Carl G. Hempel

SCIENTIFIC EXPLANATION

Essays in the Philosophy of Science

THE FREE PRESS, *NEW YORK*

COLLIER-MACMILLAN LIMITED, *LONDON*

8. THE THEORETICIAN'S DILEMMA: A STUDY IN THE LOGIC OF THEORY CONSTRUCTION

We only cover sections 5, 9, 10. Sections 1 and 2 can be read to contrast Hempel's presentations with Carnap's.

1. DEDUCTIVE AND INDUCTIVE SYSTEMATIZATION

SCIENTIFIC RESEARCH in its various branches seeks not merely to record particular occurrences in the world of our experience: it tries to discover regularities in the flux of events and thus to establish general laws which may be used for prediction, postdiction,[1] and explanation.

The principles of Newtonian mechanics, for example, make it possible, given the present positions and momenta of the celestial objects that make up the solar system, to predict their positions and momenta for a specified future time or to postdict them for a specified time in the past; similarly, those principles permit an explanation of the present positions and momenta by reference to those at some earlier time. In addition to thus accounting for particular facts, the principles of Newtonian mechanics also explain certain "general facts,"

1. This term was suggested by a passage in Reichenbach (1944), where the word 'postdictability' is used to refer to the possibility of determining "past data in terms of given observations" (p. 13). In a similar context, Ryle uses the term 'retrodict' (see for example 1949, p. 124), and Walsh speaks of the historian's business "to 'retrodict' the past: to establish, on the basis of present evidence, what the past must have been like" (1951, p. 41). According to a remark in Acton's review of Walsh's book (*Mind*, vol. 62 (1953), pp. 564–65), the word 'retrodiction' was used in this sense already by J. M. Robertson in *Buckle and his Critics* (1895).

This article is reprinted, with some changes, by kind permission of the publisher, from *Minnesota Studies in the Philosophy of Science*, vol. II. Edited by Herbert Feigl, Michael Scriven, and Grover Maxwell, University of Minnesota Press, Minneapolis. Copyright 1958 by the University of Minnesota.

i.e., empirical uniformities such as Kepler's laws of planetary motion; for the latter can be deduced from the former.[2]

Scientific explanation, prediction, and postdiction all have the same logical character: they show that the fact under consideration can be inferred from certain other facts by means of specified general laws. In the simplest case, this type of argument may be schematized as a deductive inference of the following form:

$$\frac{\begin{matrix} C_1, C_2 \ldots C_k \\ L_1, L_2 \ldots L_r \end{matrix}}{E} \qquad (1.1)$$

Here, $C_1, C_2 \ldots C_k$ are statements of particular occurrences (e.g., of the positions and momenta of certain celestial bodies at a specified time), and $L_1, L_2 \ldots L_r$ are general laws (e.g., those of Newtonian mechanics); finally, E is a sentence stating whatever is being explained, predicted, or postdicted. And the argument has its intended force only if its conclusion, E, follows deductively from the premises.[3]

While explanation, prediction, and postdiction are alike in their logical structure, they differ in certain other respects. For example, an argument of the form (1.1) will qualify as a prediction only if E refers to an occurrence at a time later than that at which the argument is offered; in the case of a postdiction, the event must occur before the presentation of the argument. These differences, however, require no fuller study here, for the purpose of the preceding discussion was simply to point out the role of general laws in scientific explanation prediction, and postdiction.

For these three types of scientific procedure, I will use the common term '*(deductive) systematization*'. More precisely, that term will be used to refer, first to any argument of the form (1.1) that meets the requirements indicated above, no matter whether it serves as an explanation, a prediction, a postdiction, or in still some other capacity; second, to the procedure of establishing arguments of the kind just characterized.

So far, we have considered only those cases of explanation, prediction, and related procedures which can be construed as deductive arguments. There are many instances of scientific explanation and prediction, however, which do not fall into a strictly deductive pattern. For example, when Johnny comes

2. More accurately: it can be deduced from the principles of Newtonian mechanics that Kepler's laws hold in approximation, namely, on the assumption that the forces exerted upon the planets by celestial objects other than the sun (especially other planets) are negligible.

3. (added in 1964). For a fuller discussion of this schema and for certain qualifications concerning the structural identity of explanatory and predictive arguments, see the essay "Aspects of Scientific Explanation" in this volume.

down with the measles, this might be explained by pointing out that he caught the disease from his sister, who is just recovering from it. The particular antecedent facts here invoked are that of Johnny's exposure and, let us assume, the further fact that Johnny had not had the measles previously. But to connect these with the event to be explained, we cannot adduce a general law to the effect that under the specified circumstances, the measles is invariably transmitted to the exposed person: what can be asserted is only a high probability (in the sense of statistical frequency) of transmission. The same type of argument can be used also for predicting or postdicting the occurrence of a case of the measles.

Similarly, in a psychoanalytic explanation of the neurotic behavior of an adult by reference to certain childhood experiences, the generalizations which might be invoked to connect the antecedent events with those to be explained can be construed at best as establishing more or less high probabilities for the connections at hand, but surely not as expressions of unexceptional uniformities.

Explanations, predictions, and postdictions of the kind here illustrated differ from those previously discussed in two important respects: The laws invoked are of a different form, and the statement to be established does not follow deductively from the explanatory statements adduced. We will now consider these differences somewhat more closely.

The laws referred to in connection with the schema (1.1), such as the laws of Newtonian mechanics, are what we will call *statements of strictly universal form*, or *strictly universal statements*. A statement of this kind is an assertion—which may be true or false—to the effect that all cases which meet certain specified conditions will unexceptionally have such and such further characteristics. For example, the statement 'All crows are black' is a sentence of strictly universal form; and so is Newton's first law of motion, that any material body which is not acted upon by an external force persists in its state of rest or of rectilinear motion at constant speed.

The laws invoked in the second type of explanatory and related arguments, on the other hand, are, as we will say, of *statistical form*; they are *statistical probability statements*. A statement of this kind is an assertion—which may be true or false—to the effect that for cases which meet conditions of a specified kind, the probability of having such and such further characteristics is so-and-so much.[4]

4. The distinction here made concerns, then, exclusively the *form* of the statements under consideration and not their truth status nor the extent to which they are supported by empirical evidence. If it were established, for example, that actually only 80 per cent of all crows are black, this would not show that 'All crows are black', or S_1 for short, was a statistical probability statement, but rather that it was a false statement of strictly universal form, and that 'The probability for a crow to be black is .8,' or S_2 for short, was a true statement of statistical form.

(Continued overleaf)

To put the distinction in a nutshell: A strictly universal statement of the simplest kind has the form 'All cases of *P* are cases of *Q*'; a statistical probability statement of the simplest kind has the form 'The probability for a case of *P* to be a case of *Q* is *r*.' While the former implies an assertion about any particular instance of *P*—namely, that it is also an instance of *Q*—the latter implies no similar assertion concerning any particular instance of *P* or even concerning any finite set of such instances.[5] This circumstance gives rise to the second distinctive characteristic mentioned above: the statement *E* describing the phenomenon being explained, predicted, or postdicted (for example, Johnny's catching the measles) is not logically deducible from the explanatory statements adduced [for example, (C_1) Johnny was exposed to the measles; (C_2) Johnny had not previously had the measles; (*L*) For persons who have not previously had the measles and are exposed to it, the probability is .92 that they will contract the disease]; rather, on the assumption that the explanatory statements adduced are true, it is very likely, though not certain, that *E* is true as well. This kind of argument, therefore, is inductive rather than strictly deductive in character: it offers the conclusion *E* on the basis of other statements which constitute only partial, if strongly supporting, grounds for it. An argument of this kind—no matter whether it is used for explanation, prediction, or postdiction, or for yet another purpose—will be called an *inductive systematization*. In particular, we will assume of an inductive systematization that the conclusion is not logically implied by the premises.[6] Again, the procedure of establishing an argument of the kind just described will also be called inductive systematization.

By way of further illustration, let us note here two explanatory arguments

5. For a fuller discussion of this point, see, for example, Nagel (1939, section 7), Reichenbach (1949, sections 63–67), Cramér (1946, Chapter 13).

6. The explanatory and predictive use of statistical laws constitutes perhaps the most important type of inductive systematization; but the occurrence of such laws among the premises is not required by our general concept of inductive systematization. And indeed, as Carnap (1950, pp. 574–75) has pointed out, it is sometimes possible to make predictions of an inductive character exclusively on the basis of information about a finite set of particular cases, without the mediation of any laws whatever. For example, information to the effect that a large sample of instances of P has been examined, that all of its elements have the characteristic Q, and that a certain case x, not included in the sample, is an instance of P, will lend high inductive support to the prediction that x, too, has the characteristic Q. Also, it is sometimes possible to base an inductive systematization on a set of premises which include one or more strictly universal statements, but no statistical laws. An example of such a systematization will be found in Section 9, in the prediction based on the formulas (9.6)-(9.12).

Furthermore, to be sure, neither S_1 nor S_2 can ever be established conclusively: they can only be more or less well supported by available evidence; each of them thus has a more or less high logical, or inductive, probability, relative to that evidence. But this again does not affect at all the fact that S_1 is of strictly universal and S_2 of statistical form.

which are of the inductive kind just characterized. They are adduced by von Mises in a statement to the effect that the everyday notion of causal explanation will eventually adjust itself to changes in the logical form of scientific theories (especially to the use of statistical probability statements as explanatory principles): "We think," von Mises says, that "people will gradually come to be satisfied by causal statements of this kind: It is *because* the die was loaded that the 'six' shows more frequently (but we do not know what the next number will be); or: *Because* the vacuum was heightened and the voltage increased, the radiation became more intense (but we do not know the precise number of scintillations that will occur in the next minute)."[7] Clearly, both of these statements can be construed as inductive explanations of certain physical phenomena.

All the cases of scientific systematization we have considered share this characteristic: they make use of general laws or general principles either of strictly universal or of statistical form. These general laws have the function of establishing systematic connections among empirical facts in such a way that with their help some empirical occurrences may be inferred, by way of explanation, prediction, or postdiction, from other such occurrences. When, in an explanation, we say that the event described by *E* occurred "because" of the circumstances detailed in $C_1, C_2 \ldots C_k$, that phrase has significance if it can be construed as referring to general laws which render $C_1, C_2 \ldots C_k$ relevant to *E* in the sense that, granted the truth of the former, they make the truth of the latter either certain (as in a deductive systematization) or inductively probable (as in an inductive systematization). It is for this reason that the establishment of general laws is of crucial importance in the empirical sciences.

2. OBSERVABLES AND THEORETICAL ENTITIES

Scientific systematization is ultimately aimed at establishing explanatory and predictive order among the bewilderingly complex "data" of our experience, the phenomena that can be "directly observed" by us. It is a remarkable fact, therefore, that the greatest advances in scientific systematization have not been accomplished by means of laws referring explicitly to *observables*, i.e., to things and events which are ascertainable by direct observation, but rather by means of laws that speak of various *hypothetical*, or *theoretical, entities*, i.e., presumptive objects, events, and attributes which cannot be perceived or otherwise directly observed by us.

7. Mises (1951, p. 188). Whether it is advisable to refer to explanations of this kind as causal is debatable: since the classical conception of causality is intimately bound up with the idea of strictly universal laws connecting cause and effect, it might be better to reserve the term 'causal explanation' for some of those explanatory arguments of form (1.1) in which all the laws invoked are of strictly universal form.

For a fuller discussion of this point, it will be helpful to refer to the familiar rough distinction between two levels of scientific systematization: the level of *empirical generalization*, and the level of *theory formation*.[8] The early stages in the development of a scientific discipline usually belong to the former level, which is characterized by the search for laws (of universal or statistical form) which establish connections among the directly observable aspects of the subject matter under study. The more advanced stages belong to the second level, where research is aimed at comprehensive laws, in terms of hypothetical entities, which will account for the uniformities established on the first level. On the first level, we find everyday physical generalizations such as 'Where there is light there is heat', 'Iron rusts in damp air', 'Wood floats on water, iron sinks in it'; but we might assign to it also such more precise quantitative laws as Galileo's, Kepler's, Hooke's, and Snell's laws, as well as botanical and zoological generalizations about the concomitance of certain observable anatomical, physical, functional, and other characteristics in the members of a given species; generalizations in psychology that assert correlations among diverse observable aspects of learning, of perception, and so forth; and various descriptive generalizations in economics, sociology, and anthropology. All these generalizations, whether of strictly universal or of statistical form, purport to express regular connections among directly observable phenomena, and they lend themselves, therefore, to explanatory, predictive, and postdictive use.

On the second level, we encounter general statements that refer to electric, magnetic, and gravitational fields, to molecules, atoms, and a variety of subatomic particles; or to ego, id, superego, libido, sublimation, fixation, and transference; or to various not directly observable entities invoked in recent learning theories.

In accordance with the distinction here made, we will assume that the (extralogical) vocabulary of empirical science, or of any of its branches, is divided into two classes: *observational terms* and *theoretical terms*. In regard to an observational term it is possible, under suitable circumstances, to decide by means of direct observation whether the term does or does not apply to a given situation.

Observation may here be construed so broadly as to include not only perception, but also sensation and introspection; or it may be limited to the perception of what in principle is publicly ascertainable, i.e., perceivable also by others. The subsequent discussion will be independent of how narrowly or how liberally the notion of observation is construed; it may be worth noting,

8. Northrop (1947, Chapters III and IV), for example, presents this distinction very suggestively; he refers to the two levels as "the natural history stage of inquiry" and "the stage of deductively formulated theory." A lucid and concise discussion of the idea at hand will be found in Feigl (1948).

however, that empirical science aims for a system of publicly testable statements, and that, accordingly, the observational data whose correct prediction is the hallmark of a successful theory are at least thought of as couched in terms whose applicability in a given situation different individuals can ascertain with high agreement, by means of direct observation. Statements which purport to describe readings of measuring instruments, changes in color or odor accompanying a chemical reaction, verbal or other kinds of overt behavior shown by a given subject under specified observable conditions—these all illustrate the use of *intersubjectively applicable* observational terms.[9]

Theoretical terms, on the other hand, usually purport to refer to not directly observable entities and their characteristics; they function, in a manner soon to be examined more closely, in scientific theories intended to explain empirical generalizations.

The preceding characterization of the two vocabularies is obviously vague; it offers no precise criterion by means of which any scientific term may be unequivocally classified as an observational term or as a theoretical one. But no such precise criterion is needed here; the questions to be examined in this essay are independent of precisely where the dividing line between the terms of the observational and the theoretical vocabularies is drawn.

3. WHY THEORETICAL TERMS?

The use of theoretical terms in science gives rise to a perplexing problem: Why should science resort to the assumption of hypothetical entities when it is interested in establishing predictive and explanatory connections among observables? Would it not be sufficient for the purpose, and much less extravagant at that, to search for a system of general laws mentioning only observables, and thus expressed in terms of the observational vocabulary alone?

Many general statements in terms of observables have indeed been formulated; they constitute the empirical generalizations mentioned in the preceding

9. In his essay on Skinner's analysis of learning (in Estes *et al.* 1945), Verplanck throws an illuminating sidelight on the importance, for the observational vocabulary (the terms of the data-language, as he calls it), of high uniformity of use among different experimenters. Verplanck argues that while much of Skinner's data-language is sound in this respect, it is "contaminated" by two kinds of term that are not suited for the description of objective scientific data. The first kind includes terms "that cannot be successfully used by many others"; the second kind includes certain terms that should properly be treated as higher-order theoretical expressions.

The nonprecise and pragmatic character of the requirement of intersubjective uniformity of use is nicely reflected in Verplanck's conjecture "that if one were to work with Skinner, and read his records with him, he would find himself able to make the same discriminations as does Skinner and hence eventually give some of them at least data-language status" (*loc. cit.*, p. 279n).

section. But, vexingly, many if not all of them suffer from definite shortcomings: they usually have a rather limited range of application; and even within that range, they have exceptions, so that actually they are not true general statements. Take for example, one of our earlier illustrations:

(3.1) Wood floats on water; iron sinks in it.

This statement has a narrow range of application in the sense that it refers only to wooden and iron objects and concerns their floating behavior only in regard to water.[10] And, what is even more serious, it has exceptions: certain kinds of wood will sink in water, and a hollow iron sphere of suitable dimensions will float on it.

