Carl G. Hempel

ASPECTS of SCIENTIFIC EXPLANATION

And Other *Essays in the Philosophy of Science*

THE FREE PRESS, *NEW YORK*

COLLIER-MACMILLAN LIMITED, *LONDON*

I. STUDIES IN THE LOGIC OF CONFIRMATION

1. OBJECTIVE OF THE STUDY[1]

THE DEFINING characteristic of an empirical statement is its capability of being tested by a confrontation with experiential findings, *i.e.* with the results of suitable experiments or focused observations. This feature distinguishes statements which have empirical content both from the statements of the formal sciences, logic and mathematics, which require no experiential test for their validation, and from the formulations of transempirical metaphysics, which admit of none.

The testability here referred to has to be understood in the comprehensive sense of "testability in principle" or "theoretical testability"; many empirical statements, for practical reasons, cannot actually be tested now. To call a statement of this kind testable in principle means that it is possible to state just what experiential findings, if they were actually obtained, would constitute favorable evidence

1. The present analysis of confirmation was to a large extent suggested and stimulated by a cooperative study of certain more general problems which were raised by Dr. Paul Oppenheim, and which I have been investigating with him for several years. These problems concern the form and the function of scientific laws and the comparative methodology of the different branches of empirical science.

In my study of the logical aspects of confirmation, I have benefited greatly by discussions with Professor R. Carnap, Professor A. Tarski, and particularly Dr. Nelson Goodman, to whom I am indebted for several valuable suggestions which will be indicated subsequently.

A detailed exposition of the more technical aspects of the analysis of confirmation presented in this essay is included in my article 'A Purely Syntactical Definition of Confirmation,' *The Journal of Symbolic Logic*, vol. 8 (1943).

This article is reprinted, with some changes, by kind permission of the editor of *Mind*, where it appeared in volume 54, pp. 1-26 and 97-121 (1945).

for it, and what findings or "data," as we shall say for brevity, would constitute unfavorable evidence; in other words, a statement is called testable in principle if it is possible to describe the kind of data which would confirm or disconfirm it.

The concepts of confirmation and of disconfirmation as here understood are clearly more comprehensive than those of conclusive verification and falsification. Thus, *e.g.*, no finite amount of experiential evidence can conclusively verify a hypothesis expressing a general law such as the law of gravitation, which covers an infinity of potential instances, many of which belong either to the as yet inaccessible future or to the irretrievable past; but a finite set of relevant data may well be "in accord with" the hypothesis and thus constitute confirming evidence for it. Similarly, an existential hypothesis, asserting, say, the existence of an as yet unknown chemical element with certain specified characteristics, cannot be conclusively proved false by a finite amount of evidence which fails to "bear out" the hypothesis; but such unfavorable data may, under certain conditions, be considered as weakening the hypothesis in question, or as constituting disconfirming evidence for it.[2]

While, in the practice of scientific research, judgments as to the confirming or disconfirming character of experiential data obtained in the test of a hypothesis are often made without hesitation and with a wide consensus of opinion, it can hardly be said that these judgments are based on an explicit theory providing general criteria of confirmation and of disconfirmation. In this respect, the situation is comparable to the manner in which deductive inferences are carried out in the practice of scientific research: this, too, is often done without reference to an explicitly stated system of rules of logical inference. But while criteria of valid deduction can be and have been supplied by formal logic, no satisfactory theory providing general criteria of confirmation and disconfirmation appears to be available so far.

In the present essay, an attempt will be made to provide the elements of a theory of this kind. After a brief survey of the significance and the present status of the problem, I propose to present a detailed critical analysis of some common conceptions of confirmation and disconfirmation and then to construct explicit definitions for these concepts and to formulate some basic principles of what might be called the logic of confirmation.

2. SIGNIFICANCE AND PRESENT STATUS OF THE PROBLEM

The establishment of a general theory of confirmation may well be regarded as one of the most urgent desiderata of the present methodology of empirical science. Indeed, it seems that a precise analysis of the concept of confirmation is

2. This point as well as the possibility of conclusive verification and conclusive falsification will be discussed in some detail in section 10 of the present paper.

a necessary condition for an adequate solution of various fundamental problems concerning the logical structure of scientific procedure. Let us briefly survey the most outstanding of these problems.

(*a*) In the discussion of scientific method, the concept of relevant evidence plays an important part. And while certain inductivist accounts of scientific procedure seem to assume that relevant evidence, or relevant data, can be collected in the context of an inquiry prior to the formulation of any hypothesis, it should be clear upon brief reflection that relevance is a relative concept; experiential data can be said to be relevant or irrelevant only with respect to a given hypothesis; and it is the hypothesis which determines what kind of data or evidence are relevant for it. Indeed, an empirical finding is relevant for a hypothesis if and only if it constitutes either favorable or unfavorable evidence for it; in other words, if it either confirms or disconfirms the hypothesis. Thus, a precise definition of relevance presupposes an analysis of confirmation and disconfirmation.

(*b*) A closely related concept is that of instance of a hypothesis. The so-called method of inductive inference is usually presented as proceeding from specific cases to a general hypothesis of which each of the special cases is an "instance" in the sense that it conforms to the general hypothesis in question, and thus constitutes confirming evidence for it.

Thus, any discussion of induction which refers to the establishment of general hypotheses on the strength of particular instances is fraught with all those logical difficulties—soon to be expounded—which beset the concept of confirmation. A precise analysis of this concept is, therefore, a necessary condition for a clear statement of the issues involved in the problem complex of induction and of the ideas suggested for their solution—no matter what their theoretical merits or demerits may be.

(*c*) Another issue customarily connected with the study of scientific method is the quest for "rules of induction." Generally speaking, such rules would enable us to infer, from a given set of data, that hypothesis or generalization which accounts best for all the particular data in the given set. But this construal of the problem involves a misconception: While the process of invention by which scientific discoveries are made is as a rule *psychologically guided and stimulated* by antecedent knowledge of specific facts, its results are *not logically determined* by them; the way in which scientific hypotheses or theories are discovered cannot be mirrored in a set of general rules of inductive inference.[3] One of the crucial

3. See the lucid presentation of this point in Karl Popper's *Logik der Forschung* (Wien, 1935), esp. sections 1, 2, 3, and 25, 26, 27; *cf.* also Albert Einstein's remarks in his lecture *On the Method of Theoretical Physics* (Oxford, 1933), 11, 12. Also of interest in this context is the critical discussion of induction by H. Feigl in "The Logical Character of the Principle of Induction," *Philosophy of Science*, vol. 1 (1934).

considerations which lead to this conclusion is the following: Take a scientific theory such as the atomic theory of matter. The evidence on which it rests may be described in terms referring to directly observable phenomena, namely to certain macroscopic aspects of the various experimental and observational data which are relevant to the theory. On the other hand, the theory itself contains a large number of highly abstract, nonobservational terms such as 'atom', 'electron', 'nucleus', 'dissociation', 'valence' and others, none of which figures in the description of the observational data. An adequate rule of induction would therefore have to provide, for this and for every other conceivable case, mechanically applicable criteria determining unambiguously, and without any reliance on the inventiveness or additional scientific knowledge of its user, all those new abstract concepts which need to be created for the formulation of the theory that will account for the given evidence. Clearly, this requirement cannot be satisfied by any set of rules, however ingeniously devised; there can be no general rules of induction in the above sense; the demand for them rests on a confusion of logical and psychological issues. What determines the soundness of a hypothesis is not the way it is arrived at (it may even have been suggested by a dream or a hallucination), but the way it stands up when tested, *i.e.* when confronted with relevant observational data. Accordingly, the quest for rules of induction in the original sense of canons of scientific discovery has to be replaced, in the logic of science, by the quest for general objective criteria determining (A) whether, and—if possible—even (B) to what degree, a hypothesis *H* may be said to be corroborated by a given body of evidence *E*. This approach differs essentially from the inductivist conception of the problem in that it presupposes not only *E*, but also *H* as given, and then seeks to determine a certain logical relationship between them. The two parts of this latter problem can be related in somewhat more precise terms as follows:

(A) To give precise definitions of the two nonquantitative relational concepts of confirmation and of disconfirmation; *i.e.* to define the meaning of the phrases '*E* confirms *H*' and '*E* disconfirms *H*'. (When *E* neither confirms nor disconfirms *H*, we shall say that *E* is neutral, or irrelevant, with respect to *H*.)

