revista mexicana de física 27 no. 4 (1981) 583-611

THE AXIOMATIC APPROACH IN PHYSICS

T.A. Brody *

Instituto de Física, UNAM

Apdo.Postal 20-364, México 20, D. F.

(recibido agosto 13,1980; aceptado febrero 20,1981)

ABSTRACT

In this paper it is shown, firstly, that physical theories contain not completely formalizable elements, the most significant being meaning assignments to formal components, approximations, unspecified elements fixed by suitable determination in specific models, and the scope (or region of validity) of the theories; secondly, that the importance of these elements depends on the type of theory as specified by a number of characteristics; and thirdly, that theories can —and must— contain various kinds of inconsistencies that are amenable to rational manipulation but preclude axiomatization within the framework of formal logic. Finally a number of aims for the axiomatic approach are outlined which are compatible with these non-formalizable elements and moreover do not exist for axiomatic methods in mathematics.

RESUMEN

Se muestra, primero, que las teorías de la física contienen elementos no completamente formalizables, de los cuales los más significativos son especificaciones de significado para elementos formales, aproximaciones, elementos por determinarse mediante modelos particulares, y el alcance de las teorías; en segundo lugar, que la importancia relativa de estos elementos determina si, y en este caso hasta donde, los métodos axiomáticos pueden aplicarse útilmente a teorías físicas; y en tercer lugar, que las teorías contienen — necesariamente — varios tipos de contradicciones e inconsistencias que se dejan manipular racionalmente pero no con las operaciones de la lógica formal. Finalmente se indican algunas metas para la axiomática en la física que son compatibles con los elementos no formalizables mencionados y que no existen en los métodos axiomáticos de las matemáticas.

The axiomatization of theoretical physics was included by Hilbert[1] in his famous list of twenty-three outstanding mathematical problems. Though since then various attempts at axiomatizing physical theories have been made, notably for analytical mechanics and for quantum mechanics, only one has had much significance for the development of physics: Carathéodory's[2] work on (classical) thermodynamics. At the present time one can find both passionate defenders[a] of the axiomatic method in physics and determined attackers[b]. Yet the mainstream of physics has ignored the problem.

In this no doubt the innate conservatism of physicists has been a factor; so also has been the abuse of axiomatic techniques in some neighbouring fields such as systems theory, where they have served to throw a glamorous mantle over the otherwise too evident theoretical poverty. But there are also more solid reasons, rooted in the essential differences between the two fields, why axiomatics has contributed so much less to physics than to mathematics, an it is the aim of this paper to have a look at these reasons and their more relevant implications.

After outlining, in section I, the nature of the axiomatic approach in mathematics and why it is of significant importance there, we examine in section II how far the different non-mathematical elements in a physical theory are formalizable; section III looks at various relevant characteristics of physical theory and draws conclusions concerning which types of theory lend themselves to axiomatization; section IV studies the specific problem of the inconsistencies in physical theory and their role; finally, in section V, some new aims for the axiomatic approach are outlined which are relevant to physics but not to mathematics.

I

The axiomatic method has, ideally, two phases.

In the first phase, the aim is to exhibit the structure of a theoretical edifice in systematic form as what we shall call an "axiom-based deductive structure", or ADS. From its axiom base, consisting of primitive concepts, axioms and rules of derivation, the theorems comprising the body of the ADS are derived by successive formal deduction, without making use of elements outside the ADS[c].

Of all the many questions associated with the axiomatic method in mathematics, only one point need be briefly examined here, namely the formal rigour it is intended to help establish. Now its desirability is not disputed even by the intuitionist school for whom the logic employed in a mathematical method is something to be discovered a posteriori. What is perhaps less generally appreciated and yet would seem to be inescapable is the quite practical reason that underlies this drive to maximal rigour. For a mathematical theory (or, for that matter, a logical system) is constructed in independence of any particular application; but if it is to work correctly in any situation where its central concepts and their interrelations adequately represent the essence of that situation, then it must have a level of internal consistency and closeness of argument which is adequate to cover the needs of any foreseeable application. The growing variety and sophistication of the applications of mathematical techniques in physics, engineering, and a rapidly increasing multitude of other fields has meant that the users of mathematical theories have become steadily more exigent over the last few centuries: it is without a doubt this need that has stimulated the continuing effort to achieve greater rigour and firmer foundations for the mathematical edifice. But what is of central importance for our present purposes is that no one of the possible applications of a given mathematical structure or theory may be permitted to fix the meaning (i.e., to establish a definite correlation between objects external to the structure and objects belonging to it) of its elements; or rather, none of the possible meanings can be used in establishing the validity of the structure. Thus the elements in the structure are reduced to the status of referenceless symbols and their interrelations must be stipulated with seeming arbitrariness: in other words, in pure mathematics we can have only formal structures, and the validity of deductions can only be based on those properties of the formal elements that have been explicitly stated. Thus the axiomatic method becomes indispensable[d].

Though in the course of the last hundred years various conflicting views on what, precisely, should constitute the basis of an axiomatic reformulation of mathematics have been propounded, this much at least is common ground. And even the intuitionists have implicitly given recognition to this state of affairs in that they have themselves attempted to formalize their methods (e.g. Heyting[3]).

The second phase of the axiomatic method consists in the exploitation of the ADS constructed during the first phase. In mathematics, the reformulation of a theory as an ADS offers a double advantage: on the one hand it allows the verification that is needed of the internal logical consistency of the theory —or at least, as far as the implications of Gödel's theorem and present-day limitations of technique allow; and on the other hand, the separation of presuppositions or postulates and deductions or theorems greatly eases the task of checking to what extent a proposed application is effectively possible. This second point is of importance because a large proportion of the applications of many mathematical theories is to other mathematical theories. Here exact equivalence of the axiom basis is both possible and desirable, and where it is achieved we shall speak of a realization of the theory[e].