As the history of science shows, flaws of this kind can often be remedied by attributing to the subject matter under study certain further characteristics which, though not open to direct observation, are connected in specified ways with its observable aspects, and which make it possible to establish systematic connections among the latter. For example, a generalization much more satisfactory than (3.1) is obtained by means of the concept of the specific gravity of a body x, which is definable as the quotient of its weight and its volume:

(3.2) Def. $s(x) = w(x)/v(x)$

Let us assume that w and v have been characterized operationally, i.e., in terms of the directly observable outcomes of specified measuring procedures, and that therefore they are counted among the observables. Then s, as determined by (3.2), might be viewed as a characteristic that is less directly observable; and, just for the sake of obtaining a simple illustration, we will classify s as a hypothetical entity. For s, we may now state the following generalization, which is a corollary of the principle of Archimedes:

(3.3) A solid body floats on a liquid if its specific gravity is less than that of the liquid.

This statement avoids, first of all, the exceptions we noted above as refuting (3.1); it predicts correctly the behavior of a piece of heavy wood and of a hollow iron sphere. Moreover, it has a much wider scope: it refers to any kind of solid object and concerns its floating behavior in regard to any liquid. Even the new

10. It should be mentioned, however, that the idea of the range of application of a generalization is here used in an intuitive sense which it would be difficult to explicate. The range of application of (3.1), for example, might plausibly be held to be narrower than here indicated: it might be construed as consisting only of wooden-objects-placed-in-water and iron-objects-placed-in-water. On the other hand, (3.1) may be equivalently restated thus: Any object whatever has the two properties of either not being wood or floating on water, and of either not being iron or sinking in water. In this form, the generalization might be said to have the largest possible range of application, the class of all objects whatsoever.

generalization has certain limitations, of course, and thus invites further improvement. But instead of pursuing this process, let us now examine more closely the way in which a systematic connection among observables is achieved by the law (3.3), which involves a detour through the domain of unobservables.

Suppose that we wish to predict whether a certain solid object b will float on a given body l of liquid. We will then first have to ascertain, by appropriate operational procedures, the weight and the volume of b and l. Let the results of these measurements be expressed by the following four statements O_1, O_2, O_3, O_4:

(3.4) O_1: $w(b) = w_1$; O_2: $v(b) = v_1$
O_3: $w(l) = w_2$; O_4: $v(l) = v_2$

where w_1, w_2, v_1, v_2, are certain positive real numbers. By means of the definition (3.2), we can infer, from (3.4), the specific gravities of b and l:

(3.5) $s(b) = w_1/v_1$; $s(l) = w_2/v_2$

Suppose now that the first of these values is less than the second; then (3.4), via (3.5) implies that

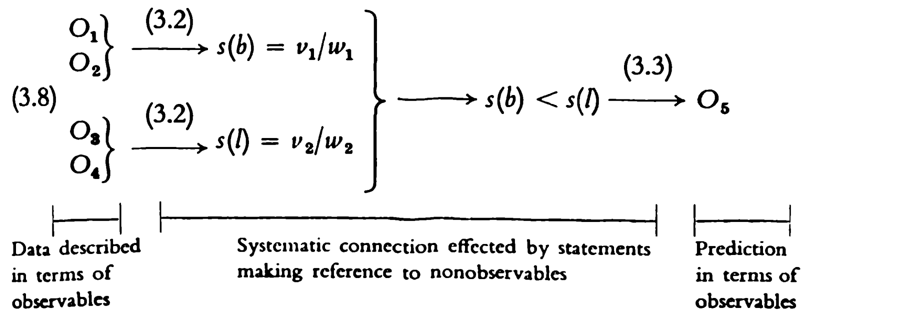
(3.6) $s(b) < s(l)$

By means of the law (3.3), we can now infer that

(3.7) b floats on l

This sentence will also be called O_5. The sentences O_1, O_2, O_3, O_4, O_5 then share the characteristic that they are expressed entirely in terms of the observational vocabulary; for on our assumption, 'w' and 'v' are observational terms, and so are 'b' and 'l', which name certain observable bodies; finally, 'floats on' is an observational term because under suitable circumstances, direct observation will show whether a given observable object floats on a given observable liquid. On the other hand, the sentences (3.2), (3.3), (3.5), and (3.6) lack that characteristic, for they all contain the term 's', which, in our illustration, belongs to the theoretical vocabulary.

The systematic transition from the "observational data" listed in (3.4) to the prediction (3.7) of an observable phenomenon is schematized in the accompanying diagram. Here, an arrow represents a deductive inference; mention,



above an arrow, of a further sentence indicates that the deduction is effected by means of that sentence, i.e., that the conclusion stated at the right end follows logically from the premises listed at the left, taken in conjunction with the sentence mentioned above the arrow. Note that the argument just considered illustrates the schema (1.1), with O_1, O_2, O_3, O_4 constituting the statements of particular facts, the sentences (3.2) and (3.3) taking the place of the general laws, and O_5 that of E.[11]

Thus, the assumption of nonobservable entities serves the purposes of systematization: it provides connections among observables in the form of laws containing theoretical terms, and this detour via the domain of hypothetical entities offers certain advantages, some of which were indicated above.

In the case of our illustration, however, brief reflection will show that the advantages obtained by the "theoretical detour" could just as well have been obtained without ever resorting to the use of a theoretical term. Indeed, by virtue of the definition (3.2), the law (3.3) can be restated as follows:

(3.3)'. A solid body floats on a liquid if the quotient of its weight and its volume is less than the corresponding quotient for the liquid.

This alternative version clearly shares the advantages we found (3.3) to have over the crude generalization (3.1); and, of course, it permits the deductive transition from O_1, O_2, O_3, O_4 to O_5 just as well as does (3.3) in conjunction with (3.2).

The question arises therefore whether the systematization achieved by general principles containing theoretical terms can always be duplicated by means of general statements couched exclusively in observational terms. To prepare for an examination of this problem, we must first consider more closely the form and function of a scientific theory.

4. STRUCTURE AND INTERPRETATION OF A THEORY

Formally, a scientific theory may be considered as a set of sentences expressed in terms of a specific vocabulary. The vocabulary, V_T, of a theory T will be understood to consist of the extralogical terms of T, i.e., those which do not belong to the vocabulary of pure logic. Usually, some of the terms of V_T are defined by means of others; but, on pain of a circle or an infinite regress, not all of them can be so defined. Hence, V may be assumed to be divided into two subsets: *primitive terms*—those for which no definition is specified—and *defined terms*. Analogously, many of the sentences of a theory are derivable from others by means of the principles of deductive logic (and the definitions of the defined terms); but, on pain of a vicious circle or an infinite regress in the deduction, not all of the theoretical sentences can be thus established. Hence, the set of sentences asserted by T falls into two subsets: *primitive sentences*, or *postulates* (also called *axioms*), and *derivative sentences*, or *theorems*. Henceforth, we will assume that theories are stated in the form of axiomatized systems as here described; i.e., by listing, first the primitive and the derivative terms and the definitions for the latter, second, the postulates. In addition, the theory will always be thought of as formulated within a linguistic framework of a clearly specified logical structure, which determines, in particular, the rules of deductive inference.

The classical paradigms of deductive systems of this kind are the axiomatizations of various mathematical theories, such as Euclidean and various forms of non-Euclidean geometry, and the theory of groups and other branches of abstract algebra;[12] but by now, a number of theories in empirical science have likewise been put into axiomatic form, or approximations thereof; among them, parts of classical and relativistic mechanics,[13] certain segments of biological theory[14] and some theoretical systems in psychology, especially in the field of learning;[15] in economic theory, the concept of utility, among others, has received axiomatic treatment.[16]

11. Since (3.2) was presented as a definition, it might be considered inappropriate to include it among the general laws effecting the predictive transition from O_1, O_2, O_3, O_4, to O_5. And indeed, it is quite possible to construe the concept of logical deduction as applied to (1.1) in such a way that it includes the use of any definition as an additional premise. In this case, (3.3) is the only law invoked in the prediction here considered. On the other hand, it is also possible to treat sentences such as (3.2), which are usually classified as purely definitional, on a par with other statements of universal form, which are qualified as general laws. This view is favored by the consideration, for example, that when a theory conflicts with pertinent empirical data, it is sometimes the "laws" and sometimes the "definitions" that are modified in order to accommodate the evidence. Our analysis of deductive systematization is neutral with respect to this issue.

12. A lucid elementary discussion of the nature of axiomatized mathematical systems may be found in Cohen and Nagel (1934), Chapter VI; also reprinted in Feigl and Brodbeck (1953). For an analysis in a similar vein, with special emphasis on geometry, see also Hempel (1945). An excellent systematic account of the axiomatic method is given in Tarski (1941, Chapters VI-X); this presentation, which makes use of some concepts of elementary symbolic logic, as developed in earlier chapters, includes several simple illustrations from mathematics. A careful logical study of deductive systems in empirical science with special attention to the role of theoretical terms, is carried out in the first three chapters of Braithwaite (1953) and a logically more advanced exposition of the axiomatic method, coupled with applications to biological theory, has been given by Woodger, especially in (1937) and (1939).

13. See, for example, Hermes (1938); Walker (1943-1949), McKinsey, Sugar, and Suppes (1953); McKinsey and Suppes (1953), Rubin and Suppes (1953), and the further references given in these publications. An important pioneer work in the field is Reichenbach (1924).

14. See expecially Woodger (1937) and (1939).

15. See for example, Hull *et al.* (1940).

16. For example, in von Neumann and Morgenstern (1947), Chapter III and Appendix.

If the primitive terms and the postulates of an axiomatized system have been specified, then the proof of theorems, i.e., the derivation of further sentences from the primitive ones—can be carried out by means of the purely formal canons of deductive logic, and thus, without any reference to the meanings of the terms and sentences at hand; indeed, for the deductive development of an axiomatized system, no meanings need be assigned at all to its expressions, primitive or derived.

However, a deductive system can function as a theory in empirical science only if it has been given an *interpretation* by reference to empirical phenomena. We may think of such interpretation as being effected by the specification of a set of *interpretative sentences*, which connect certain terms of the theoretical vocabulary with observational terms.[17] The character of these sentences will be examined in detail in subsequent sections; at present it may be mentioned as an example that interpretative sentences might take the form of so-called operational definitions, i.e., of statements specifying the meanings of theoretical terms with the help of observational ones; of special importance among these are rules for the measurement of theoretical quantities by reference to observable responses of measuring instruments or other indicators.

The manner in which a theory establishes explanatory and predictive connections among statements couched in observational terms can now be illustrated in outline by the following example. Suppose that the Newtonian theory of mechanics is used to study the motions, under the exclusive influence of their mutual gravitational attraction, of two bodies, such as the components of a double-star system, or the moon and a rocket coasting freely 100 miles above the moon's surface. On the basis of appropriate observational data, each of the two bodies may be assigned a certain mass, and, at a given instant t_0, a certain position and velocity in some specified frame of reference. Thus, a first step is taken which leads, via interpretative sentences in the form of rules of measurement, from certain statements $O_1, O_2 \ldots O_k$ which describe observable instrument readings, to certain theoretical statements, say $H_1, H_2 \ldots H_6$, which assign to each of the two bodies a specific numerical value of the theoretical quantities mass, position, and velocity. From these statements, the law of gravitation, which is couched entirely in theoretical terms, leads to a further theoretical statement, H_7, which specifies the force of the gravitational attraction the two bodies exert upon each other at t_0; and H_7 in conjunction with the preceding theoretical statements and the laws of Newtonian mechanics implies, via a deductive argument involving the principles of the calculus, certain statements H_8, H_9, H_{10}, H_{11}, which give the positions and velocites of the two objects at a specified later time, say t_1. Finally, use in reverse of the interpretative sentences leads, from the last four theoretical statements, to a set of sentences $O'_1, O'_2 \ldots O'_m$, which describe observable phenomena, namely, instrument readings that are indicative of the predicted positions and velocities.

17. Statements effecting an empirical interpretation of theoretical terms have been discussed in the methodological literature under a variety of names. For example, Reichenbach, who quite early emphasized the importance of the idea with special reference to the relation between pure and physical geometry, speaks of *coordinative definitions* (1928, section 4; also 1951, Chapter VIII); Campbell [1920, Chapter VI; an excerpt from this chapter is reprinted in Feigl and Brodbeck (1953)] and Ramsey (1931, pp. 212–36) assume a *dictionary* connecting theoretical and empirical terms. (See also Section 8 below). Margenau (1950, especially Chapter 4) speaks of *rules of correspondence*, and Carnap (1956) has likewise used the general term '*correspondence rules*.' Northrop's *epistemic correlations* (1947, especially Chapter VII) may be viewed as a special kind of interpretative statements. For a discussion of interpretation as a semantical procedure, see Carnap (1939, sections 23, 24, 25), and Hutten (1956, especially Chapter II). A fuller discussion of interpretative statements is included in sections 6, 7, 8 of the present essay.

By means of a schema analogous to (3.8), the procedure may be represented as follows:

$$(4.1)\quad \{O_1, O_2 \ldots O_k\} \xrightarrow{R} \{H_1, H_2 \ldots H_6\} \xrightarrow{G} \{H_1, H_2 \ldots H_6, H_7\} \xrightarrow{LM} \{H_8, H_9, H_{10}, H_{11}\} \xrightarrow{R} \{O'_1, O'_2 \ldots O'_m\}$$

Here, R is the set of the rules of measurement for mass, position, and velocity; these rules constitute the interpretative sentences; G is Newton's law of gravitation, and LM are the Newtonian laws of motion.

In reference to psychology, similar schematic analyses of the function of theories or of hypotheses involving "intervening variables" have repeatedly been presented in the methodological literature.[18] Here, the observational data with which the procedure starts usually concern certain observable aspects of an initial state of a given subject, plus certain observable stimuli acting upon the latter; and the final observational statements describe a response made by the subject. The theoretical statements mediating the transition from the former to the latter refer to various hypothetical entities, such as drives, reserves, inhibitions, or whatever other not directly observable characteristics, qualities, or psychological states are postulated by the theory at hand.

5. THE THEORETICIAN'S DILEMMA

The preceding account of the function of theories raises anew the problem encountered in section 3, namely, whether the theoretical detour through a domain of not directly observable things, events, or characteristics cannot be

18. A lucid and concise presentation may be found, for example, in Bergmann and Spence (1941).

entirely avoided. Assume, for example, that—as will often be the case—the interpretative sentences as well as the laws asserted by the theory have the form of equations which connect certain expressions in terms of theoretical quantities either with other such expressions, or with expressions in terms of observable quantities. Then the problem can be stated in Hull's succinct formulation: "If you have a secure equational linkage extending from the antecedent observable conditions through to the consequent observable conditions, why, even though to do so might not be positively pernicious, use several equations where one would do?"[19] Skinner makes the same point in more general form when he criticizes the construction, in psychological theories, of causal chains in which a first link consisting of an observable and controllable event is connected with a final ("third") one of the same kind by an intermediate link which usually is not open to observation and control. Skinner argues: "Unless there is a weak spot in our causal chain so that the second link is not lawfully determined by the first, or the third by the second, then the first and third links must be lawfully related. If we must always go back beyond the second link for prediction and control, we may avoid many tiresome and exhausting digressions by examining the third link as a function of the first."[20]

Problem: if there is a translation of observational terms (materials of initial observation) into theoretical terms, and then back into observational terms (materials for prediction and control), why not to exclude theoretical terms altogether?

The conclusion suggested by these arguments might be called the *paradox of theorizing*. It asserts that if the terms and the general principles of a scientific theory serve their purpose, i.e., if they establish definite connections among observable phenomena, then they can be dispensed with since any chain of laws and interpretative statements establishing such a connection should then be replaceable by a law which directly links observational antecedents to observational consequents.

By adding to this crucial thesis two further statements which are obviously true, we obtain the premises for an argument in the classical form of a dilemma:

The rigorous formulation of the dilemma

(5.1) If the terms and principles of a theory serve their purpose they are unnecessary, as just pointed out; and if they do not serve their purpose they are surely unnecessary. But given any theory, its terms and principles either serve their purpose or they do not. Hence, the terms and principles of any theory are unnecessary.

This argument, whose conclusion accords well with the views of extreme methodological behaviorists in psychology, will be called the *theoretician's dilemma*.

