3

THEORIES

The empirical sciences are systems of theories. The logic of scientific knowledge can therefore be described as a theory of theories.

Scientific theories are universal statements. Like all linguistic representations they are systems of signs or symbols. Thus I do not think it helpful to express the difference between universal theories and singular statements by saying that the latter are 'concrete' whereas theories are *merely* symbolic formulae or symbolic schemata; for exactly the same may be said of even the most 'concrete' statements.*[1]

Theories are nets cast to catch what we call 'the world': to

*[1] This is a critical allusion to a view which I later described as 'instrumentalism' and which was represented in Vienna by Mach, Wittgenstein, and Schlick (*cf.* notes *4 and 7 to section 4, and note 5 to section 27). It is the view that a theory is *nothing but* a tool or an instrument for prediction. I have analysed and criticized it in my papers 'A Note on Berkeley as a Precursor of Mach', *Brit. Journ. Philos. Science* 6, 1953, pp. 26 ff.; 'Three Views Concerning Human Knowledge', in *Contemporary British Philosophy* iii, 1956, edited by H. D. Lewis, pp. 355 ff.; and more fully in my *Postscript*, sections *11 to *15 and *19 to *26. My point of view is, briefly, that our ordinary language is full of theories: that observation is always *observation in the light of theories*; that it is only the inductivist prejudice which leads people to think that there could be a phenomenal language, free of theories, and distinguishable from a 'theoretical language'; and lastly, that the theorist is interested in explanation as such, that is to say, in testable explanatory theories: applications and predictions interest him only for theoretical reasons—because they may be used as *tests* of theories. See also the new appendix *x.

rationalize, to explain, and to master it. We endeavour to make the mesh ever finer and finer.

12 CAUSALITY, EXPLANATION, AND THE DEDUCTION OF PREDICTIONS

To give a *causal explanation* of an event means to deduce a statement which describes it, using as premises of the deduction one or more *universal laws*, together with certain singular statements, the *initial conditions*. For example, we can say that we have given a causal explanation of the breaking of a certain piece of thread if we have found that the thread has a tensile strength of 1 *lb.* and that a weight of 2 *lbs.* was put on it. If we analyse this causal explanation we shall find several constituent parts. On the one hand there is the hypothesis: 'Whenever a thread is loaded with a weight exceeding that which characterizes the tensile strength of the thread, then it will break'; a statement which has the character of a universal law of nature. On the other hand we have singular statements (in this case two) which apply only to the specific event in question: 'The weight characteristic for this thread is 1 *lb.*', and 'The weight put on this thread was 2 *lbs.*'*[1]

We have thus two different kinds of statement, both of which are necessary ingredients of a complete causal explanation. They are (1) *universal statements*, i.e. hypotheses of the character of natural laws, and (2) *singular statements*, which apply to the specific event in question and which I shall call 'initial conditions'. It is from universal statements in conjunction with initial conditions that we *deduce* the singular statement, 'This thread will break'. We call this statement a specific or singular *prediction*.*[2]

The initial conditions describe what is usually called the '*cause*' of the

*[1] A clearer analysis of this example—and one which distinguishes *two* laws as well as two initial conditions—would be the following: 'For every thread of a given structure S (determined by its material, thickness, etc.) there is a characteristic weight w, such that the thread will break if any weight exceeding w is suspended from it.'—'For every thread of the structure S_1, the characteristic weight w_1 equals 1 *lb.*' These are the two universal laws. The two initial conditions are, 'This is a thread of structure S_1' and, 'The weight to be put on this thread is equal to 2 *lbs.*'

*[2] The term 'prediction', as used here, comprises statements about the past ('retrodictions'), or even 'given' statements which we wish to explain ('*explicanda*'); cf. my *Poverty of Historicism*, 1945, p. 133 of the edition of 1957, and the *Postscript*, section *15.

event in question. (The fact that a load of 2 *lbs.* was put on a thread with a tensile strength of 1 *lb.* was the 'cause' of its breaking.) And the prediction describes what is usually called the '*effect*'. Both these terms I shall avoid. In physics the use of the expression '*causal explanation*' is restricted as a rule to the special case in which the universal laws have the form of laws of 'action by contact'; or more precisely, of *action at a vanishing distance*, expressed by differential equations. This restriction will not be assumed here. Furthermore, I shall not make any general assertion as to the universal applicability of this deductive method of theoretical explanation. Thus I shall not assert any '*principle of causality*' (or 'principle of universal causation').

The 'principle of causality' is the assertion that any event whatsoever *can* be causally explained—that it *can* be deductively predicted. According to the way in which one interprets the word 'can' in this assertion, it will be either tautological (analytic), or else an assertion about reality (synthetic). For if 'can' means that it is always logically possible to construct a causal explanation, then the assertion is tautological, since for any prediction whatsoever we can always find universal statements and initial conditions from which the prediction is derivable. (Whether these universal statements have been tested and corroborated in other cases is of course quite a different question.) If, however, 'can' is meant to signify that the world is governed by strict laws, that it is so constructed that every specific event is an instance of a universal regularity or law, then the assertion is admittedly synthetic. But in this case it is *not falsifiable*, as will be seen later, in section 78. I shall, therefore, neither adopt nor reject the 'principle of causality'; I shall be content simply to exclude it, as 'metaphysical', from the sphere of science.