Where a mathematical structure is applied in physics or another natural science, however, there can be no question of an exact adaptation; one could at best speak of approximate realizations. For here the relationships among the relatively few concepts involved in the mathematical construction cannot do more than reproduce the central features that characterize the multifarious richness of any particular segment of nature. It is only through a more or less extensive effort of abstraction that we can arrive at a morphologically simplified version which the mathematical description will fit; and in this process of abstraction certain features are lost, others exaggerated, still others distorted, so that the fit is very far from exact. If the theoretician has done his work well, the fit will be adequate for the purpose in hand —no more.

And this purpose is, of course, quite different. The realization of one mathematical structure in another has as its aim the reduction of the consistency problem for the one to that of the other; the aim of a physical theory, however, is an understanding of the underlying part of nature, or at least sufficient understanding to enable us to use and control it. It is this difference of purpose —and the resulting differences in procedures— that makes it desirable to consider the questions of axiomatization in physics from the standpoint of that science rather than from the mathematical one.

II

In what follows, we shall consider the problem in physics —with a few glances at related sciences— essentially because physics of all natural sciences has the most developed formalizable core[(6)]. We shall, moreover, restrict our considerations to theoretical physics, which is the part containing whatever formalizable elements there are; nevertheless it must not be forgotten that physics could not exist without its experimental half, whose influence on theoretical development is decisive, however complex and ill-understood it may be.

A first and all too often forgotten point is that physics is a "bootstrap" activity —one, that is to say, in which the revision of any part depends on the others, which themselves are remodeled on a basis that includes the part being revised. It is thus the complete negation of the Cartesian method of universal doubt, for it concentrates its effort on a single point at a time. As a consequence, every part in the structure is dependent on every other part, though the precise form of the dependence (which is by no means purely logical) changes dynamically. An ADS thus misrepresents the situation by singling out certain concepts and axioms as basic. Nevertheless, if we concentrate our attention on a particular theory and neglect its connection with the rest of physics, its structure at a given moment of time can be to some extent be represented by an ADS; and it is this relative and limited validity that we shall examine in what follows.

Since in physics the mathematical structure of a theory is always applied to one specific kind of object and this application is fundamental in the construction of the theory, it is clear that the theoretical edifice must contain a lot more than just the elements formalized in an ADS of the mathematical type. The kind of object to which the theory applies will here be called a physical system (or simply system) and will be taken to be a segment of the universe and therefore to have real existence. Clearly a theory applicable to only one particular system can have genuine interest only if that system is in itself sufficiently complex and important; this will be so if the system is unique -as is the case for cosmology- or if we have no direct access to others of the same kind -the situation in geophysics[g]. But equally clearly this is a fairly exceptional state of affairs; usually the number of possible systems is large and hence their variety is considerable also. The physical theory is expected to be applicable to all of them, and this condition, too, has important consequences that we examine below.

Besides the elements which correspond in kind to those in a mathematical axiom base, a physical theory will then contain a number of other components, among which the following are the most significant ones:

i) Assignments of physical meanings; that is to say, the establishment of correspondences between quantities (or sometimes non-quantative elements) in the theory and properties of or relation among the entities composing the type of physical system described by the theory.

ii) Assumptions as to what entities (elements or properties) in the system are to be neglected or treated only approximately in the theory.

iii) The scope or region of validity of the theory; that is to say, the region of phenomena over which the theory provides satisfactory explanation and sufficiently accurate prediction.

iv) Certain open positions in the theoretical structure where specific information (which can be both quantitative and qualitative) can be inserted. The presence of these "holes" guarantees the required generality of the theory. Filling in all the missing information completes the building of a specific model, intended to describe one physical system (or a restricted range of systems): the information thus characterizes the system, and includes data concerning its constitution as well as the initial conditions.

That these additional components are not sufficiently formalizable and yet indispensable to the formulation of a theory may require some discussion. Thus it has been considered (e.g. Bunge[4]) that the semantics required to establish the connection between mathematical structure and physical reality should form part of the formal axiom base. Now it is perfectly true that we can include among the axioms whatever stipulations are needed to assign clear meanings to the basic entities (i.e., the undefined concepts and quantities appearing in the axioms) of the theory. But if the semantic axioms are to play their proper role in the theory, we also need methods for deducing the meanings of derived quantities -methods, moreover, which can be stated as rules that reduce these deductions to formal schemes; for stipulating axioms is of no use unless such formal rules integrate the content of the theory into an ADS. Now there are certain aspects of semantic constructions which can be formalized in such a way; an important instance is the calculus of dimensions; and one can be misled into thinking that this holds for all the varied components of meaning. But this is not so, as we can see by means of some examples. In equilibrium thermodynamics the distinction between extensive and intensive properties[h] is extremely useful, and may moreover be formalized without difficulty, so that one can determine from the rules whether a derived quantity is extensive or intensive. But the dichotomy is valid only so long as surface effects may be neglected; when very small systems are considered - and also in many other branches of physics - there arise quantities which are intermediate or depend in more complex ways on the size of the system; and while the exact mathematical dependence may be obtained once a quantity is defined, it is by no means clear what this contributes to the <u>meaning</u> of the quantity. Extensivity restricts strongly the possible combinations of the quantities that have the property - as long as it must characterize all quantities appearing in the theory - and thus gives one part of the meaning; but in a wider theory these restrictions disappear. The calculus of dimensions -which, as we mentioned, is formalizable-contributes a far from negligible ingredient to the meaning of all those concepts whose quantitative expression has dimensionality. But physics employs a good many purely qualitative ideas; there are also many dimensionless constants and quantities of differing and unrelated significance; and that the dimensions are inadequate for specifying the nature and meaning of a physical concept is made evident by such cases as angular momentum and action —which have the same dimensions.

It will be clear from these considerations that we have tacitly extended the implications of "meaning" from the simple stipulation of a correspondence between a theoretical quantity and an experimentally characterizable component of systems, i.e. an extensional definition, to embrace the intensional aspects; but few will deny that these belong to "what we mean by meaning". The intensional aspects are, however, essential, all the more so because without them an unresolvable paradox arises: we cannot assign an extensional meaning to the basic entities of the ADS without stepping outside the bounds of the ADS itself, in a manner incompatible with the self-contained nature we require of it. This paradox reflects, of course, the misrepresentation that we mentioned above of a dynamic physics by static ADS's.