However, before yielding to glee or to gloom over the outcome of this argument, it will be well to remember that the considerations adduced so far

19. Hull (1943, p. 284).
20. Skinner (1953, p. 35).

in support of the crucial first premise were formulated rather sketchily. In order to form a more careful judgment on the issue, it will therefore be necessary to inquire whether the sketch can be filled in so as to yield a cogent argument. To this task we now turn.

6. OPERATIONAL DEFINITIONS AND REDUCTION SENTENCES

It will be well to begin by considering more closely the character of interpretative sentences. In the simplest case, such a sentence could be an *explicit definition* of a theoretical expression in terms of observational ones, as illustrated by (3.2). In this case, the theoretical term is unnecessary in the strong sense that it can always be avoided in favor of an observational expression, its definiens. If all the primitives of a theory T are thus defined, then clearly T can be stated entirely in observational terms, and all its general principles will indeed be laws that directly connect observables with observables.

This would be true, in particular, of any theory that meets the standards of operationism in the narrow sense that each of its terms is introduced by an explicit definition stating an observable response whose occurrence is necessary and sufficient, under specified observable test conditions, for the applicability of the term in question. Suppose, for example, that the theoretical term is a one-place predicate, or property term, 'Q'. Then an operational definition of the kind just mentioned would take the form

(6.1) Def. $Qx \equiv (Cx \supset Ex)$

i.e., an object x has (by definition) the property Q if and only if it is such that if it is under test conditions of kind C then it exhibits an effect, or response, of kind E. Tolman's definition of expectancy of food provides an illustration: "When we assert that a rat expects food at L, what we assert is that *if* (1) he is deprived of food, (2) he has been trained on path P, (3) he is now put on path P, (4) path P is now blocked, and (5) there are other paths which lead away from path P, one of which points directly to location L, *then* he will run down the path which points directly to location L."[21] We can obtain this formulation by replacing, in (6.1), 'Qx' by 'rat x expects food at location L', 'Cx' by the conjunction of the conditions (1), (2), (3), (4), (5) for rat x, and 'Ex' by 'x runs down the path which points directly to location L'.

However, as has been shown by Carnap in a now classical argument,[22] this manner of defining scientific terms, no matter how natural it may seem, en-

21. Tolman, Ritchie, and Kalish (1946, p. 15). See the detailed critical analysis of Tolman's characterization of expectancy in MacCorquodale and Meehl (1945, pp. 179–81).
22. See Carnap (1936–37), section 4.

counters a serious difficulty. For on the standard extensional interpretation, a conditional sentence, such as the definiens in (6.1), is false only if its antecedent is true and its consequent false. Hence, for any object which does not satisfy the test conditions *C*, and for which therefore the antecedent of the definiens is false, the definiens as a whole is true; consequently, such an object will be assigned the property Q. In terms of our illustration: of any rat not exposed to the conditions (1)-(5) just stated, we would have to say that he expected food at *L*—no matter what kind of behavior the rat might exhibit.

One way out of this difficulty is suggested by the following consideration. In saying that a given rat expects food at *L*, we intend to attribute to the animal a state or a disposition which, under circumstances (1)-(5), will cause the rat to run down the path pointing directly to *L*; hence, in a proper operational definition, *E* must be tied to *C* nomologically, i.e., by virtue of general laws of the kind expressing causal connections. The extensional 'if . . . then . . .'—which requires neither logical nor nomological necessity of connection—would therefore have to be replaced in (6.1) by a stricter, nomological counterpart which might be worded perhaps as 'if . . . then, with causal necessity, . . .'. However, the idea of causal or of nomological necessity here invoked is not clear enough at present to make this approach seem promising.[23]

Carnap[24] has proposed an alternative way of meeting the difficulty encountered by definitions of the form (6.1); it consists in providing a partial rather than a complete specification of meaning for 'Q'. This is done by means of so-called reduction sentences; in the simplest case, (6.1) would be replaced by the following *bilateral reduction sentence:*

(6.2) $Cx \supset (Qx \equiv Ex)$

This sentence specifies that if an object is under test conditions of kind *C*, then it has the property Q just in case it exhibits a response of kind *E*. Here, the use of extensional connectives no longer has the undesirable aspects it exhibited in (6.1). If an object is not under test conditions *C*, then the entire formula (6.2) is true of it, but this implies nothing as to whether the object does, or does not, have the property Q. On the other hand, while (6.1) offers a full explicit definition of 'Q', (6.2) specifies the meaning of 'Q' only partly, namely, for just those objects that meet condition *C*; for those which do not, the meaning of 'Q' is left unspecified.

23. On this point, and on the general problem of explicating the concept of a law of nature, see Braithwaite (1953), Chapter IX; Burks (1951); Carnap (1956), section 9; Goodman (1955); Hempel and Oppenheim (1948), Part III; Reichenbach (1954).

24. In his theory of reduction sentences, developed in Carnap (1936-37). There is a question, however, whether certain conditions which Carnap imposes upon reduction sentences do not implicitly invoke causal modalities. On this point, see Hempel (1963), section 3.

In our illustration, (6.2) would specify the meaning of '*x* expects food at *L*' only for rats that meet conditions (1)-(5); for them, running down the path which points to *L* would be a necessary and sufficient condition of food expectancy. In reference to rats that do not meet the test conditions (1)-(5), the meaning of '*x* expects food at *L*' would be left open; it could be further specified by means of additional reduction sentences.

In fact, it is this interpretation which is indicated for Tolman's concept of food expectancy. For while the passage quoted above seems to have exactly the form (6.1), this construal is ruled out by the following sentence which immediately follows the one quoted earlier: "When we assert that he does not expect food at location *L*, what we assert is that, under the same conditions, he will not run down the path which points directly to location *L*." The total interpretation thus given to 'rat *x* expects food at *L*' is most satisfactorily formulated in terms of a sentence of the form (6.2), in the manner outlined in the preceding paragraph.[25]

As this example illustrates, reduction sentences offer a precise formulation of the intent of operational definitions. By expressing the latter as merely partial specifications of meaning, they treat theoretical concepts as "open"; and the provision for a set of different, and mutually supplementary, reduction sentences for a given term reflects the availability, for most theoretical terms, of different operational criteria of application, pertaining to different contexts.[26]

However, while an analysis in terms of reduction sentences construes theoretical terms as not fully defined by reference to observables, it does not prove that a full explicit definition in observational terms *cannot* be achieved for theoretical expressions. And indeed, it seems questionable whether a *proof* to this effect could even be significantly asked for. The next section deals with this issue in some detail.

7. ON THE DEFINABILITY OF THEORETICAL TERMS BY MEANS OF AN OBSERVATIONAL VOCABULARY

The first, quite general, point to be made here is this: a definition of any term, say 'v', by means of a set V of other terms, say 'v_1', 'v_2' . . . 'v_n', has to specify a necessary and sufficient condition for the applicability of 'v', expressed in terms of some or all of the members of V. And in order to be able to judge whether this can be done in a given case, we will have to know how the terms

25. And in fact, the total specification of meaning effected by the passages quoted is then summarized by the authors in their "definition" DF II, which has exactly the form (6.2) of a bilateral reduction sentence for 'rat x expects food at L'. [Tolman, Ritchie, and Kalish (1946, p. 15.)]

26. For a fuller discussion, see Carnap (1936-37), section 7 and (1956), section 10.

under consideration are to be understood. For example, the vocabulary consisting of the terms 'male' and 'offspring of' permits the formulation of a necessary and sufficient condition of application for the term 'son of' in its biological, but not in its legal sense. How the given terms are to be understood can be indicated by specifying a set U of sentences which are to be considered as true, and which connect the given terms with each other and perhaps with other terms. Thus, U will be a set of sentences containing 'v', 'v_1' . . . 'v_n' and possibly also other extralogical constants. For example, in the case of the biological use of the terms 'son', 'male', and 'offspring', in reference to humans, the following set of sentences —let us call it U_1—might be given: 'Every son is male,' 'No daughter is male,' 'x is an offspring of y if and only if x is a son or a daughter of y'.

Generally, the sentences of U specify just what assumptions are to be made, in the search for a definition, concerning the concepts under consideration; and the problem of definability now turns into the question whether it is possible to formulate, in terms of $v_1, v_2 \ldots v_n$, a condition which, *in virtue of the assumptions included in U*, will be both necessary and sufficient for v. Thus, using an idea set forth and developed technically by Tarski,[27] we see that the concept of definability of 'v' by means of 'v'$_1$, 'v_2' . . . 'v_n' acquires a precise meaning only if it is explicitly relativized by reference to a set U of specifying assumptions. That precise meaning may now be stated as follows:

(7.1) 'v' is definable by means of the vocabulary $V = \{'v_1', 'v_2', \ldots, 'v_n'\}$ relative to a finite set U of statements containing, at least, 'v' and all the elements of V if from U there is deducible at least one sentence stating a necessary and sufficient condition for v in terms of no other extralogical constants than the members of V.

If all the terms under study are one-place predicates of the first order, for example, then a sentence of the required kind could most simply be stated in the form

(7.2) $v(x) \equiv D(x, v_1, v_2, \ldots, v_n)$

where the expression on the right-hand side stands for a sentential function whose only free variable is 'x', and which contains no extralogical constant other than those included in V.

Similarly, in the case of our illustration, the set U_1 specified above implies the statement:

$$x \text{ is a son of } y \equiv (x \text{ is male and } x \text{ is an offspring of } y)$$

so that, relative to U_1, 'son' is definable as 'male offspring'.

27. See Tarski (1935), especially pp. 80–83.

A definition that is not simply a convention introducing an abbreviatory notation (such as the convention to let 'x^5' be short for '$x \cdot x \cdot x \cdot x \cdot x$') is usually considered as stating the *synonymy* of two expressions, or, as it is often put, the *identity of their meanings*. Now the question of the definability of a given term 'v' by means of a set V of other terms surely is not simply one of notational fiat; and indeed it will normally be construed as concerning the possibility of expressing the meaning of the term 'v' by reference to the meanings of the members of V. If this conception is adopted, then naturally the information needed to answer the question of definability will concern the *meanings* of 'v' and of the members of V; accordingly, the statements in U which provide this information will then be required not simply to be true, but to be analytic, i.e., true by virtue of the intended meanings of the constituent terms. In this case, the statements in U would have the character of meaning postulates in the sense of Kemeny and Carnap.[28]

But in a study of the definability of theoretical expressions by means of observational terms, it is neither necessary nor even advisable to construe definition in this intensional manner. For, first of all, the idea of meaning, and related notions such as those of analyticity and synonymy, are by no means as clear as they have long been considered to be,[29] and it will be better, therefore, to avoid them when this is possible.

Secondly, even if those concepts are accepted as clearly intelligible, the definability of a theoretical term still cannot be construed exclusively as the existence of a synonymous expression containing only observational terms: it would be quite sufficient if a coextensive (rather than a strictly cointensive, or synonymous) expression in terms of observables were forthcoming. For such an expression would represent an empirically necessary and sufficient observational condition of applicability for the theoretical term; and this is all that is required for our purposes. In fact, the sentence stating the coextensiveness in question, which might have the form (7.2) for example, can then be given the status of a truth-by-definition, by a suitable reformalization of the theory at hand.

It is of interest to note here that a necessary and sufficient observational condition for a theoretical term, say 'Q', might be inductively discovered even if only a partial specification of the meaning of 'Q' in terms of observables were

28 See Kemeny (1951) and (1952); Carnap (1952).

29. On this point, see especially Quine (1951); Goodman (1949); White (1950) and (1956, Part II). The significance of the notion of analyticity in special reference to theoretical statements is critically examined, for example, in Pap (1953) and (1955) and in Hempel (1963). Arguments in defense of concepts such as analyticity and synonymy are advanced in the following articles, among others: Carnap (1952), (1955); Grice and Strawson (1956); Martin (1952); Mates (1951); Wang (1955).

available. Suppose, for example, that a set of alternative conditions of application for 'Q' has been specified by means of bilateral reduction sentences:

$$(7.3)\quad \begin{array}{l} C_1x \supset (Qx \equiv E_1x) \\ C_2x \supset (Qx \equiv E_2x) \\ \cdots\cdots\cdots\cdots \\ C_nx \supset (Qx \equiv E_nx) \end{array}$$

where all predicates except 'Q' are observational. Suppose further that suitable investigations lead to the following empirical generalizations:

$$(7.4)\quad \begin{array}{l} C_1x \supset (Ox \equiv E_1x) \\ C_2x \supset (Ox \equiv E_2x) \\ \cdots\cdots\cdots\cdots \\ C_nx \supset (Ox \equiv E_nx) \end{array}$$

where 'Ox' stands for a sentential function in 'x' which contains no nonobservational extralogical terms. These findings, in combination with (7.3), would inductively support the hypothesis

$$(7.5)\quad Qx \equiv Ox$$

which presents a necessary and sufficient observational condition for Q. However, (7.5) even if true (its acceptance involves the usual "inductive risk") clearly does not express a synonymy; if it did, no empirical investigations would be needed in the first place to establish it. Rather, it states that, as a matter of empirical fact, 'O' is coextensive with 'Q', or, that O is an empirically necessary and sufficient condition for Q. And if we wish, we may then imagine the theory-plus-interpretation at hand to be thrown into the form of a deductive system in which (7.5) becomes a definitional truth, and (7.3) assumes the character of a set of empirical statements equivalent to those listed in (7.4).

It might be mentioned here in passing that a similarly broad extensional interpretation of definability is called for also in the context of the problem whether a given scientific discipline, such as psychology, can be "reduced" to another, such as biology or even physics and chemistry.[30] For one component of this problem is the question whether the terms of the first discipline can be defined by means of those of the latter; and what is wanted for this purpose is again a set of empirical hypotheses providing for each psychological term a necessary and sufficient condition of application expressed in the vocabulary of biology, or of physics and chemistry.

30. On the problem of "reducing" the concepts of one discipline to those of another, the following publications have important bearings: Nagel (1949) and (1951); Woodger (1952, pp. 271ff); Kemeny and Oppenheim (1956).

When we say, for example, that the concepts of the various chemical elements are definable in physical terms by a characterization of the specific ways in which their molecules are composed of elementary physical particles, we are clearly referring to results of experimental research rather than of a mere analysis of what is *meant* by the terms naming the various elements. If the latter were the case, it would be quite incomprehensible why the problems pertaining to the definability of scientific terms should present any difficulty, and why they should be the objects of much conjecture and controversy.

The preceding considerations have important implications for our question whether all theoretical terms in empirical science can be defined in terms of observables. First of all, they show that the question as stated is elliptical: to complete it, we have to specify some set U of statements as referred to in (7.1). What set could reasonably be chosen for this purpose? One natural choice would be the set of all statements, in theoretical or observational terms, that are accepted as presumably true by contemporary science. Now, this pragmatic-historical characterization is by no means precise and unambiguous; there is a wide border area containing statements for which it cannot be clearly determined whether they are accepted by contemporary science. But no matter how the claims of these border-area statements are adjudicated, and no matter where—within reason—the borderline between observational and theoretical terms is drawn, it is at least an open question whether the set of presently accepted scientific statements implies for every theoretical term a necessary and sufficient condition of applicability in terms of observables. Certainly those who have asserted such definability have not supported their claim by actually deducing such conditions, or by presenting cogent general reasons for the possibility of doing so.

There is another way in which the claim of definability may be construed, namely as the assertion that as our scientific knowledge becomes more comprehensive, it will eventually be possible to deduce from it necessary and sufficient conditions of the required kind. (This is the sense in which definability is usually understood by those who claim the eventual definability of the concepts of psychology in terms of those of biology or of physics and chemistry; for that all the requisite definition statements—even in an extensional, empirical sense—cannot be deduced from current psychological, biological, physical, and chemical principles seems clear.[31]) But to assert definability of a theoretical term in this sense is to make a twofold claim: first, that the term in question will not

31. This point is discussed more fully in Hempel (1951).

be abandoned in the further development of scientific theorizing; and second, that general laws will be discovered which establish certain necessary and sufficient conditions, expressible in observational terms, for the applicability of the theoretical term at hand. Clearly, the truth of these claims cannot be established by philosophic arguments, but at best by the results of further scientific research.

Despite the precariousness of the problem, various claims and counterclaims have been advanced by philosophers of science and by methodologically interested scientists concerning the possibility of defining theoretical terms by reference to observables.