(B) (1) To lay down criteria defining a metrical concept "degree of confirmation of *H* with respect to *E*," whose values are real numbers; or, failing this,

(2) To lay down criteria defining two relational concepts, "more highly confirmed than "and "equally well confirmed as," which make possible a nonmetrical comparison of hypotheses (each with a body of evidence assigned to it) with respect to the extent of their confirmation.

Interestingly, problem B has received much more attention in methodological research than problem A; in particular, the various theories of the so-called probability of hypotheses may be regarded as concerning this problem complex; we have here adopted[4] the more neutral term' degree of confirmation' instead of 'probability' because the latter is used in science in a definite technical sense involving reference to the relative frequency of the occurrence of a given event in a sequence, and it is at least an open question whether the degree of confirmation of a hypothesis can generally be defined as a probability in this statistical sense.

The theories dealing with the probability of hypotheses fall into two main groups: the "logical" theories construe probability as a logical relation between sentences (or propositions; it is not always clear which is meant);[5] the "statistical" theories interpret the probability of a hypothesis in substance as the limit of the relative frequency of its confirming instances among all relevant cases.[6] Now it is a remarkable fact that none of the theories of the first type which have been developed so far provides an explicit general definition of the probability (or degree of confirmation) of a hypothesis *H* with respect to a body of evidence *E*; they all limit themselves essentially to the construction of an uninterpreted postulational system of logical probability.[7] For this reason, these theories fail to provide a complete solution of problem B. The statistical approach, on the other hand, would, if successful, provide an explicit numerical definition of the degree of confirmation of a hypothesis; this definition would be formulated in terms of the numbers of confirming and disconfirming instances for *H* which constitute the body of evidence *E*. Thus, a necessary condition for an adequate interpretation of degrees of confirmation as statistical probabilities is the establishment of precise criteria of confirmation and disconfirmation; in other words, the solution of problem A.

4. Following R. Carnap's use in "Testability and Meaning," *Philosophy of Science*, Vols. 3 (1936) and 4 (1937); esp. section 3 (in Vol. 3).

5. This group includes the work of such writers as Janina Hosiasson-Lindenbaum [*cf.* for instance, her article "Induction et analogie: Comparaison de leur fondement,", *Mind*, Vol. 50 (1941)], H. Jeffreys, J. M. Keynes, B.O. Koopman, J. Nicod, St. Mazurkiewicz, and F. Waismann. For a brief discussion of this conception of probability, see Ernest Nagel, *Principles of the Theory of Probability* (International Encyclopedia of United Science, Vol. I, no. 6, Chicago, 1939), esp. sections 6 and 8.

6. The chief proponent of this view is Hans Reichenbach; *cf.* especially "Ueber Induktion und Wahrscheinlichkeit," *Erkenntnis*, vol. 5 (1935), and *Experience and Prediction* (Chicago, 1938), Chap. V.

7. (Added in 1964.) Since this article was written, R. Carnap has developed a theory of inductive logic which, for formalized languages of certain types, makes it possible explicitly to define—without use of the qualitative notion of confirming instance—a quantitative concept of degree of confirmation which has the formal characteristics of a probability; Carnap refers to it as inductive, or logical, probability. For details, see especially R. Carnap, "On Inductive Logic," *Philosophy of Science*, vol. 12 (1945); *Logical Foundations of Probability* (Chicago, 1950; 2nd ed., 1962); *The Continuum of Inductive Methods* (Chicago, 1952); "The Aim of Inductive Logic" *in* E. Nagel, P. Suppes, and A. Tarski, eds., *Logic, Methodology, and Philosophy of Science. Proceedings of the 1960 International Congress* (Stanford, 1962).

However, despite their great ingenuity and suggestiveness, the attempts which have been made so far to formulate a precise statistical definition of the degree of confirmation of a hypothesis seem open to certain objections,[8] and several authors[9] have expressed doubts as to the possibility of defining the degree of confirmation of a hypothesis as a metrical magnitude, though some of them consider it as possible, under certain conditions, to solve at least the less exacting problem B (2), *i.e.* to establish standards of nonmetrical comparison between hypotheses with respect to the extent of their confirmation. An adequate comparison of this kind might have to take into account a variety of different factors;[10] but again the numbers of the confirming and of the disconfirming instances which the given evidence includes will be among the most important of those factors.

Thus, of the two problems, A and B, the former appears to be the more basic one, first, because it does not presuppose the possibility of defining numerical degrees of confirmation or of comparing different hypotheses as to the extent of their confirmation; and second because our considerations indicate that any attempt to solve problem B—unless it is to remain in the stage of an axiomatized system without interpretation—is likely to require a precise definition of the concepts of confirming and disconfirming instance of a hypothesis before it can proceed to define numerical degrees of confirmation, or to lay down nonmetrical standards of comparison.

(*d*) It is now clear that an analysis of confirmation is of fundamental importance also for the study of a central problem of epistemology, namely, the elaboration of standards of rational belief or of criteria of warranted assertibility. In the methodology of empirical science this problem is usually phrased as concerning the rules governing the test and the subsequent acceptance or rejection of empirical hypotheses on the basis of experimental or observational findings, while in its epistemological version the issue is often formulated as concerning the validation of beliefs by reference to perceptions, sense data, or the like. But no matter how the final empirical evidence is construed and in what terms it is accordingly expressed, the theoretical problem remains the same: to

8. *Cf.* Karl Popper, *Logik der Forschung* (Wien, 1935), section 80; Ernest Nagel, *l.c.*, section 8, and "Probability and the Theory of Knowledge," *Philosophy of Science*, vol. 6 (1939); C. G. Hempel, "Le problème de la vérité," *Theoria* (Göteborg), vol. 3 (1937), section 5, and "On the Logical Form of Probability Statements," *Erkenntnis*, Vol. 7 (1937-38), esp. section 5. *Cf.* also Morton White, "Probability and Confirmation," *The Journal of Philosophy*, Vol. 36 (1939).

9. See, for example, J. M. Keynes, *A Treatise on Probability* (London, 1929), esp. Chap. III; Ernest Nagel, *Principles of the Theory of Probability*, esp. p. 70; compare also the somewhat less definitely skeptical statement by Carnap, *l.c.* (note 4) section 3, p. 427.

10. See especially the survey of such factors given by Ernest Nagel in *Principles of the Theory of Probability*, pp. 66-73.

characterize, in precise and general terms, the conditions under which a body of evidence can be said to confirm, or to disconfirm, a hypothesis of empirical character; and that is again our problem A.

(*e*) The same problem arises when one attempts to give a precise statement of the empiricist and operationalist criteria for the empirical meaningfulness of a sentence; these criteria, as is well known, are formulated by reference to the theoretical testability of the sentence by means of experiential evidence,[11] and the concept of theoretical testability, as was pointed out earlier, is closely related to the concepts of confirmation and disconfirmation.[12]

Considering the great importance of the concept of confirmation, [illegible] surprising that no systematic theory of the nonquantitative relation of co[illegible]ation seems to have been developed so far. Perhaps this fact reflects th[illegible] assumption that the concepts of confirmation and of disconfirmation have sufficiently clear meaning to make explicit definitions unnecessary or at least comparatively trivial. And indeed, as will be shown below, there are certain features which are rather generally associated with the intuitive notion of confirming evidence, and which, at first, seem well suited to serve as defining characteristics of confirmation. Closer examination will reveal the definitions thus obtainable to be seriously deficient and will make it clear that an adequate definition of confirmation involves considerable difficulties.