In constructing a theory, only a finite and relatively small number of elements and characteristics of the physical systems to which the theory is to apply is taken into account; all others are either completely ignored or else incorporated in the theory in an approximate, condensed and hence distorted form. Explicit statements to this effect will essentially be of two kinds: firstly, those that stipulate the complete elimination from the theory of certain elements, and these need not figure in the axiom base for obvious reasons; and secondly, the statements that specify how to approximate or resume other elements which are not to be treated in full explicitness. Such statements are, of course, required in the theory; but since they are made use of in ways that are not very easy to formulate in a general and abstract way and are rather meant to guide the physicist's judgment, their inclusion in the axiom base does not seem justified.

Of course both kinds of statements will become necessary when we attempt to set out how the theory relates to other theories; but whether or not we consider such relations to belong to the theoretical structure, there is no place for them in the axiom base.

The scope of a theory likewise cannot be dealt with by axiomatic methods, particularly so because it is (except for certain special cases) determined experimentally, and therefore unknown at the time the theory is first formulated. It might therefore be argued, with some justice, that the scope is not properly part of the theory, and this is certainly true so long as we restrict ourselves to the purely formal aspects of the theory. But - in this unlike a mathematical theory- a physical theory is a theory _of something_; and to know precisely what it is the theory of we need to know how far its validity extends, we need to know its scope. More practically, without information about its scope a theory is useless; we do not know when and where to apply it; thus a well established theory is always described together with its scope. There is, further, a much more subtle point: as we shall see below, the ADS of a physical theory is not normally taken to include all logically possible deductions from the axiom basis, but only those that fall within its scope; for the others are in some sense aberrant, are irrelevant to any use we may make of the theory for understanding the world we live in and controlling it, and should therefore be excluded (though they would not be in a mathematical ADS). In other terms, the scope belongs to a theory much as a man's skin belongs to his body.

Now, knowledge of the scope is evidently required for establishing the (extensional) meaning of the concepts a theory organizes into a whole. If then we propose to include meaning in the theory's ADS, the same should be done with the scope; however, in view of its empirical nature, this seems peculiarly absurd: a scope can hardly be seen as axiomatizable. Nor do we have, so far, any general rules for combining scopes in order to derive the scope of a composite theory; indeed, it may be doubted whether such rules are even possible, though an upper limit for a composite theory's scope can of course be found. We conclude that scope cannot form part of an ADS, and with it the meanings it helps to delimit must be excluded; they are integral elements of a physical theory, but not of its ADS.

Finally, physical theories are, in a sense, incomplete; in order to derive definite models from them, we must complement them with the information that specifies the details of the particular physical system to which they are to be applied. These details must cover everything from the qualitative and quantitative description of the system itself to the initial values of the various quantities involved - energies, positions and velocities, an so on. The "holes" left for this purpose within the theoretical structure are, once again, not susceptible to axiomatic treatment; though we can axiomatize about numerical values not yet specified, we cannot do so for unspecified structural information; nor do we know how to do so with purely qualitative information that we do not have, and even the very way in which the theory uses some of the numbers may depend on their values. It is thus evident that the part played by the "holes" for system specification and initial conditions is very much dependent on the uses of the theory and on the approximations made in it and therefore cannot sensibly be incorporated in an ADS.

III

As well as containing non-formalizable elements in an essential way, the theories of physics differ among each other in ways which are relevant to our problem. There are several unsolved questions connected with this point which are worth stating, if only to stimulate attempts at answering them.

Theories differ in their degree of generality, in their phenomenological character or profundity, in the degree of quantitativeness, in their independence with regard to more fundamental theories, and in their structural exactness.

A theory is more general (or more fundamental) if it covers a wider range of phenomena than another theory - not merely in the sense of having a wider scope, but rather in the sense of applying to a larger variety of <u>types</u> of phenomena. The less it depends on the specification of the sort of physical system it describes, the more fundamental it will be; its generality is given by the range of differing models that one can build with it. Thus quantum mechanics is more general than, for instance, the theory of the solid state, and the latter is more general than the theory of metallic conduction. It should be noted that there is in no sense more merit or even greater profundity attached to a more general theory: the Bardeen-Cooper-Schrieffer theory of superconductivity is a highly specific one, covering as it does only the behaviour of electrical conduction electrons in certain metals at very low temperatures, - yet it is profound indeed in that it opens up a first chapter in what may well be central to the physics of the 21st century, the behaviour of systems of "not very many particles". In fact physics needs both general and specific theories, because only the combination of both types will yield useful predictions (whether usefulness is here of an applied kind or to another research problem is quite irrelevant); thus value judgment are very much beside the point.

One misconception that may be mentioned here because it is frequent even among physicist is that the various fundamental theories of physics must be independent of each other in the sense of being closed off and self-sufficient. In fact, of couse, the unity of physics is not merely one of subject matter and method but also one of mutual connections among all its theories. Thus the most general theory of all, classical mechanics, is also in a sense the most dependent on the others, for the forces that enter into Newton's equations (or the potentials in a Lagrangian formulation) are not explained or even described by it but originate in phenomena discussed by other theories. And one of the most elegantly rounded theories, Maxwell's electromagnetic theory, cannot account by itself for the stability of extended elementary charges; this has in the past led many theoreticians into accepting the view that the elementary charges must be point-like, though it is quite well known that this leads to contradictions and paradoxes.

A profound theory may be contrasted with a phenomenological one. This distinction is significant in physical research, but not easily circumscribed. The extreme of phenomenology is the abridged description of experimental results: a theoretical formulation which does not go farther has no predictive power beyond what is already known from the laboratory. Such an extreme is almost unknown to natural science, though it has its place in social science where no clear account relating the various factors may be available. Historiography, for instance, is of this sort. One step further up the theoretical ladder leads us to theories which generalize beyond experimental results and so predict new observations, but only for very similar situations. Laws such as Boyle's law connecting the volume and pressure of a gas at constant temperature were - at the time of their discovery- of this sort, though now they appear as consequences of a more elaborate theory. Situations which do not go much further than this exist, of course, in physics; in elementary-particle physics, a large number of such experimental generalizations are known— and the fact that we can often achieve an elegant and economical description of such laws by group-theoretical structures must not blind us to the fact that classifying the known particles according to SU(3), for instance, yields no understanding of why this works; and the predicted discovery of the omega particle is evidence that it works remarkably well.