Some among the philosophers have simply urged that nothing short of explicit definition in terms of a vocabulary that is clearly understood can provide an acceptable method of introducing new terms into the language of science; and the argument supporting this view is to the effect that otherwise the new terms are not intelligible.[32] To this question we will return later. The protagonists of this view do not make an assertion, then, about the actual definability of the theoretical terms used in contemporary empirical science; rather, they stress the importance of clarifying the ideas of science by restating them, as far as possible, in a language with a clear and simple logical structure, and in such a way as to introduce all theoretical terms by means of definitions.

Other writers have argued, in effect, that scientific theories and the way in which they function have certain pervasive logical or methodological characteristics which are not affected by changes in scientific knowledge, and by reference to which the question as to the definability of theoretical terms can be settled without examining all the statements accepted by contemporary science or waiting for the results of further research.

An example of this type of procedure is provided by Carnap's argument, referred to in the beginning of section 6 above, which shows that definitions of the form (6.1) cannot serve to introduce scientific concepts of the kind they are meant to specify. The argument is limited, however, in the sense that it does not show (and does not claim to show) that an explicit definition of theoretical terms by means of observational ones is generally impossible.

More recently,[33] Carnap has extended his examination of the problem in the following direction. Suppose that a given object, *b*, exhibits this kind of

32. One writer who is impelled by his "philosophical conscience" to take this view is Goodman (see 1951, Chapter I; 1955, Chapter II, section 1). A similar position was taken by Russell when he insisted that physical objects should be conceived as "logical constructions" out of sense-data, and thus as definable in terms of the latter (see, for example, 1929, Chapter VIII).

33. See Carnap (1956), especially sections 9, 10.

lawful behavior: whenever *b* is under conditions of a certain observable kind *C*, then it shows a response of a specified observable kind *E*. We then say that *b* has the disposition to react to *C* by *E*; let us call this dispositional property Q for short. Clearly, our earlier discussion in section 6 concerns the problem of precisely defining '*Q*' in terms of '*C*' and '*E*'; we noted there, following Carnap, that we will either have to resign ourselves to a partial specification of meaning for '*Q*' by means of the bilateral reduction sentence (6.2); or, if we insist on an explicit complete definition, we will have to use nomological modalities in the definiens.

But no matter which of these alternative courses is chosen, the resulting disposition term '*Q*' has this characteristic: if a given object *b* is under condition *C* and fails to show response *E*, or briefly, if *Cb* but $\sim Eb$, then this establishes conclusively that *b* lacks the property Q, or briefly that $\sim Qb$. This characteristic, Carnap argues, distinguishes "pure disposition terms," such as '*Q*', from the theoretical terms used in science; for though the latter are connected with the observational vocabulary by certain interpretative sentences—Carnap calls them *C*-rules—those rules will not, in general, permit a set of observational data (such as '*Cb*' and '$\sim Eb$' above) to constitute conclusive evidence for or against the applicability of the theoretical term in a given situation. There are two reasons for this assertion. First, the interpretative sentences for a given theoretical term provide an observational interpretation only within a certain limited range; thus, for example, in the case of the theoretical term 'mass', no *C*-rule is directly applicable to a sentence S_m ascribing a certain value of mass to a given body, if the value is either so small that the body is not directly observable or so large that the observer cannot "manipulate the body."[34]

Secondly, a direct observational interpretation for a theoretical term always involves the tacit understanding that the occurrence or absence of the requisite observable response in the specified test situation is to serve as a criterion only if there are no disturbing factors, or, provided that "the environment is in a normal state."[35] Thus, for example, a rule of correspondence might specify the deflection of a magnetic needle as an observable symptom of an electric current in a nearby wire, but with the tacit understanding that the response of the needle is to count only if there are no disturbing factors, such as, say, a sudden magnetic storm.

Generally, then, Carnap holds that "if a scientist has decided to use a certain term '*M*' in such a way, that for certain sentences about M, any possible observational results can never be absolutely conclusive evidence but at best

34. Carnap (1956), section 10.

35. Carnap (1950), section 10.

evidence yielding a high probability," then the appropriate place for 'M' is in the theoretical vocabulary.[36]

Now we should note, first of all, that if Carnap's arguments are sound, they establish that the theoretical terms of science cannot be construed as pure disposition terms, and thus even if, by the use of nomological modalities, explicit definitions of the latter should be achieved, this method would be unavailing for theoretical terms. But the arguments do not show—and are not claimed to show—that theoretical terms can in no way be explicitly defined in terms of observables. In fact, if Carnap's statement quoted in the preceding paragraph is accepted, then many terms that can be explicitly defined by means of the observational vocabulary must be qualified as theoretical. For example, let '*R*' be a two-place observational predicate, and let a one-place predicate 'M_1' be defined as follows:

(7.6) Def. $M_1x \equiv (\exists y)\, Rxy$

i.e., an object *x* has the property M_1 just in case it stands in relation *R* to at least one object *y*. If, for example, '*Rxy*' stands for '*x* is less heavy than *y*', then M_1 is the property of being exceeded in weight by at least one object, or, of not being the heaviest of all objects.

Let us assume, as customary, that the domain of objects under study is infinite or at least has not been assigned any definite maximum number of elements. Consider now the possibility of conclusive observational evidence for or against the sentence 'M_1a', which attributes M_1 to a certain object *a*. Obviously, a single observational finding, to the effect that *a* bears *R* to a certain object *b*, or that *Rab*, would suffice to verify 'M_1a' completely. But no finite set of observational data—'$\sim Raa$', '$\sim Rab$', '$\sim Rac$', and so forth—would suffice for a conclusive refutation of 'M_1a'. According to Carnap's criterion, therefore, 'M_1', though defined in terms of the observational predicate '*R*', might have to be classified as a theoretical term.

36. Carnap (1956), section 10. An idea which is similar in spirit, but not quite as clear in its content, has been put forward by Pap in (1953) and in (1955), sections 10-13 and 70, with the claim (not made by Carnap for his argument) that it establishes the "untenability" of the "thesis of explicit definability" of theoretical terms by means of observational ones. (Pap 1953, p. 8). On the other hand, Bergmann holds that many concepts of theoretical physics, including "even the particle notions of classical physics could, in principle, be introduced by explicit definitions. This, by the way, is also true of all the concepts of scientific psychology." (1951a, section 1. In the same context Bergmann mentions that the method of partial interpretation seems to be necessary in order to dissolve some of the puzzles concerning quantum theory). However, this strong assertion is supported chiefly by sketches of some sample definitions. Bergmann suggests, for example, that 'This place is in an electric field' can be defined by a sentence of the form 'If R_1 then R_2' where R_1 stands for a sentence to the effect that there is an electroscope at the place in question, and R_2 stands "for the description of the behavior of the electroscope (in an electric field)." (1951, pp. 98-99.) However, this kind of definition may be questioned on the basis of Carnap's arguments, which have just been considered. And in addition, even if unobjectionable, some examples cannot establish the general thesis at issue. Thus, the question remains unsettled.

But possibly, in the passage quoted above, Carnap meant to require of a theoretical term '*M*' that for certain sentences about *M* no observational results can be conclusively verificatory or falsificatory evidence. Yet even terms meeting this requirement can be explicitly defined in terms of observables. Let '*S*' be a three-place observational predicate; for example, '*Sxyz*' might stand for '*x* is farther away from *y* than from *z*.' And let 'M_2' be defined as follows:

(7.7) Def. $M_2x \equiv (\exists y)(z)[\sim(z = y) \supset Sxyz]$.

In our example, an object *x* has M_2 just in case there is an object *y* from which it is farther away than from any other object *z*. Consider now the sentence 'M_2a'. As is readily seen, no finite set of observational findings (all the relevant ones would have the form '*Sabc*' or '$\sim Sabc$') can be conclusive evidence, either verificatory or falsificatory, concerning 'M_2a'. Hence, though explicitly defined in terms of the observational predicate '*S*', the term 'M_2' is theoretical according to the criterion suggested by Carnap.

The preceding discussion illustrates an elementary but important point: when a term, say a one-place predicate '*Q*', is defined in terms of observables, its definiens must state a necessary and sufficient condition for the applicability of '*Q*', i.e., for the truth of sentences of the form '*Qb*'. But even though that condition is then stated completely in observational terms, it still may not enable us to decide, on the basis of a finite number of observational findings, whether '*Q*' applies to a given object *b*; for the truth condition for '*Qb*' as characterized by the definiens may not be equivalent to a truth functional compound of sentences each of which expresses a potential observational finding.

To add one more example to those given before: suppose that the property term 'iron object' and the relation terms 'attracts' and 'in the vicinity of' are included in the observational vocabulary. Then the definition

(7.8) Def. *x* is a magnet $\equiv$ *x* attracts every iron object in its vicinity

is in terms of observables; but the criterion it provides for an object *b* being a magnet cannot be expressed in terms of any finite number of observational findings; for to establish that *b* is a magnet, we would have to show that any piece of iron which, at any time whatever, is brought into the vicinity of *b*, will be attracted by *b*; and this is an assertion about an infinity of cases.

To express the idea more formally, let us assume that our observational vocabulary contains, in addition to individual names for observable objects, just first-order predicates of any degree, representing attributes (i.e., properties or relations) which are observable in the sense that a small number of direct observations will suffice, under suitable conditions, to ascertain whether a given object or group of objects exhibits the attribute in question.

Now let us adopt the following definitions: An *atomic sentence* is a sentence, such as '*Pa*', '*Rcd*', '*Sadg*', which ascribes an observable attribute to a specified object or group of objects. A *basic sentence* is an atomic sentence or the negation of an atomic sentence. A *molecular sentence* is a sentence formed from a finite number of atomic sentences by means of truth-functional connectives. Basic sentences will be considered as included among the molecular sentences.

Basic sentences can be considered as the simplest statements describing potential results of direct observation: they assert that some specified set of (one or more) objects has, or lacks, such and such an observable attribute.

Now for every molecular statement *S*, there exist certain finite classes of basic statements which imply *S*, and certain other such classes which imply the negation of *S*. Thus, the molecular sentence '*Pa* v (∼*Pa*·*Rab*)' is implied by {'*Pa*'} and also by {'∼*Pa*', '*Rab*'}, for example; whereas its negation is implied by the set {'∼*Pa*', '∼*Rab*'}. Hence, for each molecular sentence *S*, it is possible to specify a set of basic sentences whose truth would conclusively verify *S*, and also a set of basic sentences whose truth would verify the negation of *S*, and would thus conclusively refute *S*. Thus, a molecular sentence is capable both of conclusive observational verification and of conclusive observational falsification "in principle," i.e., in the sence that potential data can be described whose occurrence would verify the sentence, and others whose occurrence would falsify it; but not of course in the sense that the two kinds of data might occur jointly—indeed, they are incompatible with each other.

There are even some sentences of nonmolecular form, i.e., sentences containing quantifiers nonvacuously, which are both completely verifiable and completely falsifiable in the sense just specified.[37] For example, the sentence '(*x*) (*Px* v *Qa*)' is implied by {'*Qa*'} and its negation by {'∼*Pb*', '∼*Qa*'}. A similar argument applies to the sentence '(∃*x*) (*Px*·*Qc*)'.

As a rule, however, nonmolecular sentences are not both verifiable and falsifiable. This holds, in particular, for all nonmolecular sentences of purely general form, i.e., those containing no individual constants at all, such as '(*x*) (*Px* ⊃ *Qx*)'; but it is true also of many quantified sentences containing individual constants. Thus, if '*R*' and '*S*' are observational predicates, then sentences of the type '(∃*y*)*Ray*' are not falsifiable and sentences of the types '(*y*) (∃*z*)*Sayz*' and '(∃*y*)(*z*)*Sayz*' are neither verifiable nor falsifiable, as is readily seen.

Explicit definitions of scientific terms by means of an observational vocabulary may accordingly be divided into two kinds: those which provide *finite observational criteria of application* for the defined term, and those which do not. The

37. (added in 1964). The present paragraph, and the next few, have been modified so as to correct a mistaken statement made here in the original version of this essay, namely, that only molecular sentences are both verifiable and falsifiable.

former are simply those whose definiens, when applied to a particular case, yields a sentence that is both verifiable and falsifiable. The following definition is of this kind:

(7.9) Def. Son $xy \equiv$ Male $x \cdot$ Offspring xy

For application of the definiens to two particular individuals, say *a* and *b*, yields the sentence 'Male *a* . Offspring *a b*', which is both verifiable and falsifiable and thus provides a finite observational criterion for the application of the term 'Son' to *a* in relation to *b*. On the other hand, the definitions (7.6), (7.7), and (7.8) above are among those which afford no finite observational criteria of application for the terms they define; this was pointed out earlier.

However, the circumstance that a term, say '*M*', is originally introduced by a definition affording no finite observational criteria for its application does not preclude the possibility that '*M*' may in fact be coextensive with some observational predicate, or with a truth-functional compound of such predicates, say 'O_m'; and if this should be found to be the case, then '*M*' could, of course, be redefined by 'O_m' and could thus be provided with a finite observational criterion of application.

But granting certain plausible assumptions concerning the observational vocabulary, it can be proved that not all scientific terms are definable in a way that provides them with finite criteria of application. We will assume that the observational vocabulary is finite. It may contain individual names designating certain observable objects; first-order predicate terms with any finite number of places, representing properties and relations of observable objects; and also functors, i.e., terms expressing quantitative aspects—such as weight in grams, volume in cubic centimeters, or age in days—of observable objects. However, we will suppose that each of the functors can take on only a finite number of different values; this corresponds to the assumption that only a finite number of different weights, for example, can be ascertained and distinguished by direct observation.

In contrast to the functors in the observational vocabulary, the theoretical vocabulary of physics, for example, contains a large number of functors whose permissible values range over all real numbers or over all real numbers within a certain interval. Thus, for example, the distance between two points may theoretically have any non-negative value whatever. Now a definition of the required kind for a theoretical functor would have to specify, for each of its permissible values, a finite observational criterion of application. Thus, in the case of the theoretical functor 'length', a necessary and sufficient condition, in the form of a finite observational criterion, would have to be forthcoming for each of the infinitely many statements of the form 'The distance, in centimeters, between points *x* and *y* is *r*' or briefly, '$l(x,y) = r$', where *r* is some real number.

Hence we would have to specify for each value of 'r' a corresponding finitely ascertainable configuration of observables. But this is impossible because the limits of discrimination in direct observation allow only a finite, though very large, number of finitely observable configurations to be ascertained and distinguished.

However, if we do not require a finite observational criterion of application for each permissible value of a theoretical functor, then an infinity of different values may become available.[38] Consider, for example, the functor 'the number of cells contained in organism y'. If 'x is a cell', 'y is an organism', and 'x is contained in y' are admitted as observational expressions, then it is possible to give a separate criterion of applicability, in terms of observables, for each of the infinitely many values 1, 2, 3 . . . which that functor may theoretically assume.[39] This can be done by means of the Frege-Russell analysis of cardinal numbers. For $n = 1$, for example, the necessary and sufficient condition is the following:

(7.10) $(\exists u)\,(v)\,[y \text{ is an organism} \cdot ((v \text{ is a cell} \cdot v \text{ is contained in } y) \equiv (v = u))]$

Thus, the reach of explicit definition in terms of observables, even in the first-order functional calculus, is greatly extended if quantification is permitted in the definiens. And if stronger logical means are countenanced, considerable further extensions may be obtained. For example, the functor 'the number of cells contained in y' can be explicitly defined by the single expression

(7.11) $\hat{\alpha}\,(\alpha \text{ sim } \hat{x}\,(x \text{ is a cell} \cdot x \text{ is contained in } y))$

Here, the circumflex accent is the symbol of class abstraction, and 'sim' the symbol for similarity of classes (in the sense of one-to-one matchability of their elements).

So far, we have examined only functors whose values are integers. Can functors with rational and even irrational values be similarly defined in terms of observables? Consider, for example, the theoretical functor 'length in centimeters'. Is it possible to express, in observational terms, a necessary and sufficient condition for

(7.12) $l(x,y) = r$

for every non-negative value of r? We might try to develop a suitable definition which would correspond to the fundamental method of measuring length

38. I am grateful to Herbert Bohnert who, in a conversation, provided the stimulus for the development of the ideas here outlined concerning the definability of functors with infinitely many permissible values. Dr. Bohnert remarked on that occasion that explicit definition of such functors in terms of an observational vocabulary should be possible along lines indicated by the Frege-Russell theory of natural and of real numbers.