Now the very existence of such difficulties suggests the question whether the problem we are considering does not rest on a false assumption: Perhaps there are no objective criteria of confirmation; perhaps the decision as to whether a given hypothesis is acceptable in the light of a given body of evidence is no more subject to rational, objective rules than is the process of inventing a scientific hypothesis or theory; perhaps, in the last analysis, it is a "sense of evidence," or a feeling of plausibility in view of the relevant data, which ultimately decides whether a hypothesis is scientifically acceptable.[13] This view is comparable to the opinion that the validity of a mathematical proof or of a logical argument has to be judged ultimately by reference to a feeling of soundness or convincingness; and both theses have to be rejected on analogous grounds: they involve a con-

11. *Cf.*, for example, A. J. Ayer, *Language, Truth and Logic* (London and New York, 1936), Ch. I; R. Carnap, "Testability and Meaning," sections 1, 2, 3; H. Feigl, "Logical Empiricism" (in *Twentieth Century Philosophy*, ed. by Dagobert D. Runes, New York, 1943); P.W. Bridgman, *The Logic of Modern Physics* (New York, 1928).

12. It should be noted, however, that in his essay "Testability and Meaning," R. Carnap has constructed definitions of testability and confirmability which avoid reference to the concept of confirming and of disconfirming evidence; in fact, no proposal for the definition of these latter concepts is made in that study.

13. A view of this kind has been expressed, for example, by M. Mandelbaum in "Causal Analyses in History," *Journal of the History of Ideas*, Vol. 3 (1942); *cf.* esp. pp. 46-47.

fusion of logical and psychological considerations. Clearly, the occurrence or non-occurrence of a feeling of conviction upon the presentation of grounds for an assertion is a subjective matter which varies from person to person, and with the same person in the course of time; it is often deceptive and can certainly serve neither as a necessary nor as a sufficient condition for the soundness of the given assertion.[14] A rational reconstruction of the standards of scientific validation cannot, therefore, involve reference to a sense of evidence; it has to be based on objective criteria. In fact, it seems reasonable to require that the criteria of empirical confirmation, besides being objective in character, should contain no reference to the specific subject matter of the hypothesis or of the evidence in question; it ought to be possible, one feels, to set up purely formal criteria of confirmation in a manner similar to that in which deductive logic provides purely formal criteria for the validity of deductive inference.

With this goal in mind, we now turn to a study of the nonquantitative concept of confirmation. We shall begin by examining some current conceptions of confirmation and exhibiting their logical and methodological inadequacies; in the course of this analysis, we shall develop a set of conditions for the adequacy of any proposed definition of confirmation; and finally, we shall construct a definition of confirmation which satisfies those general standards of adequacy.

3. NICOD'S CRITERION OF CONFIRMATION AND ITS SHORTCOMINGS

We consider first a conception of confirmation which underlies many recent studies of induction and of scientific method. A very explicit statement of this conception has been given by Jean Nicod in the following passage: "Consider the formula or the law: *A entails B*. How can a particular proposition, or more briefly, a fact, affect its probability? If this fact consists of the presence of B in a case of A, it is favorable to the law '*A entails B*'; on the contrary, if it consists of the absence of B in a case of A, it is unfavorable to this law. It is conceivable that we have here the only two direct modes in which a fact can influence the probability of a law. . . . Thus, the entire influence of particular truths or facts on the probability of universal propositions or laws would operate by means of these two elementary relations which we shall call *confirmation* and *invalidation*."[15] Note that the applicability of this criterion is restricted to hypotheses of the form '*A entails B*'. Any hypothesis *H* of this kind may be expressed in the notation

14. See Popper's statement, *l.c.*, section 8.

15. Jean Nicod, *Foundations of Geometry and Induction* (transl. by P. P. Wiener), London, 1930; 219; *cf.* also R. M. Eaton's discussion of "Confirmation and Infirmation," which is based on Nicod's views; it is included in Chap. III of his *General Logic* (New York, 1931).

of symbolic logic[16] by means of a universal conditional sentence, such as, in the simplest case,

$$(x)[P(x) \supset Q(x)]$$

i.e. 'For any object *x*: if *x* is a *P*, then *x* is a Q,' or also 'Occurrence of the quality *P* entails occurrence of the quality Q.' According to the above criterion this hypothesis is confirmed by an object *a* if *a* is *P* and Q; and the hypothesis is disconfirmed by *a* if *a* is *P*, but not Q.[17] In other words, an object confirms a universal conditional hypothesis if and only if it satisfies both the antecedent (here: '$P(x)$') and the consequent (here: '$Q(x)$') of the conditional; it disconfirms the hypothesis if and only if it satisfies the antecedent, but not the consequent of the conditional; and (we add this to Nicod's statement) it is neutral, or irrelevant, with respect to the hypothesis if it does not satisfy the antecedent.

This criterion can readily be extended so as to be applicable also to universal conditionals containing more than one quantifier, such as 'Twins always resemble each other', or, in symbolic notation, '$(x)(y)(\text{Twins}(x, y) \supset \text{Rsbl}(x, y))$'. In these cases, a confirming instance consists of an ordered couple, or triple, etc., of objects satisfying the antecedent and the consequent of the conditional. (In the case of the last illustration, any two persons who are twins and resemble each other would confirm the hypothesis; twins who do not resemble each other would disconfirm it; and any two persons not twins—no matter whether they resemble each other or not—would constitute irrelevant evidence.)

We shall refer to this criterion as Nicod's criterion.[18] It states explicitly what is perhaps the most common tacit interpretation of the concept of confirmation. While seemingly quite adequate, it suffers from serious shortcomings, as will now be shown.

(*a*) First, the applicability of this criterion is restricted to hypotheses of universal conditional form; it provides no standards of confirmation for existential hypotheses (such as 'There exists organic life on other stars', or 'Poliomyelitis is caused by some virus') or for hypotheses whose explicit formulation calls for the use of both universal and existential quantifiers (such as 'Every human

16. In this essay, only the most elementary devices of this notation are used; the symbolism is essentially that of *Principia Mathematica*, except that parentheses are used instead of dots, and that existential quantification is symbolized by '(E)' instead of by the inverted 'E.'

17. (Added in 1964). More precisely we would have to say, in Nicod's parlance, that the hypothesis is confirmed by the *proposition* that *a* is both *P* and Q, and is disconfirmed by the *proposition* that *a* is *P* but not Q.

18. This term is chosen for convenience, and in view of the above explicit formulation given by Nicod; it is not, of course, intended to imply that this conception of confirmation originated with Nicod.

being dies some finite number of years after his birth', or the psychological hypothesis, ' You can fool all of the people some of the time and some of the people all of the time, but you cannot fool all of the people all of the time', which may be symbolized by '$(x)(Et)\mathrm{Fl}(x, t) \cdot (Ex)(t)\mathrm{Fl}(x, t) \cdot \sim (x)(t)\mathrm{Fl}(x, t)$', (where '$\mathrm{Fl}(x, t)$' stands for 'You can fool person x at time t'). We note, therefore, the desideratum of establishing a criterion of confirmation which is applicable to hypotheses of *any* form.[19]

(*b*) We now turn to a second shortcoming of Nicod's criterion. Consider the two sentences

$$S_1: \text{'}(x)[\mathrm{Raven}(x) \supset \mathrm{Black}(x)]\text{'};$$

$$S_2: \text{'}(x)[\sim\mathrm{Black}(x) \supset \sim \mathrm{Raven}(x)]\text{'}$$

(*i.e.* 'All ravens are black' and 'Whatever is not black is not a raven'), and let *a*, *b*, *c*, *d* be four objects such that *a* is a raven and black, *b* a raven but not black, *c* not a raven but black, and *d* neither a raven nor black. Then according to Nicod's criterion, *a* would confirm S_1, but be neutral with respect to S_2; *b* would disconfirm both S_1 and S_2; *c* would be neutral with respect to both S_1 and S_2, and *d* would confirm S_2, but be neutral with respect to S_1.