This why is an essential question, for the ability to answer it, what is commonly called its explanatory power, is precisely what makes a theory profound. The distinction between profound and phenomenological theories is not only common knowledge among physicists, it is indeed a useful and widely employed concept when formulating research aims. But among philosophers of science the concept is often held to be meaningless. The problem appears to lie in a certain confusion about the answer expected to that question, why. Let us consider an example. A violin string, suitably excited, will vibrate at certain selected frequencies, while motion at other frequencies is rapidly damped out. One kind of explanation that may be offered runs like this: the second-order differential equation describing the string's motion has stationary solutions only when suitable boundary conditions are satisfied, and this happens only for certain specific frequencies. True, but not illuminating. Another sort of explanation considers the possible ways a string fixed at either end can oscillate; each kind of oscillation has its wavelength, determined by the frequency and the mechanical properties of the string, and clearly there must be an integral number of half-wavelengths in the length of the string, or else the string will either snap or transfer all the motion's energy to the end-blocks. Hence only some frequencies correspond to vibrations that can last. Again true, but now we gain some insight into the mechanism that stabilizes certain frequencies and not others; we see what forces are at work and we can extend the model to account for frictional effects and so on.

Thus an explanation may be based on the mathematical structure of the theory alone, or it may derive from the physical significance of the concepts involved. In the latter case it will not only furnish a causal structure and hence a dynamical account for the phenomena covered by the theory, it will also provide a framework that links the meanings of the concepts in the theory, exhibiting them as far from arbitrary: in a physical theory of any profundity the meanings of its concepts, which we saw above to be an essential ingredient, cannot be assigned at will but only in such a way as to yield confirmable causal nexuses; in a phenomenological theory, this is not the case. Such distinctions depend, of course, on the explicit premiss of a world that is both real and independent of what we may happen to think; certainly it is among those who accept this premiss that we find the recognition that explanatory power is relevant - e.g., in Bhaskar [5]. Only on this premiss can we accept that the theoretical physicist's constructions are able effectively to mirror the behaviour and relations of the entities he studies: that we are, in other words, able to build functioning, dynamical models of selected aspects of our world. To deny such distinctions, as logical positivism obliges us to do through all too well known arguments, not only deprives the physicist of a useful tool but can create serious confusions. Thus the common view (repeatedly stated for instance by Bohr, as Scheibe [6] brings out clearly) that classical physics is based on the point particle, with zero extension, and the field, with infinite extension, usually leads to the conclusion that such easily visualizable (!) models are inappropriate to quantum mechanics, and that therefore only explanations of the mathematical type should be sought; and this is meaningful only if the profound/phenomenological distinction is abandoned.

The necessity for a theory in physics to be basically quantitative is by now well established, and none of the major theories are chiefly qualitative. Yet no theory is <u>purely</u> quantitative, and its qualitative features are essential to un understanding of its meaning and also to its applications. These qualitative aspects of a theory tend often to be forgotten; yet commonly they determine the field of usefulness or scope of the theory - as for instance in the case of the distinction between continuum and corpuscular theories.

The independence of a theory is connected with how fundamental it is, but is by no means identical with this property. A fundamental theory of wide scope may form the basis for many more specific theories, but need not therefore be independent. Thus a great deal of the present-day theory of the solid state, itself the generator of many detail theories, is directly dependent on quantum mechanics on the one hand, and on statistical thermodynamics on the other; the latter depends in its turn on quantum or classical mechanics and on statistical theory. At another level, the theory of general relativity is perhaps less fundamental than quantum mechanics (in the sense that its scope is more restricted and that it has generated far fewer dependent theories; we repeat that this does not mean that it is less profound); but it is more independent in that it creates the basis for its own formulation of mechanics, while quantum theory requires such a basis from outside — either Newtonian mechanics or special relativity.

Lastly, a theory may be said to be structurally exact if nowhere in its deductions the need for approximations arises. We must distinguish here between the sort of approximations that are used in deriving specific models because we do not have suitable mathematical tools, and the approximations made in order to be able to neglect what we judge to be inessential factors. Only the latter are relevant here, since they enter in an irremovable way into the framework of the theory and therefore must be considered when we attempt to create an ADS for it.

Structurally inexact theories are often theories in process of development: their central features may already be clear, but many details are lacking and with them a fully developed mathematical apparatus. In other situations the limitation is essentially experimental, as when we have quantities whose values are important but which we do not know how to measure.

Of course these various distinctions among theories are not independent of each other. Thus a general theory is mostly also profound, quantitative, and structurally exact; but as we have noted, there are exceptions, and the contrary is not usually valid.

The importance in research of the formal or formalizable part of a theory depends very much on where the theory falls along the scales of these different characteristics. To the extent that a theory is general and structurally exact, its formal part is central, and its profundity will then lend importance to the attempts at creating an ADS for it. But it is by no means clear (at the present moment, at least) how far a very specific and rather phenomenological theory can be axiomatized in a satisfactory way - i.e. without trivializing it; and in fact there is reason to doubt whether the exercise would be at all useful. A similar caution is needed with still undeveloped theories, because they tend to be structurally inexact in crucial places. For when a theory contains approximations in an essential way, all attempts at formal description will distort it beyond recognition; in fact, using an approximation is equivalent to an open invitation to use one's intuitive judgment about the validity of the procedure.

It is this situation which both justifies many of the fears that have been expressed by opponents of the axiomatic method in physics because its proponents have seen it, quite absurdly, as universally applicable - and creates the basis for selectively axiomatizing those theories for which it is meaningful. And it cannot be denied that such theories as classical mechanics or thermodynamics offer a very suitable field. But in quantum mechanics the attempts at axiomatization, though in many respects extraordinarily useful, have intensified the problems of interpretation rather than helped to resolve them.

We must conclude that the axiomatic approach is by no means universally desirable in physics; while it offers definite advantages (which we discuss below), there are clearly also some dangers that threaten.