39. If it should be objected that 'cell' and 'organism' are theoretical rather than observational terms, then they may be replaced, without affecting the crux of the argument, by terms whose observational character is less controversial, such as 'marble' and 'bag', for example.

by means of rigid rods. And indeed, if our observational vocabulary contains a name for the standard meter bar, and furthermore the (purely qualitative) terms required to describe the fundamental measuring procedure, it is possible to state, for any specified rational or irrational value of r, a necessary and sufficient condition for (7.12). However, the definiens will normally be teeming with symbols of quantification over individuals and over classes and relations of various types and will be far from providing finite observational criteria of application. I will briefly indicate how such definitions may be obtained. Expressions assumed to belong to the observational vocabulary will be italicized.

First, *the segment determined by two points x,y* will be said to have a length of 100 centimeters if it is *congruent* with (i.e., can be made to coincide with) *the segment marked off on the standard meter bar*. Next, consider the observational criterion for a rational value of length, say, $l(x,y) = .25$. It may be stated as follows: there are four *segments*, each *marked off on a rigid body*, such that (i) all four are *congruent* with each other; (ii) their *sum* (i.e., the segment obtained by placing them end to end along a straight line) is *congruent* with *the segment marked off on the standard meter bar;* (iii) each of the four *segments* is *congruent* with the *segment determined by points x,y*. Analogously, an explicit observational definiens can be formulated for any other value of n that is a rational multiple of 100, and hence, for any rational value of n.

Next, the consideration that an irrational number can be construed as the limit of a sequence of rational numbers yields the following necessary and sufficient condition for $l(x,y) = r$, where r is irrational: *the segment determined by the points x,y* contains an infinite sequence of *points* $x_1, x_2, x_3 \ldots$ such that (i) x_1 is *between* x and y, x_2 *between* x_1 and y, and so forth; (ii) given any *segment* S of rational length, there is a *point*, say x_n, in the sequence such that the *segments determined by* x_n and y, x_{n+1} and y, and so forth are all *shorter than* S, (iii) the lengths of the *segments determined by* x and x_1, x and x_2, and so forth, form a sequence of rational numbers with the limit r.

Finally, the idea underlying the preceding definition can be used to formulate an explicit definiens for the expression '$l(x,y)$' in such a way that its range of values is the set of all non-negative numbers.

Definitions of the kind here outlined are attainable only at the cost of using a strong logical apparatus, namely, a logic of sets adequate for the development of the theory of real numbers.[40] This price will be considered too high by nomin-

40. The argument can readily be extended to functors taking complex numbers or vectors of any number of components as values. Our reasoning has relied essentially on the Frege-Russell method of defining the various kinds of numbers (integers, rational, irrational, complex numbers, etc.) in terms of the concepts of the logic of sets. For a detailed outline of the procedure, see Russell (1919); fuller technical accounts may be found in works on symbolic logic.

alists, who hold that many of the logical concepts and principles here required, beginning with the general concept of set, are intrinsically obscure and should not, therefore, be used in a purported explication of the meanings of scientific terms. This is not the place to discuss the nominalistic strictures, however, and besides, it would no doubt be generally considered a worthwhile advance in clarification if for a set of theoretical scientific expressions explicit definitions in terms of observables can be constructed at all.

Another objection that might be raised against the definitional procedure here outlined is that it takes a schematic and oversimplified view of the fundamental measurement of length, and that it is rather liberal in construing as observational certain terms needed in the definiens, such as 'rigid body' and 'point'. This is quite true. By including the term 'point' in the observational vocabulary, for example, we construed points as directly observable physical objects; but our observational criterion for two points x,y determining a segment of irrational length required that there should be an infinite sequence of other points between x and y. This condition is never satisfied by the observable "points" in the form of small physical objects, or marks on rigid bodies, which are used in the fundamental measurement of length. As a consequence, the actual performance of fundamental measurement as represented in the above definition will never yield an irrational value for the length of a segment. But this does not show that no meaning has been assigned to irrational lengths; on the contrary, our outline of the definition shows that a meaning can indeed be formulated in observational terms for the assignment of any specified irrational value to the length of a physical line segment, as well as for the function 'length in centimeters' in general.

However, the concept of length thus defined is not adequate for a physical theory which incorporates geometry, say in its Euclidean form. For the latter requires that the length of certain segments which are well accessible to direct measurement—such as the diagonal of a square whose sides have a length of 100 centimeters—be an irrational number; and statements to this effect will always turn out to be false if the criterion just discussed is made strictly definitory of length; for that procedure, as we noted, will always yield a rational value for the length of a given segment.

What the preceding argument about quantitative terms (represented by functors) shows, then, is this: the fact that the set of permissible values of a theoretical functor is infinite need not preclude an explicit definition for it by means of a finite vocabulary containing only qualitative terms which are, by reasonably liberal standards, observational in character. The argument does not show, however, that such a definition is available for every functor term required by science (even our illustrative definition of 'length' turned out not to meet the needs of theoretical physics); and indeed, as was pointed out early in this section, a general proof to this effect cannot be expected.

Some writers have taken the position that even if in principle theoretical terms could be avoided in favor of observational ones, it would be practically impossible or—what is more serious—methodologically disadvantageous or even stultifying to do so.

There is, for example, the answer given by Tolman and by Spence to the problem considered by Hull, which was mentioned in section 5 above: if intervening theoretical variables can establish a secure linkage between antecedent and consequent observable conditions, why should we not use just one functional connection that directly links antecedents and consequents? Spence adduces as one reason, also suggested by Tolman,[41] the following consideration: the mathematical function required to express the connection will be so complex that it is humanly impossible to conceive of it all at once; we can arrive at it only by breaking it down into a sequence of simpler functional connections, mediated by intervening variables. This argument, then, attributes to the introduction of unobservable theoretical entities an important practical role in the context of discovering interdependencies among observables, and presumably also in the context of actually performing the calculations required for the explanation or prediction of specific occurrences on the basis of those interdependencies.

An important methodological function is attributed to hypothetical entities in an essay by Hull on intervening variables in molar behavior theory.[42] Suppose that in order to explain or predict the response of a subject in a given situation, we ascribe to him, for the time t_1 of his response, a certain habit strength, which has the status of a hypothetical entity. That strength is, in Hull's theory, "merely a quantitative representation of the perseverative after-effects" of certain earlier observable events, such as observable stimuli received in temporally remote learning situations. Consequently, if reference to habit strength were avoided by linking the subject's observable response at t_1 directly to the observable stimuli received earlier, then we would be invoking, as causal determinants for the response, certain observable events which at the time of the response have long ceased to exist. And Hull rejects this notion of causal action over a temporal distance: "it is hard to believe that an event such as stimulation in a remote learning situation can be causally active long after it has ceased to act on the receptors. I fully agree with Lewin that all the factors alleged to be causally influential in the determination of any other event must be in existence at the

41. See Tolman (1936), as reprinted in Marx (1951), p. 89; and Spence (1944), p. 65n.
42. Hull (1943).

time of such causal action."[43] Reference to the habit strength of the subject at the time t_1 of his response permits an explanation that accords with this principle.

Though the concluding part of the quoted passage sounds somewhat metaphysical, the basic import of Hull's argument is methodological. It credits the assumption of explanatory hypothetical entities with an accomplishment that is well described by Feigl in another context: "the discontinuous and historical character (action at a spatial and/or temporal distance) of the phenomenalistically restricted account vanishes and is replaced by a spatio-temporally continuous (contiguous) and nomologically coherent formulation on the level of hypothetical construction."[44] Such spatio-temporally continuous theories appear to recommend themselves for at least two reasons: first, they possess a certain formal simplicity, which at present can hardly be characterized in precise terms, but which is reflected, for example, in the possibility of using the powerful and elegant mathematical machinery of the calculus for the deduction, from the postulates of the theory, of explanatory and predictive connections among particular occurrences. And second, as was mentioned in section 3, the past development of empirical science seems to show that explanatory and predictive principles asserting discontinuous connections among (spatio-temporally separated) observable events are likely to be found to have limited scope and various exceptions. Theories in terms of hypothetical entities frequently make it possible to account for such exceptions by means of suitable assumptions concerning the hypothetical entities involved.

Another, more general, argument has been developed by Braithwaite, who gives credit to Ramsey for the basic principle.[45] Braithwaite's main contention is that "theoretical terms can only be defined by means of observable properties on condition that the theory cannot be adapted properly to apply to new situations."[46] He elaborates this idea by reference to a precisely formulated, miniature model of an interpreted theory. Without going into the details of that model, which would require too long a digression here, Braithwaite's claim can be adequately illustrated, it seems, by the following example: Suppose that the term 'temperature' is interpreted, at a certain stage of scientific research, only by reference to the readings of a mercury thermometer. If this observational criterion is taken as just a partial interpretation (namely as a sufficient but not necessary condition), then the possibility is left open of adding further partial interpretations, by reference to other thermometrical substances which are

43. Hull (1943), p. 285.
44. Feigl (1950), p. 40.
45. See the essay "Theories" in Ramsey (1931).
46. Braithwaite (1953), p. 76.

usable above the boiling point or below the freezing point of mercury; this permits a vast increase in the range of application of such laws as those connecting the temperature of a metal rod with its length or with its electric resistance, or the temperature of a gas with its pressure or its volume. If, however, the original criterion is given the status of a complete definiens, then the theory is not capable of such expansion; rather, the original definition has to be abandoned in favor of another one, which is incompatible with the first.[47]

The concept of intelligence lends itself to a similar argument: if test criteria which presuppose, on the part of the subject, the ability to read or at least to use language extensively are accorded the status of full definitions, then difficulties of the sort just indicated arise when the concept and the corresponding theory are to be extended to very young children or to animals.

However, the argument here outlined can hardly be said to establish what is claimed, namely that "A theory which it is hoped may be expanded in the future to explain more generalizations than it was originally designed to explain must allow more freedom to its theoretical terms than would be given them were they to be logical constructions out of observable entities"[48] (and thus defined in terms of the latter). For clearly, the procedure of expanding a theory at the cost of changing the definitions of some theoretical terms is not logically faulty; nor can it even be said to be difficult or inconvenient for the scientist, for the problem at hand is rather one for the methodologist or the logician, who seeks to give a clear "explication" or "logical reconstruction" of the changes involved in expanding a given theory. In the type of case discussed by Braithwaite, for example, this can be done in alternative ways—either in terms of additions to the original partial interpretation, or in terms of a total change of definition for some theoretical expressions. And if it is held that this latter method constitutes, not an expansion of the original theory, but a transition to a new one, this would raise more a terminological question than a methodological objection.

But though the above argument against definition does not have the intended systematic weight, it throws into relief an important heuristic aspect of scientific theorizing: when a scientist introduces theoretical entities such as electric currents, magnetic fields, chemical valences, or subconscious mechanisms, he intends them to serve as explanatory factors which have an existence independent of the observable symptoms by which they manifest themselves; or, to put it in more sober terms: whatever observational criteria of application the scientist may provide are intended by him to describe just symptoms or indications of the

47. This point is also made in Carnap (1936-1937), section 7, in a discussion of the advantages of reduction sentences over definitions. Feigl argues in the same vein in his essay (1951), in which the general principle is illustrated by examples from physics and psychology.
48. Braithwaite (1953), p. 76.

presence of the entity in question, but not to give an exhaustive characterization of it. The scientist does indeed wish to leave open the possibility of adding to his theory further statements involving his theoretical terms—statements which may yield new interpretative connections between theoretical and observational terms; and yet he will regard these as additional assumptions about the same hypothetical entities to which the theoretical terms referred before the expansion. This way of looking at theoretical terms appears to have definite heuristic value. It stimulates the invention and use of powerfully explanatory concepts for which only some links with experience can be indicated at the time, but which are fruitful in suggesting further lines of research that may lead to additional connections with the data of direct observation.[49].

The survey made in the present section has yielded no conclusive argument for or against the possibility of explicitly defining all theoretical terms of empirical science by means of a purely observational vocabulary; and in fact we have found strong reasons to doubt that any argument can settle the question once and for all.

As for the theoretical terms currently in use, it is impossible at present to formulate observational definientia for all of them, and thus to make them, in principle, unnecessary. In effect, therefore, most theoretical terms are presently used in science on the basis of only a partial experiential interpretation; and this use, as we noted, appears to offer distinct heuristic advantages.

In view of the importance that thus attaches to the idea of partial interpretation, we will now consider what kind of formal account might be given of it, and we will then turn to the question whether, or in what sense, the verdict of dispensability as proclaimed by the "theoretician's dilemma" applies also to theoretical terms which have been only partially interpreted, and which, therefore, cannot be dispensed with simply by virtue of definition.

8. INTERPRETATIVE SYSTEMS

Carnap's theory of reduction sentences is the first systematic study of the logic of partial definition. The introduction of a term by means of a chain of reduction sentences differs in two significant respects from the use of a chain of definitions. First, it specifies the meaning of the term only partially and thus does not provide a way of eliminating the term from all contexts in which it may occur. Second, as a rule, it does not amount just to a notational convention,

49. A concise synopsis of various arguments in favor of invoking "hypothetical constructs" will be found in Feigl (1950), pp. 38–41. Some aspects of the "semantic realism" concerning theoretical terms which Feigl presents in the same article are discussed in section 10 of the present essay.

but involves empirical assertions. If, for example, the term 'Q' is introduced by the two reduction sentences

(8.1) $C_1x \supset (Qx \equiv E_1x)$

(8.2) $C_2x \supset (Qx \equiv E_2x)$

then the following empirical law is asserted by implication:

(8.3) $(x)[(C_1x \cdot E_1x) \supset (C_2x \supset E_2x)]$

i.e., roughly speaking: any object that shows a positive response under the first test condition will, when put into the second test condition, show a positive response as well. Thus, a chain of reduction sentences for a given term normally combines two functions of language that are often considered as sharply distinct: the stipulative assignment of meaning, and the assertion or description of empirical fact.

Reduction sentences, as we saw earlier, are very well suited for the formulation of operational criteria of application as partial definitions. But they are subject to rather severe limitations as to logical form and thus they do not seem sufficient to provide a satisfactory general schema for the partial interpretation of theoretical terms.[50] A broader view of interpretation is suggested by Campbell's conception of a physical theory as consisting of a "hypothesis," represented by a set of sentences in theoretical terms, and a "dictionary," which relates the latter to concepts of experimental physics (which must be interconnected by empirical laws).[51] In contrast to the standard conception of a dictionary, Campbell's dictionary is assumed to contain, not definitions for the theoretical terms, but statements to the effect that a theoretical sentence of a certain kind is true if and only if a corresponding empirical sentence of a specified kind is true. Thus, rather than definitions, the dictionary provides rules of translation; and partial rules at that, for no claim is made that a translation must be specified for each theoretical statement or for each empirical statement.

This latter feature accords well, for example, with the consideration that a particular observable macrostate of a given physical system may correspond to a large number of theoretically distinguishable microstates; so that, for a theoretical sentence describing just one of those micro-states, the sentence describing the corresponding macrostate does not express a necessary and sufficient condition, and hence provides no translation.[52]

50. This has been pointed out by Carnap himself; see, for example, his (1956).

51. See Campbell (1920), Chapter VI. Important parts of this chapter are reprinted in Feigl and Brodbeck (1953).

52. However, this does not show that there cannot possibly be any necessary and sufficient condition in observational terms for the theoretical sentence: the problem of proving or disproving this latter claim is subject to difficulties analogous to those discussed in section 7 in regard to definability.

The statements in Campbell's dictionary evidently do not have the character of reduction sentences; they might be formulated, however, as biconditionals in which a sentence in theoretical terms is connected, by an "if and only if" clause, with a sentence in observational terms.

In other contexts, neither reduction sentences nor such biconditionals seem to be adequate. For as a rule, the presence of a hypothetical entity H, such as a certain kind of electric field, will have observable symptoms only if certain observational conditions, O_1, are satisfied, such as the presence of suitable detecting devices, which will then have to show observable responses, O_2. A sentence stating this kind of criterion would have the character of a generalized reduction sentence; it might be put into the form.

(8.4) $O_1 \supset (H \supset O_2)$

where 'O'_1 and 'O_2' are sentences—possibly quite complex ones—in terms of observables, and 'H' is a sentence which is expressed in theoretical terms.