I shall, however, propose a methodological rule which corresponds so closely to the 'principle of causality' that the latter might be regarded as its metaphysical version. It is the simple rule that we are not to abandon the search for universal laws and for a coherent theoretical system, nor ever give up our attempts to explain causally any kind of event we can describe.[1] This rule guides the scientific investigator in his

[1] The idea of regarding the principle of causality as the expression of a rule or of a decision is due to H. Gomperz, *Das Problem der Willensfreiheit*, 1907. Cf. Schlick, *Die Kausalitat in der gegenwartigen Physik, Naturwissenschaften* **19**, 1931, p. 154.

work. The view that the latest developments in physics demand the renunciation of this rule, or that physics has now established that within one field at least it is pointless to seek any longer for laws, is not accepted here.[2] This matter will be discussed in section 78.*[3]

13 STRICT AND NUMERICAL UNIVERSALITY

We can distinguish two kinds of universal synthetic statement: the 'strictly universal' and the 'numerically universal'. It is the *strictly universal statements* which I have had in mind so far when speaking of universal statements—of theories or natural laws. The other kind, the numerically universal statements, are in fact equivalent to certain singular statements, or to conjunctions of singular statements, and they will be classed as singular statements here.

Compare, for example, the following two statements: (a) Of all harmonic oscillators it is true that their energy never falls below a certain amount (*viz.* $h\nu/2$); and (b) Of all human beings now living on the earth it is true that their height never exceeds a certain amount (say

* I feel that I should say here more explicitly that the decision to search for causal explanation is that by which the theoretician adopts his aim—or the aim of theoretical science. His aim is to find *explanatory theories* (if possible, *true* explanatory theories); that is to say, theories which describe certain structural properties of the world, and which permit us to deduce, with the help of initial conditions, the effects to be explained. It was the purpose of the present section to explain, if only very briefly, what we mean by causal explanation. A somewhat fuller statement will be found in appendix *x, and in my *Postscript*, section *15. My explanation of explanation has been adopted by certain positivists or 'instrumentalists' who saw in it an attempt to explain it away—as the assertion that explanatory theories are *nothing but* premises for deducing predictions. I therefore wish to make it quite clear that I consider the theorist's interest in *explanation*—that is, in discovering explanatory theories—as irreducible to the practical technological interest in the deduction of predictions. The theorist's interest in *predictions*, on the other hand, is explicable as due to his interest in the problem whether his theories are true; or in other words, as due to his interest in testing his theories—in trying to find out whether they cannot be shown to be false. See also appendix *x, note 4 and text.

[2] The view here opposed is held for example by Schlick; he writes, *op. cit.* p. 155: '. . . this impossibility . . .' (he is referring to the impossibility of exact prediction maintained by Heisenberg) '. . . means that it is impossible to *search for* that formula.' (Cf. also note 1 to section 78.)

*[3] But see now also chapters *iv to *vi of my *Postscript*.

8 ft.). Formal logic (including symbolic logic), which is concerned only with the theory of deduction, treats these two statements alike as universal statements ('formal' or 'general' implications).[1] I think however that it is necessary to emphasize the difference between them. Statement (a) claims to be true for any place and any time. Statement (b) refers only to a finite class of specific elements within a finite individual (or particular) spatio-temporal region. Statements of this latter kind can, in principle, be replaced by a conjunction of singular statements; for given sufficient time, one can *enumerate* all the elements of the (finite) class concerned. This is why we speak in such cases of 'numerical universality'. By contrast, statement (a), about the oscillators, cannot be replaced by a conjunction of a finite number of singular statements about a definite spatio-temporal region; or rather, it could be so replaced only on the assumption that the world is bounded in time and that there exists only a finite number of oscillators in it. But we do not make any such assumption; in particular, we do not make any such assumption in defining the concepts of physics. Rather we regard a statement of type (a) as an *all-statement*, *i.e.* a universal assertion about an unlimited number of individuals. So interpreted it clearly cannot be replaced by a conjunction of a finite number of singular statements.

My use of the concept of a strictly universal statement (or 'all-statement') stands opposed to the view that every synthetic universal statement must in principle be translatable into a conjunction of a finite number of singular statements. Those who adhere to this view[2] insist that what I call 'strictly universal statements' can never be verified, and they therefore reject them, referring either to their criterion

[1] Classical logic (and similarly symbolic logic or 'logistic') distinguishes universal, particular, and singular statements. A universal statement is one referring to all the elements of some class; a particular statement is one referring to some among its elements; a singular statement is one referring to one given element—an individual. This classification is not based on reasons connected with the logic of knowledge. It was developed with an eye to the technique of inference. We can therefore identify our 'universal statements' neither with the universal statements of classical logic nor with the 'general' or 'formal' implications of logistic (cf. note 6 to section 14). *See now also appendix *x, and my *Postscript*, especially section *15.