IV

But before entering into these advantages of the axiomatic approach, we must mention an important matter which has by no means received the consideration it deserves. This is the presence of inconsistent and sometimes openly contradictory elements within the framework of physics. Not all the varieties of inconsistency in physics are directly relevant to our problem of the axiomatic method; but since they are all fairly intimately linked, it seems worthwhile to list the important ones.

A first kind of inconsistency arises because of the need to connect theory and experiment: for the construction and operation of experimental set-ups, and the interpretation of the results obtained from them, require a set of concepts that usually go well beyond what the theory under test can offer. As a result, we employ an astonishing mixture of theoretical notions with very different and often incompatible basic assumptions in order to do experimental work and link it to theory. Thus the experimental verification of relativity theory makes use of instrumentation designed on the basis of classical mechanics, of optics, of quantum theory, of electromagnetic theory, and of other branches of physics as well.

The experimental physicist is quite at home in this situation; he knows that for his purposes any theory that yields a sufficient approximation is good enough, and he has developed into a fine art the technique of combining incompatible theoretical constructs. This is not the place to examine the various epistemological and other presuppositions that make this "fine art" possible; suffice it to say that one can indeed work in a consistent fashion by combining inconsistent elements, provided certain intuitively understood constraints are observed.

A second variety of inconsistency, more directly internal to theoretical structures, is closely connected to this one: it arises whenever an approximation is made within the framework of a theory, for the basis on which such approximations are accepted is either another and usually incompatible theory or a set of experimental data, likewise obtained on theoretically unrelated foundations. The various "semi-classical" calculations so beloved of the quantum chemist are of the first sort; they exemplify the combination of incompatible theoretical contributions at its best, for they are both ingenious and remarkably successful. A second case is that of the experimentally justified estimation of relative magnitudes which allows us to neglect a small but theoretically bothersome term; again this is of frequent occurrence in physical theory.

There is a sense in which both these kinds of inconsistency are irrelevant or at least of reduced importance: they do not strike at the central core of a theory's structure- or at least they do not appear to do so. Yet only the second kind can be attributed to our human limitations; the first is clearly essential in the nature of things; and the systematic way in which both crop up places some doubt on their secondary character. A different kind of inconsistency, linked in quite central ways to the theoretical structure, arises because - as we saw above- the physical meanings incorporated in a theory are not in general the realizations (in the model-theoretic sense specified above, note [e]) of corresponding formal elements. Hence the formal part of the theory can imply consequences which go beyond what the non-formal part may justify; such consequences are dubbed "unphysical" and simply thrown out. This is the case when equations, written to describe a physical phenomenon, have solutions we do not want but are unable to get rid of in a mathematically satisfactory way. For instance, Maxwell's equations for the electromagnetic field have an advanced and a retarded solution, and we ignore one of them on the basis of a heuristic causality argument, not otherwise germane to the theory. Another case of a similar nature is that of the phase of a wave function in quantum mechanics, the absolute value of which is quite without physical meaning. Such cases appear to arise only rarely, because the majority of them is avoided by an important practical limitation we apply to an ADS: we do not allow it to include explicitly <u>all</u> possible consequences of its axiom base, but only those that do not obviously fall outside the theory's scope. This limitation is of profound significance. In the physicist's practice it is what allows him to combine incompatible theoretical constructs into one argument, as we exemplify below. From the epistemological point of view, it is quite as interesting: it is because of this limitation that the empirically expected consequences of a new conception do not extended throughout the whole edifice of knowledge, so that researchers can work each on his own problem without having to consider all the possible repercussions of any particular new idea. Moreover, one could use it actually to <u>define</u> a theory's scope as the range of phenomena over which deductions from the starting postulates may meaningfully be made. From a logical standpoint the limitation is likewise of considerable relevance; apart from the fact that it points toward a clear discrimination between the logical and the rational, it opens up a promising new field of study: logical systems in finite universes, where logical operations may connect propositions from different such universes. But to the author's knowledge, this has not yet been explored.

But the mismatch between the formal elements and the physical significance may have much more serious repercussions. Again, we find a case in classical electrodynamics, which is an excellent example of a finished and elegant physical theory. If we attempt to calculate the energy a pure electric charge has because of its interaction with the electromagnetic field it creates around itself, we get into deep trouble: if we take the particle to be point-like, then this energy (usually called the self-energy) becomes infinite, with no apparent source to provide it; if one the other hand we take a particle of a small but definite size, then we need forces to avoid its being broken up by the field, and where would these forces come from? This difficulty (which we cannot remove by any of the tricks for getting rid of unwanted "unphysical" solutions) is not improved when we go over to the much more sophisticated theory of quantum electrodynamics: here these infinities turn up in just as disturbing a fashion. (We shall not enter into the thicket of renormalization theory here.)

The mismatch may also appear between different parts of the formal structure. Such a situation may even be deliberately created. Perhaps the most famous case was Niels Bohr's 1913 theory of atomic structure, in which he simply postulated that there are certain orbits possible for the motion of an electron within the atom where no radiation is emitted; this is in flat contradiction with what classical theory, based on Maxwell's equations, predicts, namely that every charge when accelerated radiates away some of its energy. Bohr, of course, was quite well aware of this; there is in fact so much sound theory and experimental verification behind it that his introduction of a contradictory postulate was an act of remarkable physical intuition and - the term does not seem misplaced- great moral courage. For the astonishing thing is that the hybrid theory worked surprisingly well. It could naturally be bettered as soon as quantum mechanics grew of age, and above all reformulated in a more consistent way. But of course quantum mechanics has in its turn introduced some very extraordinary contradictions into physics, without altogether resolving this one (see Claverie and Diner [7]).

The last type of contradiction - or at least inconsistency- which constantly appears in physics is that between theories. Since this may surprise those who are not specialists in physics (and perhaps some who are), let us examine some examples.