But there is no good reason to limit interpretative statements to just the three types here considered. In order to obtain a general concept of partial interpretation, we will now admit as interpretative statements any sentences, of whatever logical form, which contain theoretical and observational terms. On the assumption that the theoretical and observational statements of empirical science are formulated within a specified logical framework, this idea can be stated more precisely and explicitly as follows:

(8.5) Let T be a theory characterized by a set of postulates in terms of a finite *theoretical vocabulary* V_T, and let V_B be a second set of extra-logical terms, to be called the *basic vocabulary*, which shares no term with V_T. By an *interpretative system* for T with the basis V_B we will then understand a set J of sentences which (i) is finite, (ii) is logically compatible with T, (iii) contains no extra-logical term that is not contained in V_T or V_B, (iv) contains every element of V_T and V_B essentially, i.e., is not logically equivalent to some set of sentences in which some term of V_T or V_B does not occur at all.[53]

In applying the concept here defined to the analysis of scientific theories, we will have to assume, of course, that V_B consists of terms which are antecedently understood. They might be observational terms, in the somewhat vague sense

53. The intuitive notion of interpretation, as well as the conception reflected in Campbell's idea of an interpretative dictionary, would seem to call for the following additional condition: (v) Each sentence of J contains essentially terms from V_T as well as terms from V_B. However, this requirement introduces no further restriction of the concept of interpretative system; for any system J that meets conditions (i) to (iv) can be stated in an equivalent form that satisfies (v) as well. To this end, it suffices to replace the member sentences of J by their conjunction; this yields a logically equivalent interpretative system which contains only one sentence, and which satisfies (v) since J satisfies (iv).

explained earlier, but we need not insist on this. One might well take the view, for example, that certain disposition terms such as 'malleable', 'elastic', 'hungry', and 'tired' are not strictly observation terms, and are not known to be explicitly definable by means of observation terms; and yet, such terms might be taken to be well understood in the sense that they are used with a high degree of agreement by competent observers. In this case, it would be quite reasonable to use these terms in interpreting a given theory, i.e., to admit them into V_B.

Campbell's conception of the function of his "dictionary" illustrates this possibility very well and shows that it comes closer to actual scientific procedure. Campbell specifies that the interpretation provided by the dictionary must be in terms of what he calls "concepts," such as the terms 'temperature', 'electrical resistance', 'silver', and 'iron' as used in experimental physics and chemistry. These are hardly observational in the narrow sense, for they are specifically conceived as representing clusters of empirical laws: "Thus, if we say anything about electrical resistance we assume that Ohm's Law is true; bodies for which Ohm's Law is not true, gases for example, have no electrical resistance."[54] But even though one might not wish to qualify these terms as observational, one may still consider them as well understood, and as used with high intersubjective agreement, by scientific experimenters; and thus, they might be admitted into V_B.

Interpretative systems as just defined include as special cases all the types of interpretation we considered earlier, namely, interpretation by explicit definitions for all theoretical terms, by chains of reduction sentences, by biconditional translation statements in the sense of Campbell's dictionary, and by generalized reduction sentences of the form (8.4); but of course they also allow for interpretative statements of many other forms.

Interpretative systems have the same two characteristics which distinguish chains of reduction sentences from chains of definitions: First, an interpretative system normally effects only a partial interpretation of the terms in V_T; i.e., it does not lay down (by explicit statement or by logical implication), for every term in V_T, a necessary and sufficient condition of application in terms of V_B. Second, like a chain of reduction sentences for a given theoretical term, an interpretative system will normally not be purely stipulative in character, but will imply certain statements in terms of V_B alone which are not logical truths, and which, on the conception of V_B as consisting of antecedently understood empirical terms, express empirical assertions. Thus, here again, we find a combination of the stipulative and the descriptive use of language.

54. Campbell (1920), p. 43.

But, to turn to a third point of comparison, an interpretative system need not provide an interpretation—complete or incomplete—for each term in V_T individually. In this respect it differs from a set of definitions, which specifies for each term a necessary and sufficient condition, and from a set of reduction sentences, which provides for each term a necessary and a—usually different—sufficient condition. It is quite possible that an interpretative system provides, for some or even all of the terms in V_T, no necessary or no sufficient condition in terms of V_B, or indeed neither of the two; instead, it might specify, by explicit statement or by logical implication, sufficient or necessary conditions in terms of V_B only for certain expressions containing several terms of V_T—for example, in the manner of Campbell's dictionary.

As a rule, therefore, when a theory T is interpreted by an interpretative system J, the theoretical terms are not dispensable in the narrow sense of being replaceable in all contexts by defining expressions in terms of V_B. Nor are they generally dispensable in the sense that J provides, for every sentence H that can be formed by means of V_T, a "translation" into terms of V_B, i.e., a sentence O in terms of V_B such that the biconditional $H \equiv O$[55] is logically deducible from J.

Are theoretical terms, then, altogether indispensable on this broad conception of interpretation, so that the "paradox of theorizing" formulated in section 5 no longer applies to them? We consider this question in the next section.

9. FUNCTIONAL REPLACEABILITY OF THEORETICAL TERMS

Transition from observation to prediction (or explanation)

The systematizing function of a theory T interpreted by an interpretative system J will consist in permitting inferences from given "data" in terms of V_B to certain other (e.g., predictive) statements in terms of V_B. If O_1 is the statement expressing the data, O_2 the inferred statement, then the connection may be symbolized thus:

(9.1) $(O_1 \cdot T \cdot J) \longrightarrow O_2$

Here, as in similar contexts below, 'T' stands for the set of postulates of the theory at hand; the arrow represents deductive implication.

A logically equivalent formulation

Now, (9.1) holds if and only if $T \cdot J$ implies the sentence $O_1 \supset O_2$; so that (9.1) is tantamount to

(9.2) $(T \cdot J) \longrightarrow (O_1 \supset O_2)$

55. Here, and on some subsequent occasions where there is no danger of misunderstandings, logical connectives are used autonymously; the expression '$H \equiv O$', for example, represents the sentence obtained by placing the triple-bar symbol (for 'if and only if') between the sentences of which 'H' and 'O' are names.

Whatever systematization is achieved among the V_B-sentences is clearly accomplished by T in conjunction with J. It will be convenient therefore to consider the postulates of T together with the sentences of J as the postulates of a deductive system T', which will be called an *interpreted theory*. Its vocabulary $V_{T'}$ will be the sum of V_T and V_B.

T′=T&J

What was noted in connection with (9.1) and (9.2) may now be restated thus: If an interpreted theory T' establishes a deductive transition from O_1 to O_2, i.e., if

(9.3) $(O_1 \cdot T') \longrightarrow O_2$

then

(9.4) $T' \longrightarrow (O_1 \supset O_2)$

and conversely, where T' is the set of postulates of the interpreted theory.

Logical systematization is achievable by observational terms alone

Now it can readily be shown that an interpreted theory T' establishes exactly the same deductive connections among V_B-sentences as does the set of all those theorems of T' which are expressible in terms of V_B alone; we will call this the set of *V_B-theorems*, or *V_B-consequences*, of T', and will designate it by '$O_{T'}$'. This means that for all purposes of deductive systematization, T' is, as we will say, *functionally equivalent* to the set $O_{T'}$ which contains not a single theoretical term.

The proof is as follows: The deductive transition, represented in (9.3), from O_1 to O_2 can be achieved just as well by using, instead of T', simply the sentence $O_1 \supset O_2$, which by virtue of (9.4) belongs to $O_{T'}$; for we have, by *modus ponens*,

(9.5) $[O_1 \cdot (O_1 \supset O_2)] \longrightarrow O_2$

And since $O_{T'}$ surely contains all the V_B-sentences of the form $O_1 \supset O_2$ that are implied by T', the set $O_{T'}$ suffices to effect all the deductive systematizations achievable by means of T'. On the other hand, $O_{T'}$ is no stronger in this respect than T'; for $O_{T'}$ permits the deductive transition from O_1 to O_2 only if it implies $O_1 \supset O_2$; but in this case T', too, implies $O_1 \supset O_2$, which means, in view of the equivalence of (9.4) with (9.3), that T' will permit the deductive transition from O_1 to O_2.

Thus, *the deductive systematization that an interpreted theory T' achieves among sentences expressed in terms of a basic vocabulary V_B is exactly the same as that accomplished by the set $O_{T'}$ of those statements (theorems) of T' which can be expressed in terms of V_B alone*. In this sense, the theoretical terms used in T can be dispensed with.

Any such system would be cumbersome. But we can rephrase it using just the basic vocabulary.

But $O_{T'}$ is normally an unwieldy infinite set of statements, and the question arises therefore whether there is some generally applicable method of making it more manageable and perspicuous by putting it into the form of an axiomatized theoretical system T'_B, which would be formulated in terms of V_B alone. A theorem in formal logic proved by Craig shows that this is indeed the case, provided only that T' satisfies certain extremely liberal and unconfining conditions.[56]

Craig's method

Thus Craig's theorem has a definite bearing upon the problems raised by the "paradox of theorizing," which was stated in section 5 in somewhat vague terms. The theorem indicates one way in which the "paradox" can be given a clear and precise interpretation and a rigorous proof: It shows that for any theory T' using both theoretical terms and nontheoretical, previously understood ones,

56. Craig's paper (1953) contains the first published account of this important theorem. A less condensed and less technical presentation, with explicit through brief references to applications such as the one here considered, is given in Craig (1956).

In application to the issue we are discussing, the result obtained by Craig may be briefly stated as follows: Let the set $V_{T'}$ of primitive terms of T′ and the set of postulates of T′ be specified effectively, i.e., in a manner providing a general procedure which, for any given expression, will decide in a finite number of steps whether or not the expression is a primitive term (or a postulate) of T′. Let $V_{T'}$ be divided, by an effective criterion that may otherwise be chosen at will, into two mutually exclusive vocabularies, V_T and V_B. Finally, let the rules of the logic used be such that there is an effective method of determining, for any given finite sequence of expressions, whether it is a valid deduction according to those rules.

Then there exists a general method (i.e., a method applicable in all cases meeting the conditions just outlined) of effectively constructing (i.e., effectively characterizing the postulates and the rules of inference of) a new system T'_B whose set of primitives is V_B and whose theorems are exactly those theorems of T′ which contain no extralogical constants other than those contained in V_B.

Note that the theorem permits us to draw the dividing line between V_T and V_B wherever we please, as long as the criterion used to effect the division permits us to decide in a finite number of steps to which of the two sets a given term belongs. This condition as well as the requirement of an effective characterization of $V_{T'}$ will be trivially satisfied, for example, if $V_{T'}$ is finite and its member terms as well as those of V_B and V_T are specified simply by enumerating them individually.

The further requirement of an effective characterization of the postulates and the rules of logic for T′ are so liberal that no doubt any scientific theory that has yet been considered can be formalized in a manner that satisfies them—as long as the connections between theoretical and observational expressions can be assumed to be expressible in the form of definite statements. The only important case I am aware of in which this condition would be violated is that of a theory for which no definite rules of interpretation are specified—say, on the ground that the criteria of application for theoretical expressions always have to be left somewhat vague. A conception of this kind may have been intended, for example, by A. Wald's remark "In order to apply [a scientific] theory to real phenomena, we need some rules for establishing the correspondence between the idealized objects of the theory and those of the real world. These rules will always be somewhat vague and can never form a part of the theory itself." Wald (1942), p. 1.

The conditions of Craig's theorem are satisfiable, however, if the vagueness here referred to is reflected in definite rules. Thus, for example, the interpretative sentences for a given theory might take the form of statistical probability statements (a possibility mentioned in Carnap (1956), section 5), or perhaps of logical probability statements (each specifying the logical probability of some theoretical sentence relative to a specified sentence in observational terms, or vice versa). Either of these procedures would yield an interpretation of a more general kind than that characterized by the definition of an interpretative system given in section 8 of the present essay. Yet even to theories which are interpreted in this wider sense, Craig's theorem can be applied.

there exists, under certain very widely satisfied conditions, an axiomatized theoretical system T'_B which uses only the nontheoretical terms of T' and yet is functionally equivalent with T' in the sense of effecting, among the sentences expressible in the nontheoretical vocabulary, exactly the same deductive connections as T'.

Even though replaceability is achievable, is it good?

Should empirical science then avail itself of this method and replace all its theories involving assumptions about hypothetical entities by functionally equivalent theoretical systems couched exclusively in terms which have direct observational reference or which are, at any rate, clearly understood? There are various reasons which make this inadvisable in consideration of the objectives of scientific theorizing.

Severe losses in simplicity of deductive presentation

To begin with, let us consider the general character of Craig's method. Disregarding many subtle points of detail, the procedure may be described as follows: By means of a constructive procedure, Craig arranges all the V_B-theorems of T' in a sequence. This sequence is highly redundant, for it contains, for any sentence occurring in it, also all its logical equivalents (as far as they are expressible in V_B). Craig prescribes a procedure for eliminating many, though not all, of these duplications. The remaining sequence therefore still contains each V_B-theorem of T' in at least one of its various equivalent formulations. Finally, all the sentences in this remaining sequence are made postulates of T'_B. Thus, the set of V_B-theorems of T' is "axiomatized" in T'_B only in a rather Pickwickian sense, namely by making every sentence of the set, in some of its many equivalent formulations, a postulate of T'_B. Normally, the axiomatization of a set of sentences selects as postulates just a small subset from which the rest can then be logically derived as theorems; thus, the axiomatization presents the content of the whole set "in a form which is psychologically or mathematically more perspicuous."[57] And since Craig's method in effect includes all sentences that are to be axiomatized among the postulates of T', the latter, as Craig himself puts it, "fail to simplify or to provide genuine insight."[58]

The loss in simplicity which results from discarding the theoretical terms of T' is reflected in the circumstance that the set of postulates which Craig's method yields for T'_B is always infinite. Even in cases where actually there exists some finite subset of $O_{T'}$ of V_B-theorems of T' from which all the rest can be deduced,

57. Craig (1956), p. 49. It may be well to note briefly two further points which were established by Craig, in the studies here referred to: (i) A theory T′ may have a set of V_B-consequences that cannot be axiomatized by means of a *finite* set of postulates expressible in terms of V_B. (ii) There is no general method that permits an effective decision, for every theory T′, as to whether its V_B-consequences can, or cannot, be axiomatized by means of a finite set of postulates.

58. Craig (1956), p. 49. This fact does not detract in the least, of course, from the importance and interest of Craig's result as a theorem in logic.

Craig's procedure will not yield such a subset: that is the price of its universal applicability.

Now there are cases where an infinity of postulates may not be excessively unwieldy; notably when the axioms are specified by means of axiom-schemata,[59] i.e., by stipulations to the effect that any sentence that has one of a finite number of specified forms (such as '$x = x$,' for example) is to count as an axiom. But the manner in which postulates of T'_B are specified by Craig's method is vastly more intricate, and the resulting system would be practically unmanageable—to say nothing of the loss in heuristic fertility and suggestiveness which results from the elimination of the theoretical concepts and hypotheses. For empirical science, therefore, this method of dispensing with theoretical expressions would be quite unsatisfactory.

Can the theory T'(B) be used in predictions?

So far, we have examined the eliminability of theoretical concepts and assumptions only in the context of deductive systematization: we considered an interpreted theory T' exclusively as a vehicle of establishing deductive transitions among observational sentences. However, such theories may also afford means of inductive systematization in the sense outlined in section 1; an analysis of this function will yield a further argument against the elimination of theoretical expressions by means of Craig's method.

The phosphorus example to illustrate the links between different terms

By way of illustration I will use an example which is deliberately oversimplified in order the more clearly to exhibit the essentials. Let us assume that V_T contains the term 'white phosphorus', or 'P' for short, and that the interpretative system incorporated into T' states no sufficient observational conditions of application for it, but several necessary ones. These will be taken to be independent of each other in the sense that, though in the case of white phosphorus they occur jointly, any one of them occurs in certain other cases in the absence of one or more of the others. Let those necessary conditions be the following: white phosphorus has a garlic-like odor; it is soluble in turpentine, in vegetable oils, and in ether; it produces skin burns. In symbolic notation:

(9.6) $(x)(Px \supset Gx)$
(9.7) $(x)(Px \supset Tx)$
(9.8) $(x)(Px \supset Vx)$
(9.9) $(x)(Px \supset Ex)$
(9.10) $(x)(Px \supset Sx)$

All predicates other than 'P' that occur in these sentences will belong, then, to V_B.