[2] Cf. for instance F. Kaufmann, *Bemerkungen zum Grundlagenstreit in Logik und Mathematik*, *Erkenntnis* **2**, 1931, p. 274.

of meaning, which demands verifiability, or to some similar consideration.

It is clear that on any such view of natural laws which obliterates the distinction between singular and universal statements, the problem of induction would seem to be solved; for obviously, inferences from singular statements to merely numerically universal ones may be perfectly admissible. But it is equally clear that the methodological problem of induction is not affected by this solution. For the verification of a natural law could only be carried out by empirically ascertaining every single event to which the law might apply, and by finding that every such event actually conforms to the law—clearly an impossible task.

In any case, the question whether the laws of science are strictly or numerically universal cannot be settled by argument. It is one of those questions which can be settled only by an agreement or a convention. And in view of the methodological situation just referred to, I consider it both useful and fruitful to regard natural laws as synthetic and strictly universal statements ('all-statements'). This is to regard them as non-verifiable statements which can be put in the form: 'Of all points in space and time (or in all regions of space and time) it is true that . . .'. By contrast, statements which relate only to certain finite regions of space and time I call 'specific' or 'singular' statements.

The distinction between strictly universal statements and merely numerically universal statements (*i.e.* really a kind of singular statement) will be applied to synthetic statements only. I may, however, mention the possibility of applying this distinction to analytic statements also (for example, to certain mathematical statements).[3]

14 UNIVERSAL CONCEPTS AND INDIVIDUAL CONCEPTS

The distinction between universal and singular *statements* is closely connected with that between *universal and individual concepts or names*.

It is usual to elucidate this distinction with the help of examples of

[3] Examples: (a) Every natural number has a successor. (b) With the exception of the numbers 11, 13, 17, and 19, all numbers between 10 and 20 are divisible.

the following kind: 'dictator', 'planet', 'H_2O' are universal concepts or universal names. 'Napoleon', 'the earth', 'the Atlantic' are singular or individual concepts or names. In these examples individual concepts or names appear to be characterized either by being proper names, or by having to be defined by means of proper names, whilst universal concepts or names can be defined without the use of proper names.

I consider the distinction between universal and individual concepts or names to be of fundamental importance. Every application of science is based upon an inference from scientific hypotheses (which are universal) to singular cases, *i.e.* upon a deduction of singular predictions. But in every singular statement individual concepts or names must occur.

The individual names that occur in the singular statements of science often appear in the guise of spatio-temporal co-ordinates. This is easily understood if we consider that the *application* of a spatio-temporal system of co-ordinates always involves reference to individual names. For we have to fix its points of origin, and this we can do only by making use of proper names (or their equivalents). The use of the names 'Greenwich' and 'The year of Christ's birth' illustrates what I mean. By this method an arbitrarily large number of individual names may be reduced to a very few.[1]

Such vague and general expressions as 'this thing here', 'that thing over there', etc., can sometimes be used as individual names, perhaps in conjunction with ostensive gestures of some kind; in short, we can use signs which are not proper names but which to some extent are interchangeable with proper names or with individual co-ordinates. But universal concepts too can be indicated, if only vaguely, with the help of ostensive gestures. Thus we may point to certain individual things (or events) and then express by a phrase like 'and other similar things' (or 'and so on') our intention to regard these individuals only as representatives of some class which should properly be given a universal name. There can be no doubt that *we learn the use* of universal

[1] But the units of measurements of the co-ordinate system which first were also established by individual names (the rotation of the earth; the standard metre in Paris) can be defined in principle by means of universal names, for example by means of the wavelength or frequency of the monochromatic light emitted by a certain kind of atoms treated in a certain way.

words, that is their *application* to individuals, by ostensive gestures and by similar means. The logical basis of applications of this kind is that individual concepts may be concepts not only of elements but also of classes, and that they may thus stand to universal concepts not only in a relation corresponding to that of an element to a class, but also in a relation corresponding to that of a sub-class to a class. For example, my dog Lux is not only an element of the class of Viennese dogs, which is an individual concept, but also an element of the (universal) class of mammals, which is a universal concept. And the Viennese dogs, in their turn, are not only a sub-class of the (individual) class of Austrian dogs, but also a sub-class of the (universal) class of mammals.

The use of the word 'mammals' as an example of a universal name might possibly cause misunderstanding. For words like 'mammal', 'dog', etc., are in their ordinary use not free from ambiguity. Whether these words are to be regarded as individual class names or as universal class names depends upon our intentions: it depends upon whether we wish to speak of a race of animals living on our planet (an individual concept), or of a kind of physical bodies with properties which can be described in universal terms. Similar ambiguities arise in connection with the use of concepts such as 'pasteurized', 'Linnean System', and 'Latinism', in so far as it is possible to eliminate the proper names to which they allude (or else, to define them with the help of these proper names).*[1]

The above examples and explanations should make it clear what will here be meant by 'universal concepts' and 'individual concepts'. If I were asked for definitions I should probably have to say, as above: 'An individual concept is a concept in the definition of which proper names (or equivalent signs) are indispensable. If any reference to proper names can be completely eliminated, then the concept is a universal concept.' Yet any such definition would be of very little value, since all that it does is to reduce the idea of an individual concept or name to that of a proper name (in the sense of a name of one individual physical thing).

