The nature of theoretical concepts and the role of models in an advanced science

By R. B. BRAITHWAITE

The function of every science is to establish laws —true hypotheses— which cover the behaviour of the observable things and events which are the subject-matter of the science, thereby enabling the scientist both to connect together his knowledge of particular events and to enable him to predict what events will happen under certain circumstances. If these circumstances can be produced at will, knowledge of the general laws will enable to some extent the applied scientist to control the course of nature and to construct, machines or other artefacts which will behave in known ways.

An advanced science like physics is not content only with establishing lowest-level generalisations covering physical events: it aims at, and has been largely successful in, subsuming its lowest-level hypotheses, and thus in organising its hypotheses into a hierarchical deductive system —a scientific theory— in which a hypothesis at a lower level is shown to be deducible from a set of hypotheses at a higher level. Notable examples of this are Maxwell's subsumption of the laws of optics under his very general electromagnetic equations and Einstein's subsumption of gravitational laws under his general Theory of Relativity.

The concepts which enter into the higher-level hypotheses of an advanced science are usually concepts (e. g. electric-field vector, electron, Schrödinger wave-function) which are not directly observable things or properties as are those which appear in the lowest-level generalisations of the science: instead they are 'theoretical concepts' which appear at the beginning of the deductive theory but which are eliminated in the course of the deduction. These theoretical concepts present a problem to the philosopher of science: namely, what is their epistemological status? They are clearly in some way empirical concepts —an electron or a Schrödinger wave-function is not an object of pure mathematics like a prime number; but an electron is not observable in the sense in which a flash of light or the pointer reading on a measuring scale is observable. Nevertheless the truth of propositons about electrons is tested by the observable behaviour of measuring instruments; and the question is therefore in what manner an electron or other theoretical concept is an empirical concept.

An answer to this question, an answer implicit in the writings of many philosophers of science such as Ernst Mach and Karl Pearson, was given explicitly by Bertrand Russell in his doctrine of 'logical constructions'. "The supreme maxim in scientific philosophising is this: Wherever possible, logical constructions are to be substituted for inferred entities". *(Mysticism and Logic and other essays,* 1918, p. 155). According to the logical construction view electrons, for example, are logical constructions out of the observed events and objects by which their presence can be detected; this is equivalent to saying that the word "electron" can be explicitly defined in terms of such observations. On this view every sentence containing the word "electron" is translatable, without loss of meaning, into a sentence in which there only occur words which denote entities (events, objects, properties, relations) which are directly observable. It is the business of a philosopher of science to show how these translations are to be made, and thus to show how the theoretical terms of a science can be explicitly defined by means of observable entities. A philosopher of physics should be able to make this translation in the case of the word "electron", and thus be able to exhibit the way in which electrons are logical constructions out of observable entities. Russell's programme of logical construction is similar to the 'operationalist' programme of logical construction is similar to tre 'operationalist' programme proposed by P. W. Bridgman, according to which theoretical terms must be defined by means of the empirical 'operations' involved in their measurement *(The Logic of Modern Physics,* 1927).

This 'logical construction' view of the nature of theoretical concepts was criticised by F. P. Ramsey in some notes which he wrote in 1929, a few months before his death at the age of 26, and which were published posthumously *(The Foundations of Mathematics and other logical essays,* 1931, pp. 212 ff.) Ramsey developed his criticism by constructing a simple example of a scientific theory; I have been able to construct (in my *Scientific Explanation,* 1953, Chapter III) even simpler examples which show precisely the defects of the logical construction view.

The point displayed by both Ramsey's and my examples is that, although it is always possible to define the theoretical terms occurring in the highest-level hypotheses of a theory by means of the terms denoting directly observable entities which occur in the lowest-level generalisation which the theory was propounded to explain, such a definition will prevent the theory from being expanded into a wider theory capable of explaining new lowest-level generalisation which may subsequently be established. To treat theoretical concepts as logical constructions out of observable entities would be to *ossify* the scientific theory in which they occur: the theory would be adequate to cover systematically a particular set of lowest-level generalisations already established, but there would be no hope of extending the theory to explain more generalisations than it was originally designed to explain. A scientific theory to be capable (like all good scientific theories) of this sort of growth must give more freedom of play to its theoretical concepts than the logical construction view will allow them to have

What them, if the logical construction view is inadequate, is the epistemological status of theoretical concepts? A way of answering this question, essentially that of Ramsey, is to say that the status of the theoretical concept electron is given by specifying its place in the deductive system of contemporary physics in the following way: there is a property E (called "being an electron") which is such that certain higher-level hypotheses which are true about containing the property electron there follow certain lowest-level generalisations which are empirically testable. According to this answer, nothing is asserted about the 'nature' of the property E in itself; all that is asserted is that there are instances of E, namely electrons. To say that electrons exist is to assert the truth of the physical theory in which there occurs the concept of being an electron.

There is, however, another way of answering the status-question wich is open to a philosopher of science who, with knowledge of the work of logicians such as Carnap done since Ramsey's death, would wish to make a sharper distinction than was made by Ramsey between a scientific theory arranged as a *deductive system* and the *calculus* (or language) representing the deductive system. So let us consider, not the nature of the theoretical concept electron in the deductively arranged physical theory, but the role played by the term "electron" (or other synonymous symbol) in the calculus representing the theory. This calculus consists of a series of formulae (or sentences) arranged in such a way that all the formulae, except for a small number of formulae (called "initial formulae") are derived from these initial formulae in accordance with the rules of the calculus. The calculus will be interpreted to represent the physical deductive theory by taking the highest-level hypotheses of the theory to be represented by the initial formulae and lower-level hypotheses and generalisations to be represented by derived formulae in the calculus. For this insterpretation to be possible it is necessary that the rules of the calculus should correspond to the logical and mathematical principles of deduction used in making the deductions within the deductive theory.