But S_1 and S_2 are logically equivalent; they have the same content, they are different formulations of the same hypothesis. And yet, by Nicod's criterion, either of the objects *a* and *d* would be confirming for one of the two sentences, but neutral with respect to the other. This means that Nicod's criterion makes confirmation depend not only on the content of the hypothesis, but also on its formulation.[20]

One remarkable consequence of this situation is that every hypothesis to which the criterion is applicable—*i.e.* every universal conditional—can be stated in a form for which there cannot possibly exist any confirming instances. Thus, *e.g.* the sentence

$$(x)[(\mathrm{Raven}(x) \cdot \sim \mathrm{Black}(x)) \supset (\mathrm{Raven}(x) \cdot \sim \mathrm{Raven}(x)]$$

is readily recognized as equivalent to both S_1 and S_2 above; yet no object whatever can confirm this sentence, *i.e.* satisfy both its antecedent and its consequent;

19. For a rigorous formulation of the problem, it is necessary first to lay down assumptions as to the means of expression and the logical structure of the language in which the hypotheses are supposed to be formulated; the desideratum then calls for a definition of confirmation applicable to any hypothesis which can be expressed in the given language. Generally speaking, the problem becomes increasingly difficult with increasing richness and complexity of the assumed language of science.

20. This difficulty was pointed out, in substance, in my article "Le problème de la vérité," *Theoria* (Göteborg), vol. 3 (1937), esp. p. 222.

for the consequent is contradictory. An analogous transformation is, of course, applicable to any other sentence of universal conditional form.

4. THE EQUIVALENCE CONDITION

The results just obtained call attention to the following condition which an adequately defined concept of confirmation should satisfy, and in the light of which Nicod's criterion has to be rejected as inadequate:

Equivalence condition: Whatever confirms (disconfirms) one of two equivalent sentences, also confirms (disconfirms) the other.

Fulfillment of this condition makes the confirmation of a hypothesis independent of the way in which it is formulated; and no doubt it will be conceded that this is a necessary condition for the adequacy of any proposed criterion of confirmation. Otherwise, the question as to whether certain data confirm a given hypothesis would have to be answered by saying: "That depends on which of the different equivalent formulations of the hypothesis is considered"—which appears absurd. Furthermore—and this is a more important point than an appeal to a feeling of absurdity—an adequate definition of confirmation will have to do justice to the way in which empirical hypotheses function in theoretical scientific contexts such as explanations and predictions; but when hypotheses are used for purposes of explanation or prediction,[21] they serve as premises in a deductive argument whose conclusion is a description of the event to be explained or predicted. The deduction is governed by the principles of formal logic, and according to the latter, a deduction which is valid will remain so if some or all of the premises are replaced by different but equivalent statements; and indeed, a scientist will feel free, in any theoretical reasoning involving certain hypotheses, to use the latter in whichever of their equivalent formulations are most convenient for the development of his conclusions. But if we adopted a concept of confirmation which did not satisfy the equivalence condition, then it would be possible, and indeed necessary, to argue in certain cases that it was sound scientific procedure to base a prediction on a given hypothesis if formulated in a sentence S_1, because a good deal of confirming evidence had been found for S_1; but that it was altogether inadmissible to base the prediction (say, for convenience of deduction) on an equivalent formulation S_2, because no confirming evidence for S_2 was

21. For a more detailed account of the logical structure of scientific explanation and prediction, *cf.* C. G. Hempel, "The Function of General Laws in History," *The Journal of Philosophy*, vol. 39 (1942), esp. sections 2, 3, 4. The characterization, given in that paper as well as in the above text, of explanations and predictions as arguments of a deductive logical structure, embodies an oversimplification: as will be shown in section 7 of the present essay, explanations and predictions often involve "quasi-inductive" steps besides deductive ones. This point, however, does not affect the validity of the above argument.

available. Thus, the equivalence condition has to be regarded as a necessary condition for the adequacy of any definition of confirmation.

5. THE PARADOXES OF CONFIRMATION

Perhaps we seem to have been laboring the obvious in stressing the necessity of satisfying the equivalence condition. This impression is likely to vanish upon consideration of certain consequences which derive from a combination of the equivalence condition with a most natural and plausible assumption concerning a sufficient condition of confirmation.

The essence of the criticism we have leveled so far against Nicod's criterion is that it certainly cannot serve as a necessary condition of confirmation; thus, in the illustration given in the beginning of section 3, object *a* confirms S_1 and should therefore also be considered as confirming S_2, while according to Nicod's criterion it is not. Satisfaction of the latter is therefore not a necessary condition for confirming evidence.

On the other hand, Nicod's criterion might still be considered as stating a particularly obvious and important sufficient condition of confirmation. And indeed, if we restrict ourselves to universal conditional hypotheses in one variable[22] —such as S_1 and S_2 in the above illustration—then it seems perfectly reasonable to qualify an object as confirming such a hypothesis if it satisfies both its antecedent and its consequent. The plausibility of this view will be further corroborated in the course of our subsequent analyses.

Thus, we shall agree that if *a* is both a raven and black, then *a* certainly confirms

22. This restriction is essential: In its general form which applies to universal conditionals in any number of variables, Nicod's criterion cannot even be construed as expressing a sufficient condition of confirmation. This is shown by the following rather surprising example: Consider the hypothesis:

$$S_1: (x)(y)[\sim(R(x,y) \cdot R(y,x)) \supset (R(x,y) \cdot \sim R(y,x))].$$

Let *a*, *b* be two objects such that $R(a,b)$ and $\sim R(b,a)$. Then clearly, the couple (a, b) satisfies both the antecedent and the consequent of the universal conditional S_1; hence, if Nicod's criterion in its general form is accepted as stating a sufficient condition of confirmation, (a, b) constitutes confirming evidence for S_1. But S_1 can be shown to be equivalent to

$$S_2: (x)(y)R(x, y)$$

Now, by hypothesis, we have $\sim R(b, a)$; and this flatly contradicts S_2 and thus S_1. Thus, the couple (a, b), although satisfying both the antecedent and the consequent of the universal conditional S_1, actually constitutes disconfirming evidence of the strongest kind (conclusively disconfirming evidence, as we shall say later) for that sentence. This illustration reveals a striking and—as far as I am aware—hitherto unnoticed weakness of that conception of confirmation which underlies Nicod's criterion. In order to realize the bearing of our illustration upon Nicod's original formulation, let *A* and *B* be $\sim(R(x, y) \cdot R(y, x))$ and $R(x, y) \cdot \sim(R(y, x))$, respectively. Then S_1 asserts that *A* entails *B*, and the couple (a, b) is a case of the presence of *B* in the presence of A; this should, according to Nicod, be favorable to S_1.

S_1: '(x) (Raven$(x) \supset$ Black(x))', and if *d* is neither black nor a raven, *d* certainly confirms S_2: '(x) [$\sim$ Black$(x) \supset \sim$ Raven(x)]'.

Let us now combine this simple stipulation with the equivalence condition. Since S_1 and S_2 are equivalent, *d* is confirming also for S_1; and thus, we have to recognize as confirming for S_1 any object which is neither black nor a raven. Consequently, any red pencil, any green leaf, any yellow cow, etc., becomes confirming evidence for the hypothesis that all ravens are black. This surprising consequence of two very adequate assumptions (the equivalence condition and the above sufficient condition of confirmation) can be further expanded: The sentence S_1 can readily be shown to be equivalent to S_3: '(x) [(Raven(x) v $\sim$ Raven(x)) $\supset$ ($\sim$ Raven(x) v Black(x))]', *i.e.* 'Anything which is or is not a raven is either no raven or black'. According to the above sufficient condition, S_3 is certainly confirmed by any object, say *e*, such that (1) *e* is or is not a raven and, in addition (2) *e* is not a raven or is also black. Since (1) is analytic, these conditions reduce to (2). By virtue of the equivalence condition, we have therefore to consider as confirming for S_1 any object which is either no raven or also black (in other words: any object which is no raven at all, or a black raven).

Of the four objects characterized in section 3, *a*, *c* and *d* would therefore constitute confirming evidence for S_1, while *b* would be disconforming for S_1. This implies that any nonraven represents confirming evidence for the hypothesis that all ravens are black.[23]

We shall refer to these implications of the equivalence condition and of the above sufficient condition of confirmation as the *paradoxes of confirmation.*

How are these paradoxes to be dealt with? Renouncing the equivalence condition would not represent an acceptable solution, as it is shown by the considerations presented in section 4. Nor does it seem possible to dispense with the stipulation that an object satisfying two conditions, C_1 and C_2, should be considered as confirming a general hypothesis to the effect that any object which satisfies C_1 also satisfies C_2.