*[1] 'Pasteurized' may be defined, either, as 'treated according to the advice of M. Louis Pasteur' (or something like this), or else as 'heated to 80 degrees centigrade and kept at this temperature for ten minutes'. The first definition makes 'pasteurized' an individual concept; the second makes it a universal concept. But cp. also note 4, below.

I believe that my usage corresponds fairly closely to the customary use of the expressions 'universal' and 'individual'. But whether or not this is so I certainly consider the distinction here made to be indispensable if we are not to blur the corresponding distinction between universal and singular statements. (There is a complete analogy between the problem of universals and the problem of induction.) The attempt to identify an individual thing *merely* by its universal properties and relations, which appear to belong to it alone and to nothing else, is foredoomed to failure. Such a procedure would describe not a single individual thing but the universal class of all those individuals to which these properties and relations belong. Even the use of a universal spatio-temporal system of co-ordinates would alter nothing.[2] For whether there are any individual things corresponding to a description by means of universal names, and if so how many, must always remain an open question.

In the same way, any attempt to define universal names with the help of individual names is bound to fail. This fact has often been overlooked, and it is widely believed that it is possible to rise by a process called 'abstraction' from individual concepts to universal concepts. This view is a near relation of inductive logic, with its passage from singular statements to universal statements. Logically, these procedures are equally impracticable.[3] It is true that one can obtain classes of individuals in this way, but these classes will still be individual concepts—concepts defined with the help of proper names. (Examples of such individual class-concepts are 'Napoleon's generals', and 'the inhabitants of Paris'.) Thus we see that my distinction between universal names or concepts and individual names or concepts has nothing to do with the distinction between classes and elements. Both universal names and individual names may occur as names of some classes, and also as the names of elements of some classes.

It is therefore not possible to abolish the distinction between

[2] Not 'space and time' in general but individual determinations (spatial, temporal or others) based on proper names are 'principles of individuation'.

[3] Similarly, the 'method of abstraction' used in symbolic logic is unable to accomplish the ascent from individual names to universal names. If the class defined by means of abstraction is defined extensionally with the help of individual names, then it is in its turn an individual concept.

individual concepts and universal concepts with arguments like the following of Carnap's: '. . . this distinction is not justified', he says, because '. . . every concept can be regarded as an individual or universal concept according to the point of view adopted'. Carnap tries to support this by the assertion '. . . that (almost) *all so-called individual concepts are* (names of) *classes*, just like universal concepts'.[4] This last assertion is quite correct, as I have shown, but has nothing to do with the distinction in question.

Other workers in the field of symbolic logic (at one time called 'logistics') have similarly confused the distinction between universal names and individual names with the distinction between classes and their elements.[5] It is certainly permissible to use the term 'universal name' as a synonym for 'name of a class', and 'individual name' as a synonym for 'name of an element'; but there is little to be said for this usage. Problems cannot be solved in this way; on the other hand, this usage may very well prevent one from seeing them. The situation here is quite similar to what we met before when discussing the distinction

[4] Carnap, *Der logische Aufbau der Welt*, p. 213. (Addition made in 1934 while the work was in proof.) In Carnap's *Logical Syntax of Language* (1934; Engl. ed. 1937), the distinction between individual names and universal names does not seem to have been considered; nor does this distinction seem to be expressible in the 'co-ordinate language' which he constructs. One might perhaps think that the 'co-ordinates', being signs of lowest type (cf. pp. 12 f.), are to be interpreted as *individual* names (and that Carnap uses a co-ordinate system defined with the help of individuals). But this interpretation will not do, since Carnap writes (p. 87; see also p. 12 Engl. ed., p. 97, para. 4) that in the language which he uses '. . . all expressions of lowest type are numerical expressions' in the sense that they denote what would fall under Peano's undefined primitive sign 'number' (cf. pp. 31 and 33). This makes it clear that the number signs appearing as co-ordinates are not to be thought of as proper names or individual co-ordinates, but as universals. (They are 'individual' only in a Pickwickian sense, cf. note 3 (b) to section 13.)