When the deductive system to be represented by the calculus is a pure one, i. e. a system containing only logically necessary propositions (e. g. a deductive system of arithmetic starting with Peano's axioms), the calculus is interpreted all in a piece. Meanings are attached to the 'primitive' symbols occurring in the initial formulae representing the axioms (e. g. to "number", "successor of", "zero" in Peano's arithmetical calculus), and meanings are attached to all the other symbols used in the calculus by the use of formulae of the calculus interpreted as explicit definitions of these other symbols in terms of the primitive symbols. (These definitions may, of course, be "contextual definitions", i. e. definitions of a whole formula rather than of a separate term.) The meanings attached to the symbols do not depend upon the order of the formulae in the calculus representing the pure deductive system: the meaning, for example, of "prime number" does not depend upon the order in wich theorems about prime numbers are deduced from the axioms of arithmetic.

The situation is quite different in the case in which the deductive system to be represented by the calculus is a scientific theory containing empirical hypotheses at different levels. Here, on my view, the calculus is not interpreted all in a piece: meanings are first attached to the symbols denoting the directly observable properties and relations which occur in the directly testable lowest-level generalisations of the theory. Meaning is then attached to the symbols which are to denote the theoretical concepts of the theory merely by virtue of the fact that they occur in the initial formulae of the calculus representing the theory. The formulae of the calculus are arranged in an order corresponding to the deductive arrangement of the hypotheses of the theory, with the initial formulae corresponding to the highest level hypotheses. The initial formulae are interpreted as representing propositions from which directly testable propositions logically follow, and the symbols (e. g. "electron") which occur in these initial formulae are interpreted as being essential parts of these formulae. No direct meaning is attached to the term "electron": it is given a meaning indirectly by the function it plays in the calculus which is interpreted as representing the physical theory.

On this view the status of the concept electron, and the question "Do electrons really exist?", can only be discussed in terms of the role played by the word "electron" in the exposition of physical theory. This view resembles Russell's logical construction view in that in both cases what is in question is the meaning of the word or other symbol. But whereas Russell would say that the word is given a meaning by being *explicitly* defined by means of words denoting observable entities, I would say that it is given a meaning by showing its place in a calculus representing a scientific theory. This may be called giving an *implicit* or a *contextual* definition of the theoretical terms by means of words denoting observable entities occurring in the formulae calculus which represent the directly testable lowest-level generalisations of the theory; and my account may be taken as an elucidation of such indirect or contextual definition. (It will give a wider sense of "contextual definition" than the usual one, for it is the whole interpreted calculus and not only one sentence (or type of sentence) which will form the 'context' for the purpose of the contextual definition.)

A direction of attention upon the calculus representing a deductive scientific theory will also throw light upon the use of a *model* in thinking about the scientific theory. Suppose that the calculus which is interpreted as representing the theory can also be interpreted as a deductive system in such a way that a direct meaning is given to all the symbols occurring in its initial formulae. In this second interpretation of the calculus the propositions represented will contain no 'theoretical' concepts, and the calculus will be interpreted all in a piece —as in the case of a calculus representing a pure mathematical deductive system—. Thus this interpretation does not present the epistemological difficulty presented by the original interpretation of the calculus as representing the scientific theory. The deductive system which is this second interpretation of the calculus may be regarded as a *model* for the theory. A theory and a model for it have the same formal structure, since they are both represented by one and the same calculus. There is a one-one correlation between the propositions of the theory and those of the model, and the deductive arrangement of the propositions in the theory corresponds to that of the correlated propositions in the model (and vice versa). But the epistemological structure of the theory and of model for it are different: in order to give the model the calculus is directly interpreted all in a piece, whereas to give the theory it is derived formulae of the calculus which are directly interpreted, the earlier formulae being interpreted indirectly by virtue of their place in the calculus.

The fact that a model has the same logical struc ture as, but a simpler epistemological structure than, the theory for which it is a model explains the use of models in thinking about a theory in an advanced science. For thinking about the model will, for many purposes, serve as a substitute for thinking explicitly about the calculus of which both theory and model are interpretations, since the model is a quite straightforward interpretation of the calculus. So to think of the model instead of the calculus in connection with the theory avoids the self-consciousness required in thinking at the same time of a theory and of the language in wich it is expounded, and thus allows of a philosophically unsophisticated approach to an understanding of the logical structure of a scientific deductive theory.

The danger of thinking of a theory by way of thinking of a model for it are, first, that we may, if we are not careful, suppose that there are concepts involved in the theory which correspond to *all* the properties of the objects in the model; we may think, for example, that the electrons in an atom have all the spatial properties of the balls in a 'solar system' model of an atom. A second —more subtle— danger is that it may well happen that some of the propositions in the model which are the interpretations of the initial formulae of the calculus are logically necessary propositions (indeed they may all be logically necessary propositions, in which case the model is a pure mathematical model). We may then be tempted illicitly to transfer the logical necessity of these propositions in the model on to the correlated propositions in the theory, and thus to suppose that some or all of the highest-level hypotheses of the theory are logically necessary instead of contingent. In using models we must never forget that we are engaging in *as-if* thinking: the theoretical concepts in a scientific theory behave as if they were elements in the model, but only in certain respects. To forget the limitations is to misuse the valuable aid to thought provided by the model.

The topics of this paper are discussed at greater length in Chapters III and IV of my book *Scientific Explanation: a study of the function of theory, probability and law in science* (Cambridge: at the University Press: 1953).

King's College
CAMBRIDGE