The first —and conceptually the simplest- is offered by the plethora of theories that make up the attempts to build models of the atomic nucleus. In principle, we could solve the Schrödinger equation describing the motion of the particles within the nucleus; in practice this proves impossible, partly because we do not know the force acting between the nuclear particles with sufficient detail, partly because the equation is too complex to yield to presently known mathematical techniques —even with the aid of computers. So we construct models on the basis of simplifying assumptions. Perhaps the best known of these is the shell model, where each particle is taken to move in a common field of force created by its seeing, so to say, the average effect of the other particles; because of this averaging, the motion of each particle appears independent of that of the others. The other extreme in nuclear model-building is the liquid-drop model: here the nucleus is treated as if it were a continuous fluid, and we forget about the existence of individual particles. Both these models have been very successful, each in its own sphere. A number of nuclear properties which neither could explain easily have been treated by means of yet a third model, the so-called collective model, a betwixt-and-between construction which takes many features of the shell model but allows the motion of the nuclear particles to deform the shape of the common field of force. There are still other models, each with its own usefulness: the optical model, the alpha-particle model, the cluster model, the statistical model [i].

Though these are full-fledged theories, the nuclear physicist calls them models, to signal his awareness of the unsatisfactory state of affairs the need for such a multiplicity of theories represents. They are not merely different, they are incompatible to such an extent that the basic assumptions of no two of them agree; thus some even ignore the essentially quantal nature of the nuclear particles, though most draw central features precisely from quantum mechanics. And in spite of the intense and ingenious efforts that have gone into all this model building, every nuclear theorist would welcome the appearance of a genuine nuclear theory that could sweep them all into the dustbin of history. Yet so long as no such theory is visible on the horizon, we must go on using these diverse models. And here lies the awkwardness of the situation for the philosopher of science: for in many cases the theoretical explanation offered for an observed behaviour of nuclei is built on the judicious combination of several such models, in spite of their conflicting bases. In fact, some of the models themselves might be described as just such mixtures of theoretical oil and water. Yet they work, and often very successfully indeed.

A second example is furnished by the relation between thermodynamics and statistical mechanics. Here the inconsistency is much more subtle, and may indeed for most practical purposes be ignored; yet it amply repays analysis. The situation is as follows. Thermodynamics is a theoretical structure whose remarkable internal clarity - put in evidence by the axiomatic reformulation first achieved by Carathéodory [2] (see also Falk and Jung[8]) — cannot hide what we might call its lack of intuitiveness. A significant aspect of this is that it does not appear to have a direct link to other physical theories, while at the same time it is so fundamental that for instance the direction of the flow of time for all of classical physics, at least, is derived from it. The link to the rest of physics is established by underpinning it with statistical mechanics. This enterprise is complete but for one small loophole: the proof that the ensembles of statistical mechanics actually have averages of the required kind depends on the so-called ergodic hypothesis; but for all physically significant types of system this hypothesis remains no more than a postulate with a posteriori justification. A related difficulty arises for the concept of equilibrium: in thermodynamics the notion seems quite clear, but in statistical mechanics its exact meaning is very hard to pin down.

Now because of the practical success of statistical mechanics, and because all discrepancies between the predictions of thermodynamics and statistical mechanics are well understood,many and perhaps most physicist are content to acept without further ado the validity of the ergodic hypothesis. Yet there remains this small but very deep conceptual gap in our understanding; and such problems as the origin of the macroscopic irreversibility for many-body systems in which each body follows a fully reversible microscopic mechanics are evidently related to it[j].

These two examples concerned situations where more than one physical theory existed within the same field; to complete the picture, we will briefly mention a third example of contradiction between theories in different fields, though this case is not really relevant to our theme. It concerns the relations between quantum mechanics and relativity theory —the two great generalizations that twentieth-century physics has to offer. The need for combining them is obvious: there are too many situations of physical (not to mention astrophysical) interest in which subatomic particles, subject to quantum behaviour, move at speeds or through distances such that ordinary Newtonian mechanics is no longer a good approximation and relativity theory must be invoked. But the difficulties in the way of this endeavour have so far won out. So long as we remain within Einstein's special theory the technical problems have largely been solved; at least in its applications to electromagnetic radiation and its interaction with matter —quantum electrodynamics, that is to say— the combination has proved spectacularly successful and has provided us with some of the most accurate predictions of any theory in physics; yet it is still true, as was written eighteen years ago, that "the fusion of these requirements [of special relativity and quantum mechanics] into a non-contradictory theory (in four dimensions) is well known to be a problem whose solution has not been achieved in a non-trivial way even in a model" (Jost [9]). And when we go over to the theory of general relativity, the picture is much bleaker: in spite of an enormous amount of effort concentrated on the problem, no satisfactory way of quantizing it in any of its forms has been found. For there is here a basic conceptual conflict: relativity theory is essentially a non-linear theory of continuously variable quantities. Quantum mechanics, on the other hand, yields discrete spectra and is basically linear; this is pointed up by the central role in it of the superposition principle, which states that the sum of two solutions of the wave equation is again a physically meaningful solution[k].

To conclude this long discussion of conflicting elements within physical theory, it must be noted that the situation is in no sense only temporary. Admittedly, any one of the inconsistencies or contradictions we have mentioned will only persist for a certain time, until the moment, in fact, when theoretical development makes a decisive step forward; so that indeed the elimination of these difficulties may be said to be one of the aims of the physicists' work. Yet every advance, every new idea, every reformulation of theory brings in its wake a number of new difficulties and conflicts which it substitutes for those it solves: for this contradictoriness is at the very least essential to our understanding of nature, built up as it is of a series of part views, each appropriate to its own purposes but not suited for any other one. Hence different theoretical structures must be incompatible to achieve their incompatible aims. It can be argued that this fact merely reflects, at the cognitive level, even deeper reasons; but we shall not pursue this matter here. Certainly these inconsistencies are so much a part of the physicist's way of life that he sees them as normal and may even be unaware of them[ℓ]. And in the present context they are of course relevant, because they make otiose the attempt to axiomatize any theory that exhibits them.