[5] The distinction drawn by Russell and Whitehead between individuals (or particulars) and universals has also nothing to do with the distinction here introduced between individual names and universal names. According to Russell's terminology, in the sentence 'Napoleon is a French general', 'Napoleon' is, as in my scheme, an individual, but 'French general' is a universal; but conversely, in the sentence 'Nitrogen is a non-metal', 'non-metal' is, as in my scheme, a universal, but 'nitrogen' is an individual. Moreover, what Russell calls 'descriptions' does not correspond to my 'individual names' since *e.g.* the class of 'geometrical points falling within my body', is for me an individual concept, but cannot be represented by means of a 'description'. Cf. Whitehead and Russell *Principia Mathematica* (2nd edition 1925, vol. I), Introduction to the second edition, II 1, pp. xix, f.

between universal and singular statements. The instruments of symbolic logic are no more adequate for handling the problem of universals than for handling the problem of induction.[6]

15 STRICTLY UNIVERSAL AND EXISTENTIAL STATEMENTS

It is not enough, of course, to characterize universal statements as statements in which no individual names occur. If the word 'raven' is used as a universal name, then, clearly, the statement 'all ravens are black' is a strictly universal statement. But in many other statements such as 'many ravens are black' or perhaps 'some ravens are black' or 'there are black ravens', etc., there also occur only universal names; yet we should certainly not describe such statements as universal.

Statements in which only universal names and no individual names occur will here be called 'strict' or 'pure'. Most important among them are the *strictly universal* statements which I have already discussed. In addition to these, I am especially interested in statements of the form 'there are black ravens', which may be taken to mean the same as 'there exists at least one black raven'. Such statements will be called *strictly or purely existential statements* (or *'there-is' statements*).

The negation of a strictly universal statement is always equivalent to a strictly existential statement and *vice versa*. For example, 'not all ravens are black' says the same thing as 'there exists a raven which is not black', or 'there are non-black ravens'.

[6] The difference between universal and singular statements can also not be expressed in the system of Whitehead and Russell. It is not correct to say that the so-called 'formal' or 'general' implications must be universal statements. For every singular statement can be put in the form of a general implication. For example, the statement 'Napoleon was born in Corsica' can be expressed in the form, $(x)\ (x = N \rightarrow \phi x)$, in words: it is true for all values of x that, if x is identical with Napoleon, then x was born in Corsica.

A *general implication* is written, '$(x)\ (\phi x \rightarrow fx)$', where the 'universal operator', '(x)', can be read: 'It is true for all values of x'; 'ϕx' and 'fx' are '*propositional functions*': (*e.g.* 'x was born in Corsica', without its being said who x is; a propositional function can be neither true nor false). '$\rightarrow$' stands for: 'if it is true that . . . then it is true that . . .' the propositional function ϕx preceding '$\rightarrow$' may be called the *antecedent* or the *conditioning propositional function*, and fx the *consequent propositional function* or the *prediction*; and the *general implication*, $(x)\ (\phi x \rightarrow fx)$, asserts that all values of x which satisfy ϕ also satisfy f.

The theories of natural science, and especially what we call natural laws, have the logical form of strictly universal statements; thus they can be expressed in the form of negations of strictly existential statements or, as we may say, in the form of *non-existence statements* (or 'there-is-not' statements). For example, the law of the conservation of energy can be expressed in the form: 'There is no perpetual motion machine', or the hypothesis of the electrical elementary charge in the form: 'There is no electrical charge other than a multiple of the electrical elementary charge'.

In this formulation we see that natural laws might be compared to 'proscriptions' or 'prohibitions'. They do not assert that something exists or is the case; they deny it. They insist on the non-existence of certain things or states of affairs, proscribing or prohibiting, as it were, these things or states of affairs: they rule them out. And it is precisely because they do this that they are *falsifiable*. If we accept as true one singular statement which, as it were, infringes the prohibition by asserting the existence of a thing (or the occurrence of an event) ruled out by the law, then the law is refuted. (An instance would be, 'In such-and-such a place, there is an apparatus which is a perpetual motion machine'.)

Strictly existential statements, by contrast, cannot be falsified. No singular statement (that is to say, no 'basic statement', no statement of an observed event) can contradict the existential statement, 'There are white ravens'. Only a universal statement could do this. On the basis of the criterion of demarcation here adopted I shall therefore have to treat strictly existential statements as non-empirical or 'metaphysical'. This characterization may perhaps seem dubious at first sight and not quite in accordance with the practice of empirical science. By way of objection, it might be asserted (with justice) that there are theories even in physics which have the form of strictly existential statements. An example would be a statement, deducible from the periodic system of chemical elements, which asserts the existence of elements of certain atomic numbers. But if the hypothesis that an element of a certain atomic number exists is to be so formulated that it becomes testable, then much more is required than a purely existential statement. For example, the element with the atomic number 72 (Hafnium) was not discovered merely on the basis of an isolated purely existential statement. On the contrary, all attempts to find it were in vain until Bohr succeeded in predicting several of its properties by deducing them from his theory. But Bohr's theory and those of its conclusions which were relevant to this element and which helped to bring about its discovery are far from being isolated purely existential statements.*[1] They are strictly universal statements. That my decision to regard strictly existential statements as non-empirical—because they are not falsifiable—is helpful, and also in accordance with ordinary usage, will be seen from its application to probability statements and to the problem of testing them empirically. (Cf. sections 66–68.)

