deny the first possibility because of their skepticism concerning the second. In point of fact, however, the two issues are logically independent. In particular, although it may never become possible to manufacture living organisms artificially, it does not follow that vital phenomena are therefore incapable of being explained mechanistically. A glance at the achievements of the physical sciences will suffice to establish this claim. We do not possess the power to manufacture nebulae or solar systems, despite the fact that we do possess physicochemical theories in terms of which nebulae and planetary systems are tolerably well understood. Moreover, while modern physics and chemistry provide competent explanations for various properties of chemical elements in terms of the electronic structure of the atoms, there are no compelling reasons for believing that, for example, men will some day be capable of manufacturing hydrogen by putting together artificially the subatomic components of the substance. On the other hand, the human race possessed skills (e.g., in the construction of dwellings, in the manufacture of alloys, and in the preparation of foods) long before adequate explanations for the traits of the artificially constructed articles were available.

Nonetheless, organismic biologists often develop their critique of the mechanistic program in biology as if its realization were equivalent to the acquisition of techniques for literally taking apart living things and then overtly reconstituting the original organisms out of their dismembered and independent parts. However, the conditions for achieving mechanistic explanations for vital phenomena are quite different from the requirements for the artificial manufacture of living organisms. The former task is contingent on the construction of factually warranted theories of physicochemical substances; the latter task depends on the availability of suitable physicochemical materials, and on the invention of effective techniques for combining and controlling them. It is perhaps unlikely that living organisms will ever be synthesized in the laboratory except with the help of mechanistic theories of vital processes; in the absence of such theories, the artificial manufacture of living things, were this ever accomplished, would be the outcome of a fortunate but improbable accident. But in any event, the conditions for achieving these patently different tasks are not identical, and either may some day be realized without the other. Accordingly, a denial of the possibility of mechanistic explanations in biology on the tacit supposition that these conditions do coincide, is not a cogently reasoned thesis.

The main conclusion of this discussion is that organismic biologists have not established the absolute autonomy of biology or the inherent impossibility of physicochemical explanations of vital phenomena. Nevertheless, the stress they place on the hierarchical organization of living things and on the mutual dependence of organic parts is not a misplaced one. For, although organismic biology has not convincingly secured all its claims, it has demonstrated the important point that the pursuit of mechanistic explanations for vital processes is not a *sine qua non* for valuable and fruitful study of such processes. There is no more reason for rejecting a biological theory (e.g., the gene theory of heredity) because it is not a mechanistic one (in the sense of "mechanistic" we have been employing) than there is for discarding some physical theory (e.g., modern quantum theory) on the ground that it is not reducible to a theory in another branch of physical science (e.g., to classical mechanics). A wise strategy of research may indeed require that a given discipline be cultivated as a relatively independent branch of science, at least during a certain period of its development, rather than as an appendage to some other discipline, even if the theories of the latter are more inclusive and better established than are the explanatory principles of the former. The protest of organismic biology against the dogmatism often associated with the mechanistic standpoint in biology is salutary.

There is, however, an obverse side to the organismic critique of that dogmatism. Organismic biologists sometimes write as if any analysis of vital processes into the operation of distinguishable parts of living things entails a seriously distorted view of these processes. For example, E. S. Russell has maintained that in analyzing the activities of an organism into elementary processes "something is lost, for the action of the whole has a certain unifiedness and completeness which is left out of account in the process of analysis." [20] Analogously, J. S. Haldane claimed that we cannot apply mathematical reasoning to vital processes, since a mathematical treatment assumes a separability of events in space "which does not exist for life as such. We are dealing with an indivisible whole when we are dealing with life." [21] And H. Wildon Carr, a professional philosopher who subscribed to the organismic standpoint and wrote as one of its exponents, declared that "Life is individual; it exists only in living beings, and each living being is indivisible, a whole not constituted of parts." [22]

Such pronouncements exhibit an intellectual temper that is as much an obstacle to the advancement of biological inquiry as is the dogmatism of intransigent mechanists. In biology as in other branches of science knowledge is acquired only by analysis or the use of the so-called "abstractive method"—by concentrating on a limited set of properties things possess and ignoring (at least for a time) others, and by investigating the traits selected for study under controlled conditions. Organismic biologists also proceed in this way, despite what they may say, for there is no effective alternative to it. For example, although J. S. Haldane formally proclaimed the "indivisible unity" of living things, his studies on respiration and the chemistry of the blood were not conducted by considering the body as an indivisible whole. His researches involved the examination of relations between the behavior of one part of the

[20] E. S. Russell, *The Interpretation of Development and Heredity,* p. 171.
[21] J. S. Haldane, *The Philosophical Basis of Biology,* London, 1931, p. 14.
[22] Quoted in L. Hogben, *The Nature of Living Matter,* London, 1930, p. 226.

body (e.g., the quantity of carbon dioxide taken in by the lungs) and the behavior of another part (the chemical action of the red blood cells). Like everyone else who contributes to the advance of knowledge, organismic biologists must be abstractive and analytical in their research procedures. They must study the operations of various prescinded parts of living organisms under selected and often artificially instituted conditions—on pain of mistaking unenlightening statements liberally studded with locutions like 'wholeness,' 'unifiedness,' and 'indivisible unity' for expressions of genuine knowledge.