But the deduction of the above paradoxical results rests on one other assumption which is usually taken for granted, namely, that the meaning of general empirical hypotheses, such as that all ravens are black, or that all sodium salts burn yellow, can be adequately expressed by means of sentences of universal

23. (Added in 1964). The following further "paradoxial" consequence of our two conditions might be noted: Any hypothesis of universal conditional form can be equivalently rewritten as another hypothesis of the same form which, even if true, can have no confirming instances in Nicod's sense at all, since the proposition that a given object satisfies the antecedent and the consequent of the second hypothesis is self-contradictory. For example, '$(x)\,[P(x) \supset Q(x)]$' is equivalent to the sentence '$(x)\,[(P(x) \cdot \sim Q(x)) \supset (P(x) \cdot \sim P(x))]$', whose consequent is true of nothing.

conditional form, such as '(x) [Raven(x) ⊃ Black(x)]' and '(x) (Sod. Salt(x) ⊃ Burn Yellow (x))', etc. Perhaps this customary mode of presentation has to be modified; and perhaps such a modification would automatically remove the paradoxes of confirmation? If this is not so, there seems to be only one alternative left, namely to show that the impression of the paradoxical character of those consequences is due to misunderstanding and can be dispelled, so that no theoretical difficulty remains. We shall now consider these two possibilities in turn: Subsections 5.11 and 5.12 are devoted to a discussion of two different proposals for a modified representation of general hypotheses; in subsection 5.2, we shall discuss the second alternative, *i.e.* the possibility of tracing the impression of paradoxicality back to a misunderstanding.

5.11. It has often been pointed out that while Aristotelian logic, in agreement with prevalent everyday usage, confers existential import upon sentences of the form 'All *P*'s are *Q*'s', a universal conditional sentence, in the sense of modern logic, has no existential import; thus, the sentence

$$\text{`}(x)\ [\text{Mermaid}(x) \supset \text{Green}(x)]\text{'}$$

does not imply the existence of mermaids; it merely asserts that any object either is not a mermaid at all, or a green mermaid; and it is true simply because of the fact that there are no mermaids. General laws and hypotheses in science, however —so it might be argued— are meant to have existential import; and one might attempt to express the latter by supplementing the customary universal conditional by an existential clause. Thus, the hypothesis that all ravens are black would be expressed by means of the sentence S_1: '[(x) (Raven(x) ⊃ Black(x)]·(Ex)Raven(x)'; and the hypothesis that no nonblack things are ravens by S_2: '(x)[∼Black(x) ⊃ ∼ Raven(x)] · (*Ex*) ∼ Black(x)'. Clearly, these sentences are not equivalent, and of the four objects *a*, *b*, *c*, *d* characterized in section 3, part (*b*), only *a* might reasonably be said to confirm S_1, and only *d* to confirm S_2. Yet this method of avoiding the paradoxes of confirmation is open to serious objections:

(*a*) First of all, the representation of every general hypothesis by a conjunction of a universal conditional and an existential sentence would invalidate many logical inferences which are generally accepted as permissible in a theoretical argument. Thus, for example, the assertions that all sodium salts burn yellow, and that whatever does not burn yellow is no sodium salt are logically equivalent according to customary understanding and usage, and their representation by universal conditionals preserves this equivalence; but if existential clauses are added, the two assertions are no longer equivalent, as is illustrated above by the analogous case of S_1 and S_2.

(*b*) Second, the customary formulation of general hypotheses in empirical

science clearly does not contain an existential clause, nor does it, as a rule, even indirectly determine such a clause unambiguously. Thus, consider the hypothesis that if a person after receiving an injection of a certain test substance has a positive skin reaction, he has diphtheria. Should we construe the existential clause here as referring to persons, to persons receiving the injection, or to persons who, upon receiving the injection, show a positive skin reaction? A more or less arbitrary decision has to be made; each of the possible decisions gives a different interpretation to the hypothesis, and none of them seems to be really implied by the latter.

(*c*) Finally, many universal hypotheses cannot be said to imply an existential clause at all. Thus, it may happen that from a certain astrophysical theory a universal hypothesis is deduced concerning the character of the phenomena which would take place under certain specified extreme conditions. A hypothesis of this kind need not (and, as a rule, does not) imply that such extreme conditions ever were or will be realized; it has no existential import. Or consider a biological hypothesis to the effect that whenever man and ape are crossed, the offspring will have such and such characteristics. This is a general hypothesis; it might be contemplated as a mere conjecture, or as a consequence of a broader genetic theory, other implications of which may already have been tested with positive results; but unquestionably the hypothesis does not imply an existential clause asserting that the contemplated kind of cross-breeding referred to will, at some time, actually take place.

5.12. Perhaps the impression of the paradoxical character of the cases discussed in the beginning of section 5 may be said to grow out of the feeling that the hypothesis that all ravens are black is about ravens, and not about nonblack things, nor about all things. The use of an existential clause was one attempt at exhibiting this presumed peculiarity of the hypothesis. The attempt has failed, and if we wish to express the point in question, we shall have to look for a stronger device. The idea suggests itself of representing a general hypothesis by the customary universal conditional, supplemented by the indication of the specific "field of application" of the hypothesis; thus, we might represent the hypothesis that all ravens are black by the sentence '(x) [Raven(x) ⊃ Black(x)]' or any one of its equivalents, plus the indication 'Class of ravens', characterizing the field of application; and we might then require that every confirming instance should belong to the field of application. This procedure would exclude the objects *c* and *d* from those constituting confirming evidence and would thus avoid those undesirable consequences of the existential-clause device which were pointed out in 5.11 (*c*). But apart from this advantage, the second method is open to objections similar to those which apply to the first: (*a*) The way in which general hypotheses are used in science never involves the statement of a field of application; and the choice of the latter in a symbolic formulation of a given hypothesis thus introduces again a considerable measure of arbitrariness. In particular, for a scientific hypothesis to the effect that all *P*'s are Q's, the field of application cannot simply be said to be the class of all *P*'s; for a hypothesis such as that all sodium salts burn yellow finds important application in tests with negative results; *e.g.*, it may be applied to a substance of which it is not known whether it contains sodium salts, nor whether it burns yellow; and if the flame does not turn yellow, the hypothesis serves to establish the absence of sodium salts. The same is true of all other hypotheses used for tests of this type. (*b*) Again, the consistent use of a field of application in the formulation of general hypotheses would involve considerable logical complications, and yet would have no counterpart in the theoretical procedure of science, where hypotheses are subjected to various kinds of logical transformation and inference without any consideration that might be regarded as referring to changes in the fields of application. This method of meeting the paradoxes would therefore amount to dodging the problem by means of an *ad hoc* device which cannot be justified by reference to actual scientific procedure.

5.2 We have examined two alternatives to the customary method of representing general hypotheses by means of universal conditionals; neither of them proved an adequate means of precluding the paradoxes of confirmation. We shall now try to show that what is wrong does not lie in the customary way of construing and representing general hypotheses, but rather in our reliance on a misleading intuition in the matter: The impression of a paradoxical situation is not objectively founded; it is a psychological illusion.

(*a*) One source of misunderstanding is the view, referred to before, that a hypothesis of the simple form 'Every *P* is a Q', such as 'All sodium salts burn yellow', asserts something about a certain limited class of objects only, namely, the class of all *P*'s. This idea involves a confusion of logical and practical considerations: Our interest in the hypothesis may be focussed upon its applicability to that particular class of objects, but the hypothesis nevertheless asserts something about, and indeed imposes restrictions upon, *all* objects (within the logical type of the variable occurring in the hypothesis, which in the case of our last illustration might be the class of all physical objects). Indeed, a hypothesis of the form 'Every *P* is a Q' forbids the occurrence of any objects having the property *P* but lacking the property Q; *i.e.* it restricts all objects whatsoever to the class of those which either lack the property *P* or also have the property Q. Now, every object either belongs to this class or falls outside it, and thus, every object—and not only the P's—either conforms to the hypothesis or violates it; there is no object which is not implicitly referred to by a hypothesis of this type. In particular, every object which either is no sodium salt or burns yellow conforms to, and thus bears out, the hypothesis that all sodium salts burn yellow; every other object violates that hypothesis.