Strict or pure statements, whether universal or existential, are not limited as to space and time. They do not refer to an individual, restricted, spatio-temporal region. This is the reason why strictly existential statements are not falsifiable. We cannot search the whole world in order to establish that something does not exist, has never existed, and will never exist. It is for precisely the same reason that strictly universal statements are not verifiable. Again, we cannot search the whole world in order to make sure that nothing exists which the law forbids. Nevertheless, both kinds of strict statements, strictly existential and strictly universal, are in principle empirically decidable; each, however, in *one way only*: they are *unilaterally decidable*. Whenever it is found that something exists here or there, a strictly existential statement may thereby be verified, or a universal one falsified.

The asymmetry here described, with its consequence, the one-sided falsifiability of the universal statements of empirical science, may now perhaps seem less dubious than it did before (in section 6). We now see that no asymmetry of any purely *logical* relationship is involved. On the contrary, the logical relationships show symmetry. Universal and

*[1] The word 'isolated' has been inserted to avoid misinterpretation of the passage though its tendency, I feel, was clear enough: an *isolated* existential statement is never falsifiable; but if taken *in context* with other statements, an existential statement *may in some cases* add to the empirical content of the whole context: it may enrich the theory to which it belongs, and may add to its degree of falsifiability or testability. In this case, the theoretical system including the existential statement in question is to be described as scientific rather than metaphysical.

existential statements are constructed symmetrically. It is only*2 the line drawn by our criterion of demarcation which produces an asymmetry.

16 THEORETICAL SYSTEMS

Scientific theories are perpetually changing. This is not due to mere chance but might well be expected, according to our characterization of empirical science.

Perhaps this is why, as a rule, only *branches* of science—and these only temporarily—ever acquire the form of an elaborate and logically well-constructed system of theories. In spite of this, a tentative system can usually be quite well surveyed as a whole, with all its important consequences. This is very necessary; for a severe test of a system presupposes that it is at the time sufficiently definite and final in form to make it impossible for new assumptions to be smuggled in. In other words, the system must be formulated sufficiently clearly and definitely to make every new assumption easily recognizable for what it is: a modification and therefore a *revision* of the system.

This, I believe, is the reason why the form of a rigorous system is aimed at. It is the form of a so-called '*axiomatized system*'—the form which Hilbert, for example, was able to give to certain branches of theoretical physics. The attempt is made to collect all the assumptions which are needed, but no more, to form the apex of the system. They are usually called the 'axioms' (or 'postulates', or 'primitive propositions'; no claim to truth is implied in the term 'axiom' as here used). The axioms are chosen in such a way that all the other statements belonging to the theoretical system can be derived from the axioms by purely logical or mathematical transformations.

A theoretical system may be said to be axiomatized if a set of statements, the axioms, has been formulated which satisfies the following four fundamental requirements. (a) The system of axioms must be *free*

*2 The word 'only' here should not be taken too seriously. The situation is quite simple. If it is characteristic of empirical science to look upon *singular* statements as test-statements, then the asymmetry arises from the fact that, *with respect to singular statements*, universal statements are falsifiable only and existential statements verifiable only. See also section *22 of my *Postscript*.

from contradiction (whether self-contradiction or mutual contradiction). This is equivalent to the demand that not every arbitrarily chosen statement is deducible from it.[1] (b) The system must be *independent*, *i.e.* it must not contain any axiom deducible from the remaining axioms. (In other words, a statement is to be called an axiom only if it is not deducible within the rest of the system.) These two conditions concern the axiom system as such; as regards the relation of the axiom system to the bulk of the theory, the axioms should be (c) *sufficient* for the deduction of all statements belonging to the theory which is to be axiomatized, and (d) *necessary*, for the same purpose; which means that they should contain no superfluous assumptions.[2]

In a theory thus axiomatized it is possible to investigate the mutual dependence of various parts of the system. For example, we may investigate whether a certain part of the theory is derivable from some part of the axioms. Investigations of this kind (of which more will be said in sections 63 and 64, and 75 to 77) have an important bearing on the problem of falsifiability. They make it clear why the falsification of a logically deduced statement may sometimes not affect the whole system but only some part of it, which may then be regarded as falsified. This is possible because, although the theories of physics are in general not completely axiomatized, the connections between its various parts may yet be sufficiently clear to enable us to decide which of its sub-systems are affected by some particular falsifying observation.*[1]

17 SOME POSSIBILITIES OF INTERPRETING A SYSTEM OF AXIOMS

The view of classical rationalism that the 'axioms' of certain systems, *e.g.* those of Euclidean geometry, must be regarded as immediately or intuitively certain, or self-evident, will not be discussed here. I will only mention that I do not share this view. I consider two different interpretations of any system of axioms to be admissible. The axioms

[1] Cf. section 24.

[2] Regarding these four conditions, and also the following section, see, for example, the somewhat different account in Carnap's *Abriss der Logistik*, 1929, pp. 70 ff.