V

What, then, will be the tasks for the axiomatic approach in physics? From what has been said above, it should be apparent that the question must be asked explicitly, since the aims usually proposed for mathematical axiomatization cannot be relevant to physics. Thus the completeness of an axiom base is not of practical interest because, as we have noted, a physical theory is of limited scope; hence only a limited range of consequences drawn from the axiom base is meaningful, and the physicist knows that he is on speculative grounds when he steps outside these limits. If the adding of further axioms to the base causes inconsistencies outside the scope of the theory, this is merely irrelevant; their elimination may be justified under Occam's razor within the scope but cannot constitute a logical requirement. In fact, of the mathematically interesting purposes of axiomatization only the somewhat pedestrian need to check the deductive soundness of the theoretical structure survives. Important though this may be in practice, we need hardly discuss it further here; not only does everyone agree on this matter, but it is essentially a matter of scientific technique rather than fundamental principle.

Are there any other purposes, specific to physics (or, perhaps, the natural sciences), for which the axiomatic method could prove useful? Most writers on axiomatization appear to take it as read that such purposes are self-evident and need not be stated; what follows then are suggestions for the future rather than conclusions from already existing studies.

Firstly, exhibiting the formal structure of a theory —and in particular making explicit its fundamental postulates— may be of great help in understanding it. What is important here, of course, is formulating adequately the theory rather than the axiomatic approach as such, which is only one way of achieving this, though a convenient one. The significance of the Carathéodory approach to thermodynamics lay largely in the fact that it stated unequivocally the concepts fundamental to the theory. This was of great usefulness in determining where it would apply and where not, and led to many fruitful extensions. In a similar way the two distinct but related axiomatizations that von Neumann [(10)] proposed for quantum mechanics retain their significance in all the heated discussions about the interpretation of this theory.

Note here that the recognition of alternative axiomatizations opens the way to a many-sided and flexible understanding. Thus in the case of classical mechanics, we may build an ADS along the lines laid down by Newton; we start with space, time and particles of fixed mass, and take his three laws to make up the vital part of the axiom base. In this way we obtain a very direct access to solving the simpler problems, at the price of facing non-trivial difficulties when going over to, say, continuum mechanics. Alternatively, we may adopt a Lagrangian formalism. This is less intuitive in that the necessary axioms are no longer easily linked to everyday experience; the simpler mechanical problems already require some mathematical sophistication for their solution, in consequence. But we achieve a convenient way for solving the more difficult problems, we are led to new insights concerning the conservation laws by way of Noether's theorem, and the precise meaning of the transition to relativity or quantum mechanics now becomes clear. Both approaches thus have their justification. We also see that there can be alternative axiomatizations in a new sense, wider than the mathematical one which requires demonstrable logical equivalence: the "physical" equivalence of two ADS's is also possible, meaning thereby that within the scope of the theory all deducible consequences are indistinguishable.

Such a situation arises, for instance, in statistical mechanics (see Jaynes [11], Farquhar [12] and Penrose [13]), where several essentially different axiomatizations have been given. The present argument suggests that each of them highlights certain features of the theory: so far from being regarded as competitors, among which a "best one" should be chosen, the alternatives complement each other, and their relationships ought to be studied from this point of view. In quantum mechanics we also have a number of competing axiomatizations (see, for instance, Gudder [14]); but the situation is not quite the same, for they have all been constructed for fairly similar ends, and thus the study of their interrelation (though important enough in its own right) will not shed much light on the vexed questions of the interpretation of quantum mechanics.

Related to the increased understanding of a theory that axiomatizing it may bring about is the usefulness of exhibiting its formal structure for determining its scope, or region of applicability. As I have attempted to show in a forthcoming paper, talking about the probability of a theory's being true is not helpful; the problems raised by this approach are removed or turned into something useful by the scope concept; and the scope is of course essential in discussing any practical application of a theory. Let us look at this a little more closely.

The question whether building an ADS may be of assistance in finding the scope of the theory cannot arise, naturally, in axiomatic mathematics and is peculiar to the experimental sciences. This and certain other aspects of the question where axiomatizing a physical theory could prove important have not so far been taken very seriously; it may be suspected that this is because the great differences between physical and mathematical axiomatics have not been fully appreciated.

How, then, does and ADS help in determining the scope? What in practice <u>can</u> be directly found is the scope of the individual laws deduced from the theory (we are using "law" in the sense of a relation between experimentally measurable quantities which derives from the theory). Experimental work directly decides where the law "works" and where it does not. Now different laws in the same theory need not have the same scope; and we can clearly define an inner scope for the theory which is the intersection of the scopes of all its laws, and an outer scope which is the union —both these terms in their set-theoretical meaning. The inner scope is the range over which all of the theory is valid, the outer scope is the region where at least some of the theory works. But we cannot say that we have examined all possible laws to be derived from the theory, for their number is presumably infinite. Yet if we exhibit the place of those laws whose scope we know is the ADS, it becomes at once evident from this hierarchical structure where a new law might alter the theory's two scope limits and where we may go on deducing as many new laws as could be interesting without affecting the situation. Thus the ADS aids us in determining how far we can state that a theory's scope (either inner or outer) is already well established.

But we can go further. The ADS allows us to see which of the axioms are involved in the deduction of each of the laws whose scopes are experimentally known. We can therefore also find the scopes of the axioms individually (here only the inner scope, the intersection of the scopes of those laws that require the axiom for their deduction, appears to be meaningful). In general the axiom scopes will not coincide, and the "topology" of the situation can be quite complex. Let us examine only one case: where the inner scope of the theory as a whole coincides with the scope of one of the axioms (or perhaps a small set of the axioms). Here it is clear how a better theory, one of wider scope, could be found: by substituting for the limiting axiom another one of ampler scope. In this case, then, the axiomatization turns out to be useful because it suggests a direction for further development; just how this improved theory could be built is, of course, not to be answered by axiomatic or any other techniques, for this requires an effort of the creative imagination. But at least we know along what line to search.

This case will not always be the one we have: but to achieve it we may make use of the freedom we have in setting up the axiom base for the ADS and search for alternative ones which are either logically or at least "physically" equivalent (meaning thereby that within the theory's scope we cannot distinguish their consequences). If in one of these alternatives we can pick out a scope-limiting axiom of the sort just described, then we can obtain a hint towards improving the theory.