Now let V_T contain just one term in addition to 'P', namely 'has an ignition temperature of 30° C', or 'I' for short; and let there be exactly one interpretative sentence for 'I', to the effect that if an object has the property I then it will burst

59. On this method, first used by von Neumann, see Carnap (1937), pp. 29-30 and p. 96, where further references to the literature are given.

into flame if surrounded by air in which a thermometer shows a reading above 30° C. This property will be considered as observable and will be represented by the predicate 'F' in V_B. The interpretative sentence for 'I', then, is

(9.11) $(x)(Ix \supset Fx)$

Finally, we will assume that the theoretical part of T' contains one single postulate, namely,

(9.12) $(x)(Px \supset Ix)$

which states that white phosphorus has an ignition temperature of 30° C. Let the seven sentences (9.6)-(9.12) represent the total content of T'.

Then, as is readily seen, T' has no consequences in terms of V_B except for purely logical truths; consequently, T' will permit a deductive transition from one V_B-sentence to another only if the latter is logically implied by the former, so that T' is not required to establish the connection. In other words: T' effects no deductive systematization among V_B-sentences at all. Nevertheless, T' may play an essential role in establishing certain explanatory or predictive connections of an inductive kind among the V_B-sentences. Suppose, for example, that a certain object b has been found to have all the characteristics G, T, V, E, S. In view of the sentences (9.6)-(9.10), according to which these characteristics are symptomatic of P, it might then well be inferred that b is white phosphorus. This inference would be inductive rather than deductive, and part of its strength would derive from the mutual independence which we assumed to exist among those five observable symptoms of white phosphorus. The sentence 'Pb' which has thus been inductively accepted leads, via (9.12), to the prediction 'Ib', which in turn, in virtue of (9.11), yields the forecast 'Fb'. Thus, T' permits the transition from the observational data 'Gb', 'Tb', 'Vb', 'Eb', 'Sb' to the observational prediction 'Fb'. But the transition requires an inductive step, consisting of the acceptance of 'Pb' on the strength of the five data sentences, which support, but do not logically imply, 'Pb'.

On the other hand, the system T'_B obtained by Craig's method does not lend itself to this inductive use; in fact, all its sentences are logical truths and thus T'_B makes no empirical assertion at all, for, as was noted above, all the V_B-theorems of T' are logically true statements.

See the description in page 213

Thus, if the systematizing use of an interpreted theory T' is conceived as involving inductive as well as deductive procedures, then the corresponding system T'_B cannot, in general, replace T'.

Pragmatic reasons to prefer T'

An intuitively simpler method of obtaining a functional equivalent, in observational terms, of a given interpreted theory T' is provided by an idea of Ramsey's. In effect, the method amounts to treating all theoretical terms as existentially quantified variables, so that all the extralogical constants that occur in Ramsey's manner of formulating a theory belong to the observational vocab-

Ramsey's method

ulary.[60] Thus, the interpreted theory determined by the formulas (9.6)-(9.12) would be expressed by the following sentence, which we will call the *Ramsey-sentence associated with the given theory*:

Ramsey sentence

(9.13) $(\exists\phi)(\exists\psi)(x)[(\phi x \supset (Gx \cdot Tx \cdot Vx \cdot Ex \cdot Sx)) \cdot (\psi x \supset Fx) \cdot (\phi x \supset \psi x)]$

This sentence is equivalent to the expression obtained by conjoining the sentences (9.6)-(9.12), replacing 'P' and 'I' throughout by the variables 'ϕ' and 'ψ' respectively, and prefixing existential quantifiers with regard to the latter. Thus, (9.13) asserts that there are two properties, ϕ and ψ, otherwise unspecified, such that any object with the property ϕ also has the observable properties G, T, V, E, S; any object with the property ψ also has the observable property E; and any object with the property ϕ also has the property ψ.

Ramsey sentence mirrors the pragmatic benefits of T and T(B)

An interpreted theory T' is not, of course, logically equivalent with its associated Ramsey-sentence any more than it is logically equivalent with the associated Craig-system T'_B; in fact, each of the two is implied by, but does not in turn imply, T'. But though the Ramsey-sentence contains, apart from variables and logical constants, only terms from V_B, it can be shown to imply exactly the same V-sentences as does T'; hence, it establishes exactly the same deductive transitions among V_B-sentences as does T'. In this respect then, the Ramsey-sentence associated with T' is on a par with the Craig-system T'_B obtainable from T'. But its logical apparatus is more extravagant than that required by T' or by T'_B. In our illustration, for example, T' and T'_B contain variables and quantifiers only with respect to individuals (physical objects), whereas the Ramsey-sentence (9.13) contains variables and quantifiers also for properties of individuals; thus, while T' and T'_B require only a first-order functional calculus, the Ramsey-sentence calls for a second-order functional calculus.

But it is committed to properties

And hence to unobservables

But this means that the Ramsey-sentence associated with an interpreted theory T' avoids reference to hypothetical entities only in letter—replacing Latin constants by Greek variables—rather than in spirit. For it still asserts the existence of certain entities of the kind postulated by T', without guaranteeing any more than does T' that those entities are observables or at least fully characterizable in terms of observables. Hence, Ramsey-sentences provide no satisfactory way of avoiding theoretical concepts.

And indeed, Ramsey himself made no such claim. Rather, his construal of theoretical terms as existentially quantified variables appears to have been motivated by considerations of the following kind: If theoretical terms are treated as constants which are not fully defined in terms of antecedently understood observational terms, then the sentences that can formally be constructed out of them do not have the character of assertions with fully specified meanings,

60. Ramsey (1931), pp. 212-15, 231.

which can be significantly held to be either true or false; rather, their status is comparable to that of sentential functions, with the theoretical terms playing the role of variables. But of a theory we want to be able to predicate truth or falsity, and the construal of theoretical terms as existentially quantified variables yields a formulation which meets this requirement and at the same time retains all the intended empirical implications of the theory.

This consideration raises a further problem, which will be discussed in the next section.

10. ON MEANING AND TRUTH OF SCIENTIFIC THEORIES

The problem suggested by Ramsey's approach is this: If, in the manner of section 8, we construe the theoretical terms of a theory as extralogical constants for which the system J provides only a partial interpretation in terms of the antecedently understood vocabulary V_B, can the sentences formed by means of the theoretical vocabulary nevertheless be considered as meaningful sentences which make definite assertions, and which are either true or false?

The question might seem to come under the jurisdiction of semantics, and more specifically, of the semantical theory of truth. But this is not the case. What the semantical theory of truth provides (under certain conditions) is a general definition of truth for the sentences of a given language L. That definition is stated in a suitable metalanguage, M, of L and permits the formulation of a necessary and sufficient condition of truth for any sentence S of L. This condition is expressed by a translation of S into M.[61] (To be suited for its purpose, M must therefore contain a translation of every sentence of L and must meet certain other conditions which are specified in the semantical theory of truth.) But if the truth criteria thus stated in M are to be intelligible at all, then clearly all the translations of L-statements into M must be assumed to be significant to begin with. Instead of deciding the question as to the meaningfulness of L-sentences, the semantical definition of truth presupposes that it has been settled antecedently.

For analogous reasons, semantics does not enable us to decide whether the theoretical terms in a given system T' do or do not have semantical, or factual, or ontological reference—a characteristic which some writers have considered as distinguishing genuinely theoretical constructs from auxiliary or intervening theoretical terms.[62] One difficulty with the claims and counterclaims that have been made in this connection lies in the failure of the discussants to indicate clearly what they wish to assert by attributing ontological reference to a given

61. See Tarski (1944), section 9.

62. On this point, see for example, MacCorquodale and Meehl (1948); Lindzey (1953); Feigl (1950), (1950a); Hempel (1950); Rozeboom (1956).

term. From a purely semantical point of view, it is possible to attribute semantical reference to any term of a language L that is taken to be understood: the referent can be specified in the same manner as the truth condition of a given sentence in L, namely by translation into a suitable metalanguage. For example, using English as a metalanguage, we might say, in reference to Freud's terminology, that 'Verdraengung' designates repression, 'Sublimierung', sublimation, and so on. Plainly, this kind of information is unilluminating for those who wish to use existential reference as a distinctive characteristic of a certain kind of theoretical term; nor does it help those who want to know whether, or in what sense, the entities designated by theoretical terms can be said actually to exist—a question to which we will return shortly.

Semantics, then, does not answer the question raised at the beginning of this section; we have to look elsewhere for criteria of significance for theoretical expressions.

Generally speaking, we might qualify a theoretical expression as intelligible or significant if it has been adequately explained in terms which we consider as antecedently understood. In our earlier discussion, such terms were represented by the vocabulary V_B (plus the terms of logic). But now the question arises: What constitutes an "adequate" explanation? No generally binding standards can be specified: the answer is ultimately determined by one's philosophical conscience. The logical and epistemological puritan might declare intelligible only what has been explicitly defined in terms of V_B; and he might impose further restrictions—in a nominalistic vein, for example—on the logical apparatus that may be used in formulating the definitions. Others will find terms introduced by reduction sentences quite intelligible, and still others will even countenance an interpretation as tenuous as that afforded by an interpretative system. One of the most important advantages of definition lies in the fact that it ensures the possibility of an equivalent restatement of any theoretical sentence in terms of V_B. Partial interpretation does not guarantee this; consequently it does not provide, for every sentence expressible in theoretical terms, a necessary and sufficient condition of truth that can be stated in terms which are antecedently understood. This, no doubt, is the basic difficulty that critics find with the method of partial interpretation.

In defense of partial interpretation, on the other hand, it can be said that to understand an expression is to know how to use it, and in a formal reconstruction the "how to" is expressed by means of rules. Partial interpretation as we have construed it provides such rules. These show, for example, what sentences in terms of V_B alone may be inferred from sentences containing theoretical terms; and thus they specify a set of V_B-sentences that are implied, and hence indirectly asserted, by an interpreted theory T'. (If the set is empty, the theory does not fall within the domain of empirical science.) Conversely, the rules also show what sentences in theoretical terms may be inferred from V_B-sentences. Thus, there are close resemblances between our theoretical sentences and those sentences which are intelligible in the narrower sense of being expressible entirely in terms of V_B—a circumstance which militates in favor of admitting theoretical sentences into the class of significant statements.

It should be mentioned that if this policy is adopted, then we will have to recognize as significant (though not, of course, as interesting or worth investigating) certain interpreted systems which surely would not qualify as potential scientific theories. For example, let L be the conjunction of some finite number of empirical generalizations about learning behavior, formulated in terms of an observational vocabulary V_B, and let P be the conjunction of a finite number of arbitrary sentences formed out of a set V_T of arbitrarily chosen uninterpreted terms (for example, P might be the conjunction of the postulates of some axiomatization of elliptic geometry). Then, by making P the postulates of T and by choosing the sentence $P \supset L$ as the only member of our interpretative system J, we obtain an interpreted theory T' which explains in a trivial way all the given empirical generalizations, since $T \cdot J$ plainly implies L. Yet, needless to say, T' would not be considered a satisfactory learning theory.[63] The characteristic here illustrated does not vitiate our analysis of partial interpretation, since the latter does not claim that every partially interpreted theoretical system is a potentially interesting theory; and indeed, even the requirement of full definition of all theoretical terms by means of V_B still leaves room for similarly unrewarding "theories." Examples like our mock "learning theory" simply remind us that, in addition to having an empirical interpretation (which is necessary if there are to be any empirically testable consequences) a good scientific theory must satisfy various important further conditions; its V_B-consequences must be empirically well confirmed; it must effect a logically simple systematization of the pertinent V_B-sentences, it must suggest further empirical laws, and so forth.

If the sentences of a partially interpreted theory T' are granted the status of significant statements, they can be said to be either true or false. And then the question, touched upon earlier in this section, as to the factual reference of theoretical terms, can be dealt with in a quite straightforward manner: To assert that the terms of a given theory have factual reference, that the entities they purport to refer to actually exist, is tantamount to asserting that what the theory tells us is true; and this in turn is tantamount to asserting the theory.

63. It is of interest to note here that if in addition to the conditions specified in section 8, an interpreted theory were also required to meet the criteria of significance for theoretical terms and sentences that have recently been proposed by Carnap (1956 sections 6, 7, 8), then the terms and the sentences of our mock "learning theory" would be ruled out as nonsignificant.

When we say, for example, that the elementary particles of contemporary physical theory actually exist, we assert that there occur in the universe particles of the various kinds indicated by physical theory, governed by specified physical laws, and showing certain specific kinds of observable symptoms of their presence in certain specified circumstances. But this is tantamount to asserting the truth of the (interpreted) physical theory of elementary particles. Similarly, asserting the existence of the drives, reserves, habit strengths, and the like postulated by a given theory of learning amounts to affirming the truth of the system consisting of the statements of the theory and its empirical interpretation.[64]

Thus understood, the existence of hypothetical entities with specified characteristics and interrelations, as assumed by a given theory, can be examined inductively in the same sense in which the truth of the theory itself can be examined, namely, by empirical tests of its V_B-consequences.

According to the conception just outlined, we have to attribute factual reference to all the (extra-logical) terms of a theory if that theory is true; hence, this characteristic provides no basis for a semantical dichotomy in the theoretical vocabulary. Also, the factual reference, as here construed, of theoretical terms does not depend on whether those terms are avoidable in favor of expressions couched in terms of V_B alone. Even if all the theoretical terms of a theory T' are explicitly defined in terms of V_B, so that their use affords a convenient shorthand way of saying what could also be said by means of V_B alone, they will still have factual reference if what the theory says is true.

The preceding observations on truth and factual reference in regard to partially interpreted theories rest on the assumption that the sentences of such theories are accorded the status of statements. For those who find this assumption unacceptable, there are at least two other ways of construing what we have called an interpreted theory. The first of these is Ramsey's method, which was described in the previous section. It has the very attractive feature of representing an interpreted theory in the form of a bona fide statement, which contains no extra-logical constants other than those contained in V_B, and which has exactly the same V_B-consequences as the theory stated in terms of incompletely interpreted theoretical constants. It is perhaps the most satisfactory way of conceiving the logical character of a scientific theory, and it will be objectionable mainly, or perhaps only, to those who, on philosophical grounds, are opposed to the ontological commitments[65] involved in countenancing variables that range over domains other than that of the individuals of the theory (such as, for example the set of all quantitative characteristics of physical objects, or the set of all dyadic relations among them, or sets of such sets, and so forth).

64. More precisely, the assertion that there exist entities of the various kinds (such as hypothetical objects and events and their various qualitative and quantitative properties and relations) postulated by an interpreted theory T′ is expressed by the Ramsey-sentence associated with T′. It is obtained by replacing all theoretical constants in the conjunction of the postulates of T′ by variables and binding all these by existential quantifiers placed before the resulting expression. The sentence thus obtained is a logical consequence of the postulates of T′; but the converse does not hold; hence strictly speaking, the assertion of the existence of the various hypothetical entities assumed in a theory is logically weaker than the theory itself.

For suggestive observations on the question of the reality of theoretical entities, see, for example, Toulmin (1953), pp. 134–139 and Smart (1956).

Those finally, who, like the contemporary nominalists, reject such strong ontological commitments, may adopt a conception of scientific theories, not as significant statements, but as intricate devices for inferring, from intelligible initial statements, expressed in terms of an antecedently understood vocabulary V_B, certain other, again intelligible, statements in terms of that vocabulary.[66] The nominalistically inclined may then construe theoretical terms as meaningless auxiliary marks, which serve as convenient symbolic devices in the transition from one set of experiential statements to another. To be sure, the conception of laws and theories as extralogical principles of inference does not reflect the way in which they are used by theoretical scientists. In publications dealing with problems of theoretical physics, or biology, or psychology, for example, sentences containing theoretical terms are normally treated on a par with those which serve to describe empirical data: together with the latter, they function as premises and as conclusions of deductive and of inductive arguments. And indeed, for the working scientist the actual formulation and use of theoretical principles as complex extralogical rules of inference would be a hindrance rather than a help. However, the purpose of those who suggest this conception is not, of course, to facilitate the work of the scientist but rather to clarify the import of his formulations; and from the viewpoint of a philosophical analyst with nominalistic inclinations the proposed view of scientific sentences which by his standards are not admissible as statements does represent an advance in clarification.