*[1] The point is more fully discussed in my *Postscript*, especially section *22.

may be regarded either (i) as *conventions*, or they may be regarded (ii) as empirical or scientific *hypotheses*.

(i) If the axioms are regarded as *conventions* then they tie down the use or meaning of the fundamental ideas (or primitive terms, or concepts) which the axioms introduce; they determine what can and what cannot be said about these fundamental ideas. Sometimes the axioms are described as '*implicit definitions*' of the ideas which they introduce. This view can perhaps be elucidated by means of an analogy between an axiomatic system and a (consistent and soluble) system of equations.

The admissible values of the 'unknowns' (or variables) which appear in a system of equations are in some way or other determined by it. Even if the system of equations does not suffice for a unique solution, it does not allow every conceivable combination of values to be substituted for the 'unknowns' (variables). Rather, the system of equations characterizes certain combinations of values or value-systems as admissible, and others as inadmissible; it distinguishes the class of admissible value systems from the class of inadmissible value systems. In a similar way, systems of concepts can be distinguished as admissible or as inadmissible by means of what might be called a 'statement-equation'. A statement-equation is obtained from a propositional function or statement-function (*cf.* note 6 to section 14); this is an incomplete statement, in which one or more 'blanks' occur. Two examples of such propositional functions or statement functions are: 'An isotope of the element x has the atomic weight 65'; or '$x + y = 12$'. Every such statement-function is transformed into a *statement* by the substitution of certain values for the blanks, x and y. The resulting statement will be either true or false, according to the values (or combination of values) substituted. Thus, in the first example, substitution of the word 'copper' or 'zinc' for 'x' yields a true statement, while other substitutions yield false ones. Now what I call a 'statement-equation' is obtained if we decide, with respect to some statement-function, to admit only such values for substitution as turn this function into a *true statement*. By means of this statement-equation a definite class of admissible value-systems is defined, namely the class of those which satisfy it. The analogy with a mathematical equation is clear. If our second example is interpreted, not as a statement-function but as a statement-equation, then it becomes an equation in the ordinary (mathematical) sense.

Since its undefined fundamental ideas or primitive terms can be regarded as blanks, an axiomatic system can, to begin with, be treated as a system of statement-functions. But if we decide that only such systems or combinations of values may be substituted as will satisfy it, then it becomes a system of statement-equations. As such it implicitly defines a class of (admissible) systems of concepts. Every system of concepts which satisfies a system of axioms can be called a *model of that system of axioms*.*[1]

The interpretation of an axiomatic system as a system of (conventions or) implicit definitions can also be expressed by saying that it amounts to the decision: only models may be admitted as substitutes.*[2] But if a model is substituted then the result will be a system of analytic statements (since it will be true by convention). An axiomatic system interpreted in this way cannot therefore be regarded as a system of empirical or scientific hypotheses (in our sense) since it cannot be refuted by the falsification of its consequences; for these too must be analytic.

(ii) How then, it may be asked, can an axiomatic system be interpreted as a system of empirical or scientific *hypotheses*? The usual view is that the primitive terms occurring in the axiomatic system are not to be regarded as implicitly defined, but as 'extra-logical constants'. For example, such concepts as 'straight line' and 'point', which occur in every axiom system of geometry, may be interpreted as 'light ray' and 'intersection of light rays'. In this way, it is thought, the statements of the axiom system become statements about empirical objects, that is to say, synthetic statements.

At first sight, this view of the matter may appear perfectly satisfactory. It leads, however, to difficulties which are connected with the problem of the empirical basis. For it is by no means clear what would

*[1] See note *2.

*[2] Today I should clearly distinguish between the *systems of objects* which satisfy an axiom system and the *system of names of these objects* which may be substituted in the axioms (rendering them true); and I should call only the first system a 'model'. Accordingly, I should now write: 'only names of objects which constitute a model may be admitted for substitution'.

be an *empirical way of defining a concept*. It is customary to speak of 'ostensive definitions'. This means that a definite empirical meaning is assigned to a concept by *correlating it* with certain objects belonging to the real world. It is then regarded as a symbol of those objects. But it should have been clear that only individual names or concepts can be fixed by ostensively referring to 'real objects'—say, by pointing to a certain thing and uttering a name, or by attaching to it a label bearing a name, etc. Yet the concepts which are to be used in the axiomatic system should be universal names, which cannot be defined by empirical indications, pointing, etc. They can be defined if at all only explicitly, *with the help of other universal names*; otherwise they can only be left undefined. That some universal names should remain undefined is therefore quite unavoidable; and herein lies the difficulty. For these undefined concepts can always be used in the non-empirical sense (i), *i.e.* as if they were implicitly defined concepts. Yet this use must inevitably destroy the empirical character of the system. This difficulty, I believe, can only be overcome by means of a methodological decision. I shall, accordingly, adopt a rule not to use undefined concepts as if they were implicitly defined. (This point will be dealt with below in section 20.)