But there are certain questions which require further investigation. Does the existence of a scope-limiting axiom have some epistemological significance? Is it desirable that the scopes of the axioms in an ADS should approximately coincide? This last cannot of course be a primary criterion in judging a theory, since it is satisfied for a theory all of whose axioms have null scope; but if its relevance were understood, it could be helpful as a secondary criterion.

In conclusion, then, we see that the axiomatic approach does have a definite role to play in physics; but it differs markedly from the one it has in mathematics, and we must clearly understand the nature of the theory under study before deciding whether to axiomatize or not. Where it is appropriate, axiomatization can achieve significant results; perhaps more so than has been realized so far; but to apply it indiscriminately is to invite disaster. The situation might be summed up by saying that in mathematics we can axiomatize in order to understand; in physics we must understand in order to axiomatize.

NOTES

* A first draft of this paper was written while on sabbatical leave at Oxford University. The support of the Royal Society under its exchange agreement with the Academia de la Investigación Científica is gratefully acknowledged.

a. "The essential or primitive concepts of a theory cannot be discerned with clarity and certainty unless the theory is axiomatized... And as long as there is no clarity concerning the building stones (primitive concepts and axioms) of a theory, discussions on fundamental problems are likely to be confused, because immature".(Bunge[15]; see also Suppes[16]).

b. "In particular, any "axiomatization" of the theory can at best help to avoid trivial contradictions or redundancies in its formal apparatus, but is incapable of throwing any light on the adequacy of the theory as a mode of description of experience. This last problem belongs to what the scholastics pointedly called "real logic", in contradistinction to "formal logic"... Kept within proper bounds, axiomatization of the formalism would not make any difference... When, however, exaggerated claims are made about its powers, disastrous results follow."(Rosenfeld[17]).

c. A rigorous description of axiomatic techniques will be found in many textbooks, e.g. Hilbert & Bernays[18] and Kneebone[19].

d. This must not blind us to its limitations in mathematical research; these are lucidly discussed by Hao Wang[20].

e. The metamathematical literature uses the term "model". I prefer here the word "realization", to avoid confusion with "model" as used below in connection with physical theories.

f. That the author is, professionally, a physicist may also have something to do with it...

g. Note, however, that the situation is not welcome to the geophysicist: the very small amount of information satellite exploration has so far yielded concerning the other planets has had a major impact on geophysics, essentially because it removes the restriction of having only one object to study.

h. An extensive quantity is proportional to the size of the system, i.e. to the total amount of matter in it, while an intensive one is independent of the size.

i. Details concerning these models will be found in the textbooks of nuclear physics (e.g. Bohr and Mottelson[21] or Brown[22]).

j. Ergodic theory (see e.g. Farquhar[12]) is the most widely studied but not the only way to create a physical justification for the use of ensembles in statistical mechanics; others - which generate different but no less intricate problems — are those of Tolman[23], Jaynes[11] and Penrose[13]. For the question or irreversibility see e.g. Mehra and Sudarshan[24].

k. A more detailed discussion of inconsistencies among theories, though from a different point of view, will be found in Tisza[25].

ℓ. To a physicist the assertion that, say, classical mechanics formally contradicts quantum theory, even if true, will appear trivial or even ridiculous. He will prefer to see classical mechanics as a limiting form of quantum behaviour — though the problems of going to this limit are far from trivial.

REFERENCES

1. Hilbert, D., Nachr. K. Ges. Wiss., Göttingen, math.-phys. Kl. (1900) 253.
2. Carathéodory, C., Math. Annalen, 67 (1909) 355.
3. Heyting, A., Sitzber. preuss. Akad. Wiss., phys.-math. Kl. (1930) 42, 57, 158.
4. Bunge, M., Philosophy of Science, Reidel, Dordrecht (1973).
5. Bhaskar, R., A Realist Theory of Science, Harvester Press, Sussex (1978).
6. Scheibe, E., The Logical Analysis of Quantum Mechanics, Pergamon Press, Oxford (1973) p. 14.
7. Claverie, P. and Diner, S., in O. Chalvet et al. (eds.), Localization and Delocalization in Quantum Chemistry, Vol. II, Reidel, Dordrecht (1976) pp. 395, 449, 461.
8. Falk, G. and Jung, H., in S. Flügge (ed.), Handbuch der Physik, Vol III/2, Springer, Berlin (1959) p. 119.
9. Jost, R., in M. Fierz & V.F. Wesskopf (eds.), Theoretical Physics in the Twentieth Century, Interscience (1960) p. 107.
10. Neumann, J.v., Mathematische Grundlagen der Quantenmechanik, Springer, Berlin (1932).
11. Jaynes, E.T., Phys. Rev., 106 (1957) 171.
12. Farquhar, I., Ergodic Theory in Statistical Mechanics, Wiley, New York (1964).
13. Penrose, O., Foundations of Statistical Mechanics, Pergamon, Oxford (1970).
14. Gudder, S.P., in W.C. Price & S.S. Chissick (eds.), The Uncertainty Principle and Foundations of Quantum Mechanics, Wiley, London (1977) p. 247.
15. Bunge, M., in I. Lakatos & A. Musgrave (eds.), Problems in the Philosophy of Physics, North-Holland, Amsterdam (1968) p. 120.
16. Suppes, P., Philos. Sci. 21 (1954) 242.
17. Rosenfeld, L., Nucl. Phys., A108 (1968) 241.
18. Hilbert, D. and Bernays, P. Grundlagen der Mathematik, J. Springer, Berlin (1934/39).
19. Kneebone, G.T., Mathematical Logic and the Foundations of Mathematics, Van Nostrand, London (1963).
20. Hao Wang, A Survey of Mathematical Logic, North-Holland, Amsterdam, ch. III (1963).

21. Bohr, A. and Mottelson, B.R., Nuclear Structure, 2 vols., Benjamin, New York (1969/75).
22. Brown, G.E., Unified Theory of Nuclear Models and Forces, 3rd. edn., North-Holland, Amsterdam (1971).
23. Tolman, R.C., The Principles of Statistical Mechanics, Oxford University Press, Oxford (1938).
24. Mehra, J. and Sudarshan, E.C. G., Nuovo Cimento, 11B (1972) 215.
25. Tisza, L., Rev. Mod. Phys., 35 (1963) 151.