The weakness of the idea under consideration is evidenced also by the observation that the class of objects about which a hypothesis is supposed to assert something is in no way clearly determined, and that it changes with the context, as was shown in 5.12 (*a*).

(*b*) A second important source of the appearance of paradoxicality in certain cases of confirmation is exhibited by the following consideration.

Suppose that in support of the assertion 'All sodium salts burn yellow' somebody were to adduce an experiment in which a piece of pure ice was held into a colorless flame and did not turn the flame yellow. This result would confirm the assertion, 'Whatever does not burn yellow is no sodium salt' and consequently, by virtue of the equivalence condition, it would confirm the original formulation. Why does this impress us as paradoxical? The reason becomes clear when we compare the previous situation with the case where an object whose chemical constitution is as yet unknown to us is held into a flame and fails to turn it yellow, and where subsequent analysis reveals it to contain no sodium salt. This outcome, we should no doubt agree, is what was to be expected on the basis of the hypothesis that all sodium salts burn yellow—no matter in which of its various equivalent formulations it may be expressed; thus, the data here obtained constitute confirming evidence for the hypothesis. Now the only difference between the two situations here considered is that in the first case we are told beforehand the test substance is ice, and we happen to "know anyhow" that ice contains no sodium salt; this has the consequence that the outcome of the flame-color test becomes entirely irrelevant for the confirmation of the hypothesis and thus can yield no new evidence for us. Indeed, if the flame should not turn yellow, the hypothesis requires that the substance contain no sodium salt — and we know beforehand that ice does not; and if the flame should turn yellow, the hypothesis would impose no further restrictions on the substance: hence, either of the possible outcomes of the experiment would be in accord with the hypothesis.

The analysis of this example illustrates a general point: In the seemingly paradoxical cases of confirmation, we are often not actually judging the relation of the given evidence *E* alone to the hypothesis *H* (we fail to observe the methodological fiction, characteristic of every case of confirmation, that we have no relevant evidence for *H* other than that included in *E*); instead, we tacitly introduce a comparison of *H* with a body of evidence which consists of *E* in conjunction with additional information that we happen to have at our disposal; in our illustration, this information includes the knowledge (1) that the substance used in the experiment is ice, and (2) that ice contains no sodium salt. If we assume this additional information as given, then, of course, the outcome of the experiment can add no strength to the hypothesis under consideration. But if we are careful to avoid this tacit reference to additional knowledge (which entirely changes the character of the problem), and if we formulate the question as to the confirming character of the evidence in a manner adequate to the concept of confirmation as used in this paper, we have to ask: Given some object *a* (it happens to be a piece of ice, but this fact is not included in the evidence), and given the fact that *a* does not turn the flame yellow and is no sodium salt: does *a* then constitute confirming evidence for the hypothesis? And now—no matter whether *a* is ice or some other substance—it is clear that the answer has to be in the affirmative; and the paradoxes vanish.

So far, in section (*b*), we have considered mainly that type of paradoxical case which is illustrated by the assertion that any nonblack nonraven constitutes confirming evidence for the hypothesis, 'All ravens are black.' However, the general idea just outlined applies as well to the even more extreme cases exemplified by the assertion that any nonraven as well as any black object confirms the hypothesis in question. Let us illustrate this by reference to the latter case. If the given evidence *E*—*i.e.* in the sense of the required methodological fiction, all data relevant for the hypothesis—consists only of one object which, in addition, is black, then *E* may reasonably be said to support even the hypothesis that all objects are black, and *a fortiori E* supports the weaker assertion that all ravens are black. In this case, again, our factual knowledge that not all objects are black tends to create an impression of paradoxicality which is not justified on logical grounds. Other paradoxical cases of confirmation may be dealt with analogously. Thus it turns out that the paradoxes of confirmation, as formulated above, are due to a misguided intuition in the matter rather than to a logical flaw in the two stipulations from which they were derived.[24,25]

24. The basic idea of section (*b*) in the above analysis is due to Dr. Nelson Goodman, to whom I wish to reiterate my thanks for the help he rendered me, through many discussions, in clarifying my ideas on this point.

25. The considerations presented in section (*b*) above are also influenced by, though not identical in content with, the very illuminating discussion of the paradoxes by the Polish methodologist and logician Janina Hosiasson-Lindenbaum; *cf.* her article "On Confirmation," *The Journal of Symbolic Logic*, vol. 5 (1940), especially section 4. Dr. Hosiasson's attention had been called to the paradoxes by my article "Le problème de la vérité" (*cf.* note 20) and by discussions with me. To my knowledge, hers has so far been the only publication which presents an explicit attempt to solve the problem. Her solution is based on a theory of degrees of confirmation, which is developed in the form of an uninterpreted axiomtaic system, and most of her arguments presuppose that theoretical framework. I have profited, however, by some of Miss Hosiasson's more general observations which proved relevant for the analysis of the paradoxes of the nongraduated or qualitative concept of confirmation which forms the object of the present study.

One point in those of Miss Hosiasson's comments which rest on her theory of degrees of confirmation is of particular interest, and I should like to discuss it briefly. Stated in reference to the raven hypothesis, it consists in the suggestion that the finding of one nonblack object which is no raven, while constituting confirming evidence for the hypothesis, would increase the degree of confirmation of the hypothesis by a smaller amount than the finding of one raven which is black. This is said to be so because the class of all ravens is much less numerous than that of all nonblack objects, so that—to put the idea in suggestive though somewhat misleading terms—the finding of one black raven confirms a larger portion of the total content of the hypothesis than the finding of one nonblack nonraven. In fact, from the basic assumptions of her theory, Miss Hosiasson is able to derive a theorem according to which the above statement about the relative increase in degree of confirmation will hold provided that actually the number of all ravens is small compared with the number of all nonblack objects. But is this last numerical assumption actually warranted in the present case and analogously in all other "paradoxical" cases? The answer depends in part upon the logical structure of the language of science. If a "coordinate language" is used, in which, say, finite space-time regions figure as individuals, then the raven hypothesis assumes some such form as 'Every space-time region which contains a raven contains something black'; and even if the total number of ravens ever to exist is finite, the class of space-time regions containing a raven has the power of the continuum, and so does the class of space-time regions containing something nonblack; thus, for a coordinate language of the type under consideration, the above numerical assumption is not warranted. Now the use of a coordinate language may appear quite artificial in this particular illustration; but it will seem very appropriate in many other contexts, such as, *e.g.*, that of physical field theories. On the other hand, Miss Hosiasson's numerical assumption may well be justified on the basis of a "thing language," in which physical objects of finite size function as individuals. Of course, even on this basis, it remains an empirical question, for every hypothesis of the form 'All *P*'s are *Q*'s', whether actually the class of non-*Q*'s is much more numerous than the class of *P*'s; and in many cases this question will be very difficult to decide.

6. CONFIRMATION CONSTRUED AS A RELATION BETWEEN SENTENCES

Our analysis of Nicod's criterion has so far led to two main results: The rejection of that criterion in view of several deficiencies, and the emergence of the equivalence condition as a necessary condition of adequacy for any proposed definition of confirmation. Another aspect of Nicod's criterion requires consideration now. In our formulation of the criterion, confirmation was construed as a dyadic relation between an object or an ordered set of objects, representing the evidence, and a sentence, representing the hypothesis. This means that confirmation was conceived of as a semantical relation[26] obtaining between certain extra-linguistic objects[27] on one hand and certain sentences on the other. It is possible, however, to construe confirmation in an alternative fashion as a relation between two sentences, one describing the given evidence, the other expressing the hypothesis. Thus, instead of saying that an object *a* which is both a raven and black (or the fact of *a* being both a raven and black) confirms the hypothesis that all ravens are black, we may say that the evidence sentence,

26. For a detailed account of this concept, see C. W. Morris, *Foundations of the Theory of Signs* (Internat. Encyclopedia of Unified Science, vol. I, No. 2, Chicago, 1938) and R. Carnap *Introduction to Semantics* (Cambridge, Mass., 1962), esp. sections 4 and 37.