Here I may perhaps add that it is usually possible for the primitive concepts of an axiomatic system such as geometry to be correlated with, or interpreted by, the concepts of another system, *e.g.* physics. This possibility is particularly important when, in the course of the evolution of a science, one system of statements is being *explained* by means of a new—a more general—system of hypotheses which permits the deduction not only of statements belonging to the first system, but also of statements belonging to other systems. In such cases it may be possible to define the fundamental concepts of the new system with the help of concepts which were originally used in some of the old systems.

18 LEVELS OF UNIVERSALITY. THE *MODUS TOLLENS*

We may distinguish, within a theoretical system, statements belonging to various levels of universality. The statements on the highest level of universality are the axioms; statements on the lower levels can be deduced from them. Higher level empirical statements have always the character of hypotheses relative to the lower level statements deducible from them: they can be falsified by the falsification of these less universal statements. But in any hypothetical deductive system, these less universal statements are themselves still strictly universal statements, in the sense here understood. Thus they too must have the character of *hypotheses*—a fact which has often been overlooked in the case of lower-level universal statements. Mach, for example, calls[1] Fourier's theory of heat conduction a 'model theory of physics' for the curious reason that 'this theory is founded not on a *hypothesis* but on an *observable fact*'. However, the 'observable fact' to which Mach refers is described by him by the statement. ' . . . the velocity of the levelling out of temperature differences, provided these differences of temperature are small, is proportional to these differences themselves'—an all-statement whose hypothetical character should be sufficiently conspicuous.

I shall say even of some singular statements that they are hypothetical, seeing that conclusions may be derived from them (with the help of a theoretical system) such that the falsification of these conclusions may falsify the singular statements in question.

The falsifying mode of inference here referred to—the way in which the falsification of a conclusion entails the falsification of the system from which it is derived—is the *modus tollens* of classical logic. It may be described as follows:*[1]

Let p be a conclusion of a system t of statements which may consist of theories and initial conditions (for the sake of simplicity I will not distinguish between them). We may then symbolize the relation of derivability (analytical implication) of p from t by '$t \rightarrow p$' which may

[1] Mach, *Principien der Wärmelehre*, 1896, p. 115.

*[1] In connection with the present passage and two later passages (cf. notes *1 to section 35 and *1 to section 36) in which I use the symbol ' $\rightarrow$ ', I wish to say that when writing the book, I was still in a state of confusion about the distinction between a conditional statement (if-then-statement; sometimes called, somewhat misleadingly, 'material implication') and a statement about deducibility (or a statement asserting that some conditional statement is logically true, or analytic, or that its antecedent entails its consequent)—a distinction which I was taught to understand by Alfred Tarski, a few months after the publication of the book. The problem is not very relevant to the context of the book; but the confusion should be pointed out nevertheless. (These problems are discussed more fully, for example, in my paper in *Mind*, **56**, 1947, pp. 193 ff.)

be read: 'p follows from t'. Assume p to be false, which we may write '$\bar{p}$', to be read 'not-p'. Given the relation of deducibility, $t \rightarrow p$, and the assumption $\bar{p}$, we can then infer $\bar{t}$ (read 'not-t'); that is, we regard t as falsified. If we denote the conjunction (simultaneous assertion) of two statements by putting a point between the symbols standing for them, we may also write the falsifying inference thus: $((t \rightarrow p).\bar{p}) \rightarrow \bar{t}$, or in words: 'If p is derivable from t, and if p is false, then t also is false'.

By means of this mode of inference we falsify *the whole system* (the theory as well as the initial conditions) which was required for the deduction of the statement p, *i.e.* of the falsified statement. Thus it cannot be asserted of any one statement of the system that it is, or is not, specifically upset by the falsification. Only if p is *independent* of some part of the system can we say that this part is not involved in the falsification.[2] With this is connected the following possibility: we may, in some cases, perhaps in consideration of the *levels of universality*, attribute the falsification to some definite hypothesis—for instance to a newly introduced hypothesis. This may happen if a well-corroborated theory, and one which continues to be further corroborated, has been deductively explained by a new hypothesis of a higher level. The attempt will have to be made to test this new hypothesis by means of some of its consequences which have not yet been tested. If any of these are falsified, then we may well attribute the falsification to the new hypothesis alone. We shall then seek, in its stead, other high-level generalizations, but we shall not feel obliged to regard the old system, of lesser generality, as having been falsified. (Cf. also the remarks on 'quasi-induction' in section 85.)

[2] Thus we cannot at first know which among the various statements of the remaining sub-system t' (of which p is not independent) we are to blame for the falsity of p; which of these statements we have to alter, and which we should retain. (I am not here discussing interchangeable statements.) It is often only the scientific instinct of the investigator (influenced, of course, by the results of testing and re-testing) that makes him guess which statements of t' he should regard as innocuous, and which he should regard as being in need of modification. Yet it is worth remembering that it is often the modification of what we are inclined to regard as obviously innocuous (because of its complete agreement with our normal habits of thought) which may produce a decisive advance. A notable example of this is Einstein's modification of the concept of simultaneity.