65. The concept is used here in Quine's sense, according to which a theory is ontologically committed to those entities which must be included in the domains over which its bound variables range if the theory is to be true. Quine develops and defends this idea in several of the essays comprising his book (1953).

66. The conception of laws or theories as inferential principles has been suggested, but by no means generally from a nominalistic point of view, by several authors; among them Schlick (1931), pp. 151 and 155; Ramsey (1931), p. 241; Ryle (1949), especially pp. 120–25; and Toulmin (1953), Chapters III and IV. (Toulmin remarks, however, that to think of laws of nature as rules or licenses "reflects only a part of their nature" (*loc. cit.*, p. 105).) See also Braithwaite's discussion of the issue in (1953), pp. 85–87. Finally, Popper's essay (1956) contains several critical and constructive comments that bear on this issue and on some of the other questions discussed in the present study.

However, the question posed by the theoretician's dilemma can be raised also in regard to the two alternative conceptions of the status of a theory. Concerning Ramsey's formulation, we may ask whether it is not possible to dispense altogether with the existentially quantified variables which represent the theoretical terms, and thus to avoid the ontological commitment they require, without sacrificing any of the deductive connections that the Ramsey-sentence establishes among V_B-sentences. And in regard to theories conceived as inferential devices, we may ask whether they cannot be replaced by a functionally equivalent set of rules—i.e., one establishing exactly the same inferential transitions among V_B-sentences—which uses none of the "meaningless marks."

To both questions, Craig's theorem gives an affirmative answer by providing a general method for constructing the desired kind of equivalent. But again, in both cases, the result has the shortcomings mentioned in section 8. First, the method would replace the Ramsey-sentence by an infinite set of postulates, or the body of inferential rules by an infinite set of rules, in terms of V_B, and would thus lead to a loss of economy. Second, the resulting system of postulates or of inferential rules would not lend itself to inductive prediction and explanation. And third, it would have the pragmatic defect, partly reflected already in the second point, of being less fruitful heuristically than the system using theoretical terms.

Our argument (5.1), the theoretician's dilemma, took it to be the sole purpose of a theory to establish deductive connections among observation sentences. If this were the case, theoretical terms would indeed be unnecessary. But if it is recognized that a satisfactory theory should provide possibilities also for inductive explanatory and predictive use and that it should achieve systematic economy and heuristic fertility, then it is clear that theoretical formulations cannot be replaced by expressions in terms of observables only; the theoretician's dilemma, with its conclusion to the contrary, is seen to rest on a false premise.

REFERENCES

Bergmann, Gustav. "The Logic of Psychological Concepts," *Philosophy of Science*, 18:93–110 (1951).

Bergmann, Gustav. "Comments on Professor Hempel's 'The Concept of Cognitive Significance'," *Proceedings of the American Academy of Arts and Sciences*, 80 (No. 1): 78–86 (1951). Reprinted in Gustav Bergmann, *The Metaphysics of Logical Positivism*. New York: Longmans, Green and Co., 1954. Referred to in this essay as (1951a).

Bergmann, Gustav, and Kenneth Spence. "Operationism and Theory in Psychology," *Psychological Review*, 48:1–14 (1941). Reprinted in Marx (1951).

Braithwaite, R. B. *Scientific Explanation*. Cambridge, England: Cambridge University Press, 1953.

Bridgman, P. W. *The Logic of Modern Physics*. New York: Macmillan, 1927.

Burks, Arthur W. "The Logic of Causal Propositions," *Mind*, 60:363–382 (1951).

Campbell, Norman R. *Physics: The Elements*. Cambridge: Cambridge University Press, 1920. Republished under the title *Foundations of Science*. New York: Dover, 1957.

Carnap, Rudolf. "Testability and Meaning," *Philosophy of Science*, 3:420–468 (1936); 4:1–40 (1937). Reprinted as a monograph by Whitlock's Inc., New Haven, Conn., 1950. Excerpts reprinted in Feigl and Brodbeck (1953).

Carnap, Rudolf. *The Logical Syntax of Language*. London: Routledge and Kegan Paul, 1937.

Carnap, Rudolf. *Foundations of Logic and Mathematics*. Chicago: University of Chicago Press, 1939.

Carnap, Rudolf. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 1950.

Carnap, Rudolf. "Meaning Postulates," *Philosophical Studies*, 3:65–73 (1952).

Carnap, Rudolf. "Meaning and Synonymy in Natural Languages," *Philosophical Studies*, 6:33–47 (1955).

Carnap, Rudolf. "The Methodological Character of Theoretical Concepts," in H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, pp. 38–76. Minneapolis: University of Minnesota Press, 1956.

Cohen, M. R., and E. Nagel. *Introduction to Logic and Scientific Method*. New York: Harcourt, Brace, 1934.

Craig, William. "On Axiomatizability within a System," *Journal of Symbolic Logic*, 18:30–32 (1953).

Craig, William. "Replacement of Auxiliary Expressions," *Philosophical Review*, 65:38–55 (1956).

Cramér, Harald. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

Estes, W. K., S. Koch, K. MacCorquodale, P. E. Meehl, C. G. Mueller, W. S. Schoenfeld, and W. S. Verplanck. *Modern Learning Theory*. New York; Appleton-Century-Crofts, 1954.

Feigl, Herbert. "Some Remarks on the Meaning of Scientific Explanation," (A slightly modified version of comments first published in *Psychological Review*, 52(1948)), in Feigl and Sellars (1949), pp. 510–514.

Feigl, Herbert. "Existential Hypotheses," *Philosophy of Science*, 17:35–62 (1950).

Feigl, Herbert. "Logical Reconstruction, Realism, and Pure Semiotic," *Philosophy of Science*, 17:186–195 (1950). Referred to in this essay as (1950a).

Feigl, Herbert. "Principles and Problems of Theory Construction in Psychology," in W. Dennis (ed.), *Current Trends in Psychological Theory*, pp. 179–213. Pittsburgh: University of Pittsburgh Press, 1951.

Feigl, Herbert, and May Brodbeck (eds.). *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts, 1953.

Feigl, Herbert, and Wilfred Sellars (eds.). *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts, 1949.

Goodman, Nelson. "On Likeness of Meaning," *Analysis*, 10:1–7 (1949). Reprinted in a revised form in Linsky (1952).

Goodman, Nelson. *The Structure of Appearance*. Cambridge, Mass.: Harvard University Press, 1951.

Goodman, Nelson. *Fact, Fiction, and Forecast*. Cambridge, Mass.: Harvard University Press, 1955.

Grice, H. P., and P. F. Strawson. "In Defense of a Dogma," *Philosophical Review*, 65:141-158 (1956).

Hempel, Carl G. "Geometry and Empirical Science," *American Mathematical Monthly*, 52:7-17 (1945). Reprinted in Feigl and Sellars (1949), in Wiener (1953), and in James R. Newman (ed.), *The World of Mathematics*. New York: Simon and Schuster, 1956.

Hempel, Carl G. "A Note on Semantic Realism," *Philosophy of Science*, 17: 169-173 (1950).

Hempel, Carl G. "General System Theory and the Unity of Science," *Human Biology*, 23:313-322 (1951).

Hempel, Carl G. *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press, 1952.

Hempel, Carl G. "Implications of Carnap's Work for the Philosophy of Science," in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. La Salle, Ill.: Open Court Publishing Co., 1963.

Hempel, Carl G. "The Concept of Cognitive Significance: A Reconsideration," *Proceedings of the American Academy of Arts and Sciences*, 80 (No. 1):61-77 (1951).

Hempel, Carl G., and Paul Oppenheim. "Studies in the Logic of Explanation," *Philosophy of Science*, 15:135-175 (1948). Reprinted in the present volume.

Hermes, H. "Eine Axiomatisierung der allgemeinen Mechanik," *Forschungen zur Logik und Grundlegung der exakten Wissenschaften*. Neue Folge, Heft 3. Leipzig, 1938.

Hull, C. L. "The Problem of Intervening Variables in Molar Behavior Theory," *Psychological Review*, 50:273-291 (1943). Reprinted in Marx (1951).

Hull, C. L., C. I. Hovland, R. T. Ross, M. Hall, D. T. Perkins, and F. B. Fitch. *Mathematico-Deductive Theory of Rote Learning*. New Haven: Yale University Press, 1940.

Hutten, Ernest H. *The Language of Modern Physics*: An *Introduction to the Philosophy of Science*. London and New York: Macmillan, 1956.

Kemeny, John G. Review of Carnap (1950). *The Journal of Symbolic Logic*, 16:205-207 (1951).

Kemeny, John G. "Extension of the Methods of Inductive Logic," *Philosophical Studies*, 3:38-42 (1952).

Kemeny, John G., and Paul Oppenheim. "On Reduction," *Philosophical Studies*, 7:6-19 (1956).

Lindzey, Gardner. "Hypothetical Constructs, Conventional Constructs, and the Use of Physiological Data in Psychological Theory," *Psychiatry*, 16:27-33 (1953).

Linsky, Leonard (ed.). *Semantics and the Philosophy of Language*. Urbana, Ill.: University of Illinois Press, 1952.

MacCorquodale, K., and P. Meehl. "On a Distinction between Hypothetical Constructs and Intervening Variables," *Psychological Review*, 55:95-107 (1948). Reprinted in Feigl and Brodbeck (1953) and, with omissions, in Marx (1951).

MacCorquodale, K., and P. Meehl. "Edward C. Tolman," in Estes *et al.* (1954), 177-266.

Margenau, Henry. *The Nature of Physical Reality*. New York: McGraw-Hill Book Co., 1950.

Martin, R. M. "On 'Analytic'," *Philosophical Studies*, 3:42-47 (1952).

Marx, Melvin H. (ed.). *Psychological Theory*. New York: Macmillan, 1951.

Mates, Benson. "Analytic Sentences," *Philosophical Review*, 60:525-534 (1951).

McKinsey, J. C. C., A. C. Sugar, and P. Suppes. "Axiomatic Foundations of Classical Particle Mechanics," *Journal of Rational Mechanics and Analysis*, 2:253-272 (1953).

McKinsey, J. C. C., and P. Suppes. "Transformations of Systems of Classical Particle Mechanics," *Journal of Rational Mechanics and Analysis*, 2:273-289 (1953).

Mises, R. von. *Positivism: A Study in Human Understanding*. Cambridge, Mass.: Harvard University Press, 1951.

Nagel, Ernest. *Principles of the Theory of Probability*. Chicago: University of Chicago Press, 1939.

Nagel, Ernest. "The Meaning of Reduction in the Natural Sciences," in Robert C. Stauffer (ed.), *Science and Civilization*. Madison, Wis.: University of Wisconsin Press, 1949. Reprinted in Wiener (1953).

Nagel, Ernest. "Mechanistic Explanation and Organismic Biology," *Philosophy and Phenomenological Research*. 11:327-338 (1951).

Neumann, John von, and Oskar Morgenstern. *Theory of Games and Economic Behavior*, 2d ed. Princeton: Princeton University Press, 1947.

Northrop, F. S. C. *The Logic of the Sciences and the Humanities*. New York: Macmillan, 1947.

Pap, Arthur. "Reduction Sentences and Open Concepts," *Methodos*, 5:3-28 (1953).

Pap, Arthur. *Analytische Erkenntnistheorie*. Wien: J. Springer, 1955.

Popper, Karl. *Logik der Forschung*. Wien: J. Springer, 1935.

Popper, Karl. *The Open Society and its Enemies*. London: G. Routledge & Sons, 1945.

Popper, Karl. "Three Views Concerning Human Knowledge," in H. D. Lewis (ed.), *Contemporary British Philosophy: Personal Statements*. New York: Macmillan, 1956.

Quine, W. V. "Two Dogmas of Empiricism," *Philosophical Review*, 60:20-43 (1951). Reprinted in Quine (1953).

Quine, W. V. *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press, 1953.

Ramsey, Frank Plumpton. *The Foundation of Mathematics and other Logical Essays*. London: Kegan Paul, and New York: Harcourt Brace, 1931.

Reichenbach, Hans. *Axiomatik der relativistischen Raum-Zeit-Lehre*. Braunschweig: F. Vieweg & Sohn, 1924.

Reichenbach, Hans. *Philosophie der Raum-Zeit-Lehre*. Berlin: W. de Gruyter & Co., 1928.

Reichenbach, Hans. *Philosophic Foundations of Quantum Mechanics*. Berkeley and Los Angeles: University of California Press, 1944.

Reichenbach, Hans. *The Theory of Probability*. Berkeley: University of California Press, 1949.

Reichenbach, Hans. *The Rise of Scientific Philosophy*. Berkeley and Los Angeles: University of California Press, 1951.

Reichenbach, Hans. *Nomological Statements and Admissible Operations*. Amsterdam: North Holland Publishing Co., 1954.

Rozeboom, William W. "Mediation Variables in Scientific Theory," *Psychological Review*, 63:249-264 (1956).

Rubin, H., and P. Suppes. *Transformations of Systems of Relativistic Particle Mechanics*. Technical Report No. 2. Prepared under contract for Office of Naval Research. Stanford University, Stanford, 1953.

Russell, Bertrand. *Introduction to Mathematical Philosophy*. London and New York: Macmillan, 1919.

Russell, Bertrand. *Mysticism and Logic*. New York: W. W. Norton & Co., 1929.

Ryle, Gilbert. *The Concept of Mind*. London: Hutchinson's University Library, 1949.

Schlick, M. "Die Kausalitaet in der gegenwaertigen Physik," *Die Naturwissenschaften*, 19:145-162 (1931).

Skinner, B. F. *Science and Human Behavior*. New York: Macmillan, 1953.

Smart, J. J. C. "The Reality of Theoretical Entities," *Australasisan Journal of Philosophy*, 34:1-12 (1956).

Spence, Kenneth W. "The Nature of Theory Construction in Contemporary Psychology," *Psychological Review*, 51:47-68 (1944). Reprinted in Marx (1951).

Tarski, Alfred. "Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe," *Erkenntnis* 5:80-100 (1935). English translation in Tarski (1956).

Tarski, Alfred. *Introduction to Logic and to the Methodology of Deductive Sciences*. New York: Oxford University Press, 1941.

Tarski, Alfred, "The Semantic Conception of Truth," *Philosophy and Phenomenological Research*, 4:341-375 (1944). Reprinted in Feigl and Sellars (1949) and in Linsky (1952).

Tarski, Alfred, *Logic, Semantics, Metamathematics*. Tr. by J. H. Woodger. Oxford: The Clarendon Press, 1956.

Tolman, E. C. "Operational Behaviorism and Current Trends in Psychology," *Proceedings of the 25th Anniversary Celebration of the Inauguration of Graduate Study*, Los Angeles, 1936, pp. 89-103. Reprinted in Marx (1951).

Tolman, E. C., B. F. Ritchie, and D. Kalish. "Studies in Spatial Learning. I. Orientation and the Short-Cut." *Journal of Experimental Psychology*, 36: 13-24 (1946).

Toulmin, Stephen. *The Philosophy of Science*. London: Hutchinson's University Library, 1953.

Verplanck, W. S. "Burrhus F. Skinner," in Estes *et al.* (1954), 267-316.

Wald, A. *On the Principles of Statistical Inference*. Notre Dame: University of Notre Dame Press, 1942.

Walker, A. G. "Foundations of Relativity: Parts I and II," *Proceedings of the Royal Society of Edinburgh*, 62:319-335 (1943-1949).

Walsh, W. H. *An Introduction to Philosophy of History*. London: Hutchinson's University Library, 1951.

Wang, Hao. "Notes on the Analytic-Synthetic Distinction," *Theoria*, 21:158-178 (1955).

White, Morton G. "The Analytic and the Synthetic: An Untenable Dualism," in S. Hook (ed.), *John Dewey: Philosopher of Science and of Freedom*. New York: Dial Press, 1950. Reprinted in Linsky (1952).

White, Morton G. *Toward Reunion in Philosophy*. Cambridge, Mass.: Harvard University Press, 1956.

Wiener, Philip P. (ed.). *Readings in Philosophy of Science*. New York: Scribner, 1953.

Woodger, J. H. *The Axiomatic Method in Biology*. Cambridge: Cambridge University Press, 1937.

Woodger, J. H. *The Technique of Theory Construction*. Chicago: University of Chicago Press, 1939.

Woodger, J. H. *Biology and Language*. Cambridge: Cambridge University Press, 1952.