27. Instead of making the first term of the relation an object or a sequence of objects we might construe it as a state of affairs (or perhaps as a fact, or a proposition, as Nicod puts it), such as that state of affairs which consists in *a* being a black raven, etc.

'*a* is a raven and *a* is black', confirms the hypothesis-sentence (briefly, the hypothesis), 'All ravens are black'. We shall adopt this conception of confirmation as a relation between sentences here for the following reasons: First, the evidence adduced in support or criticism of a scientific hypothesis is always expressed in sentences, which frequently have the character of observation reports; and second, it will prove very fruitful to pursue the parallel, alluded to in section 2 above, between the concepts of confirmation and of logical consequence. And just as in the theory of the consequence relation, *i.e.* in deductive logic, the premises of which a given conclusion is a consequence are construed as sentences rather than as "facts," so we propose to construe the data which confirm a given hypothesis as given in the form of sentences.

The preceding reference to observation reports suggests a certain restriction which might be imposed on evidence sentences. Indeed, the evidence adduced in support of a scientific hypothesis or theory consists, in the last analysis, in data accessible to what is loosely called direct observation, and such data are expressible in the form of "observation reports." In view of this consideration, we shall restrict the evidence sentences which form the domain of the relation of confirmation to sentences of the character of observation reports. In order to give a precise meaning to the concept of observation report, we shall assume that we are given a well-determined "language of science," in terms of which all sentences under consideration, hypotheses as well as evidence sentences, are formulated. We shall further assume that this language contains, among other terms, a clearly delimited "observational vocabulary" which consists of terms designating more or less directly observable attributes of things or events, such as, say, 'black,' 'taller than', 'burning with a yellow light', etc., but no theoretical constructs such as 'aliphatic compound', 'circularly polarized light', 'heavy hydrogen', etc.

We shall now understand by a *hypothesis* any sentence which can be expressed in the assumed language of science, no matter whether it is a generalized sentence, containing quantifiers, or a particular sentence referring only to a finite number of particular objects. An *observation report* will be construed as a finite class (or a conjunction of a finite number) of observation sentences; and an observation sentence as a sentence which either asserts or denies that a given object has a certain observable property (*e.g.* '*a* is a raven', '*d* is not black'), or that a given sequence of objects stand in a certain observable relation (*e.g.* '*a* is between *b* and *c*').

Now the concept of observability itself obviously is relative to the techniques of observation used. What is unobservable to the unaided senses may well be observable by means of suitable devices such as telescopes, microscopes, polariscopes, lie detectors, Gallup polls, etc. If by direct observation we mean such observational procedures as do not make use of auxiliary devices, then such

property terms as 'black', 'hard', 'liquid', 'cool', and such relation terms as 'above', 'between', 'spatially coincident', etc., might be said to refer to directly observable attributes; if observability is construed in a broader sense, so as to allow for the use of certain specified instruments or other devices, the concept of observable attribute becomes more comprehensive. If, in our study of confirmation, we wanted to analyze the manner in which the hypotheses and theories of empirical science are ultimately supported by "evidence of the senses," then we should have to require that observation reports refer exclusively to directly observable attributes. This view was taken, for simplicity and concreteness, in the preceding parts of this section. Actually, however, the general logical characteristics of that relation which obtains between a hypothesis and a group of empirical statements which support it, can be studied in isolation from this restriction to direct observability. All we will assume here is that in the context of the scientific test of a given hypothesis or theory, certain specified techniques of observation have been agreed upon; these determine an observational vocabulary, namely, a set of terms designating properties and relations observable by means of the accepted techniques. For our purposes it is entirely sufficient that these terms, constituting the observational vocabulary, be given. An observation sentence is then defined simply as a sentence affirming or denying that a given object, or sequence of objects, possesses one of those observable attributes.[28]

Let it be noted that we do not require an observation sentence to be true, nor to be accepted on the basis of actual observations; rather, an observation sentence expresses something that is decidable by means of the accepted techniques of

28. The concept of observation sentence has, in the context of our study, a status and a logical function closely akin to that of the concepts of protocol statement or basis sentence, etc., as used in many recent studies of empiricism. However, the conception of observation sentence which is being proposed in the present study is more liberal in that it renders the discussion of the logical problems of testing and confirmation independent of various highly controversial epistemological issues; thus, *e.g.*, we do not stipulate that observation reports must be about psychic events, or about sense perceptions (*i.e.* that they have to be expressed in terms of a vocabulary of phenomenology, or of introspective psychology). According to the conception of observation sentence adopted in the present study, the "objects" referred to in an observation sentence may be construed in any one of the senses just referred to, or in various other ways; for example, they might be space-time regions, or again physical objects such as stones, trees, etc. (most of the illustrations given throughout this article represent observation sentences belonging to this kind of "thing language"); all that we require is that the few very general conditions stated above be satisfied.

These conditions impose on observation sentences and on observation reports certain restrictions with respect to their form; in particular, neither kind of sentence may contain any quantifiers. This stipulation recommends itself for the purposes of the logical analysis here to be undertaken; but we do not wish to claim that this formal restriction is indispensable. On the contrary, it is quite possible and perhaps desirable also to allow for observation sentences containing quantifiers: our simplifying assumption is introduced mainly in order to avoid considerable logical complications in the definition of confirmation.

observation. In other words, an observation sentence describes a possible outcome of the accepted observational techniques; it asserts something that might conceivably be established by means of those techniques. Possibly, the term "observation-type sentence" would be more suggestive; but for convenience we give preference to the shorter term. An analogous comment applies, of course, to our definition of an observation report as a class or a conjunction of observation sentences. The need for this broad conception of observation sentences and observation reports is readily recognized: Confirmation as here conceived is a logical relationship between sentences, just as logical consequence is. Now whether a sentence S_2 is a consequence of a sentence S_1 does not depend on whether or not S_1 is true (or known to be true); and analogously, the criteria of whether a given statement, expressed in terms of the observational vocabulary, confirms a certain hypothesis cannot depend on whether the statements in the report are true, or based on actual experience, or the like. Our definition of confirmation must enable us to indicate what kind of evidence *would* confirm a given hypothesis *if* it were available; and clearly the sentence characterizing such evidence can be required only to express something that *might* be observed, but not necessarily something that has actually been established by observation.

It may be helpful to carry the analogy between confirmation and consequence one step further. The truth or falsity of S_1 is irrelevant for the question of whether S_2 is a consequence of S_1 (whether S_2 can be validly inferred from S_1); but in a logical inference which justifies a sentence S_2 by showing that it is a logical consequence of a conjunction of premises, S_1, we can be certain of the truth of S_2 only if we know S_1 to be true. Analogously, the question of whether an observation report stands in the relation of confirmation to a given hypothesis does not depend on whether the report states actual or fictitious observational findings; but for a decision as to the soundness or acceptability of a hypothesis which is confirmed by a certain report, it is of course necessary to know whether the report is based on actual experience or not. Just as a conclusion of a logical inference, shown to be true, must be (a1) validly inferred from (a2) a set of true premises, so a hypothesis, to be scientifically acceptable, must be (b1) formally confirmed by (b2) reliable reports on observational findings.

The central problem of this essay is to establish general criteria for the formal relation of confirmation as referred to in (b1); the analysis of the concept of a reliable observation report, which belongs largely to the field of pragmatics,[29] falls outside the scope of the present study. One point, however, deserves mention here. A statement in the form of an observation report (for example, about the position of the pointer of a certain thermograph at 3 A.M.) may be accepted or

29. An account of the concept of pragmatics may be found in the publications listed in note 26.

rejected in science either on the basis of direct observation, or because it is indirectly confirmed or disconfirmed by other accepted observation sentences (in the example, these might be sentences describing the curve traced by the pointer during the night); and because of this possibility of indirect confirmation, our study has a bearing also on the question of the acceptance of hypotheses which have themselves the form of observation reports.

The conception of confirmation as a relation between sentences analogous to that of logical consequence suggests yet another requirement for the attempted definition of confirmation: While logical consequence has to be conceived of as a basically semantical relation between sentences, it has been possible, for certain languages, to establish criteria of logical consequence in purely syntactical terms. Analogously, confirmation may be conceived of as a semantical relation between an observation report and a hypothesis; but the parallel with the consequence relation suggests that it should be possible, for certain languages, to establish purely syntactical criteria of confirmation. The subsequent considerations will indeed eventuate in a definition of confirmation based on the concept of logical consequence and other purely syntactical concepts.

The interpretation of confirmation as a logical relation between sentences involves no essential change in the central problem of the present study. In particular, all the points made in the preceding sections can readily be rephrased in accordance with this interpretation. Thus, for example, the assertion that an object *a* which is a swan and white confirms the hypothesis '(x) [Swan$(x) \supset$ White(x)]' can be expressed by saying that the observation report 'Swan$(a)\cdot$White(a)' confirms that hypothesis. Similarly, the equivalence condition can be reformulated as follows: If an observation report confirms a certain sentence, then it also confirms every sentence which is logically equivalent with the latter. Nicod's criterion as well as our grounds for rejecting it can be reformulated along the same lines. We presented Nicod's concept of confirmation as referring to a relation between nonlinguistic objects on one hand and sentences on the other because this approach seemed to approximate most closely Nicod's own formulations,[30] and because it enabled us to avoid certain technicalities which are actually unnecessary in that context.

7. THE PREDICTION-CRITERION OF CONFIRMATION AND ITS SHORTCOMINGS

We are now in a position to analyze a second conception of confirmation,

30. (Added in 1964.) Actually this is not correct; *cf.* note 17 above. But, as is readily seen, the objections raised in this article against Nicod's criterion remain in force also when that criterion is understood as taking general hypotheses to be confirmed or disconfirmed by propositions rather than by objects.

which is reflected in many methodological discussions and which can claim a great deal of plausibility. Its basic idea is very simple: General hypotheses in science as well as in everyday use are intended to enable us to anticipate future events; hence, it seems reasonable to count any prediction that is borne out by subsequent observation as confirming evidence for the hypothesis on which it is based, and any prediction that fails as disconfirming evidence. To illustrate: Let H_1 be the hypothesis that all metals, when heated, expand; symbolically: '(x) [(Metal (x) · Heated (x)) ⊃ Exp(x)]'. If we have an observation report to the effect that a certain object *a* is metallic and is heated, then by means of H_1 we can derive the prediction that *a* expands. Suppose that this is borne out by observation and described in an additional observation statement. We should then have the total observation report: {Metal(*a*), Heated(*a*), Exp(*a*)}.[31] This report would be qualified as confirming evidence for H_1 because its last sentence bears out what could be predicted, or derived, from the first two by means of H_1; more explicitly, because the last sentence can be derived from the first two in conjunction with H_1. Now let H_2 be the hypothesis that all swans are white; symbolically: '(x) [Swan (x) ⊃ White(x)]'; and consider the observation report {Swan(*a*), ∼ White(*a*)}. This report would constitute disconfirming evidence for H_2 because the second of its sentences contradicts (and thus fails to bear out) the prediction 'White(*a*)' which can be deduced from the first sentence in conjunction with H_2; or, symmetrically, because the first sentence contradicts the consequence '∼Swan(*a*)' which can be derived from the second in conjunction with H_2. Obviously, either of these formulations implies that H_2 is incompatible with the given observation report. These illustrations suggest the following general definition of confirmation:

Prediction criterion of Confirmation: Let *H* be a hypothesis, *B* an observation report, *i.e.* a class of observation sentences. Then

(*a*) *B* is said to confirm *H* if *B* can be divided into two mutually exclusive subclasses B_1 and B_2 such that B_2 is not empty, and every sentence of B_2 can be logically deduced from B_1 in conjunction with *H*, but not from B_1 alone;

(*b*) *B* is said to disconfirm *H* if *H* logically contradicts *B*;[32]

31. An (observation) report, it will be recalled, may be represented by a conjunction or by a class of observation sentences: in the latter case, we characterize it by writing the sentences between braces; the single quotes which normally would be used to mention the sentences are, for convenience, assumed to be absorbed by the braces.

32. It might seem more natural to stipulate that *B* disconfirms *H* if it can be divided into two mutually exclusive classes B_1 and B_2 such that the denial of at least one sentence in B_2 can be deduced from B_1 in conjunction with *H*; but this condition can be shown to be equivalent to (*b*) above.

(*c*) *B* is said to be neutral with respect to *H* if it neither confirms nor disconfirms *H*.[33]

But while this criterion is quite sound as a statement of sufficient conditions of confirmation for hypotheses of the type illustrated above, it is considerably too narrow to serve as a general definition of confirmation. Generally speaking, this criterion would serve its purpose if all scientific hypotheses could be construed as asserting regular connections between observable features of the subject matter under investigation; *i.e.* if they all were of the form "Whenever the observable characteristic *P* is present in an object or a situation, then the observable characteristic Q is present as well." But actually, most scientific hypotheses and laws are not of this simple type; as a rule, they express regular connections of characteristics which are not observable in the sense of direct observability, nor even in a much more liberal sense. Consider, for example, the following hypothesis: 'Whenever plane-polarized light of wave length λ traverses a layer of quartz of thickness *d*, then its plane of polarization is rotated through an angle α which is proportional to d/λ'. Let us assume that the observational vocabulary, by means of which our observation reports have to be formulated, contains exclusively terms referring to directly observable attributes. Then, since the question of whether a given ray of light is plane-polarized and has the wave length λ cannot be decided by means of direct observation, no observation report of the kind here admitted could afford information of this type. This in itself would not be crucial if at least we could assume that the fact that a given ray of light is plane-polarized, etc., could be logically inferred from some possible observation report; for then, from a suitable report of this kind, in conjunction with the given hypothesis, one would be able to predict a rotation of the plane of polarization; and from this prediction, which itself is not yet expressed in exclusively observational terms, one might expect to derive further predictions in the form of genuine observation sentences. But actually, a hypothesis to the effect that a given ray of light is plane-polarized has to be considered as a general hypothesis which entails an unlimited number of observation sentences; thus it cannot be logically inferred from, but at best be confirmed by, a suitable set of observational findings. The logically essential point can best be exhibited by

33. The following quotations from A. J. Ayer's book *Language, Truth and Logic* (London, 1936) formulate in a particularly clear fashion the conception of confirmation as successful prediction (although the two are not explicitly identified by definition): ". . . the function of an empirical hypothesis is to enable us to anticipate experience. Accordingly, if an observation to which a given proposition is relevant conforms to our expectations, . . . that proposition is confirmed" (*loc. cit.* pp. 142–43); ". . . it is the mark of a genuine factual proposition . . . that some experiential propositions can be deduced from it in conjunction with certain premises without being deducible from those other premises alone." (*loc. cit.* p. 26).

reference to a very simple abstract case: Let us assume that R_1 and R_2 are two relations of a kind accessible to direct observation, and that the field of scientific investigation contains infinitely many objects. Consider now the hypothesis

$$(H) \qquad (x)[(y)R_1(x, y) \supset (Ez)R_2(x, z)]$$

i.e.: Whenever an object x stands in R_1 to every object y, then it stands in R_2 to at least one object z. This simple hypothesis has the following property: However many observation sentences may be given, H does not enable us to derive any new observation sentences from them. Indeed—to state the reason in suggestive though not formally rigorous terms— in order to make a prediction concerning some specific object a, we should first have to know that a stands in R_1 to every object; and this necessary information clearly cannot be contained in any finite number, however large, of observation sentences, because a finite set of observation sentences can tell us at best for a finite number of objects that a stands in R_1 to them. Thus an observation report, which always involves only a finite number of observation sentences, can never provide a sufficiently broad basis for a prediction by means of H.[34] Besides, even if we did know that a stood in R_1 to every object, the prediction derivable by means of H would not be an observation sentence; it would assert that a stands in R_2 to *some* object, without specifying which, and where to find it. Thus, H is an empirical hypothesis that contains, besides purely logical terms, only expressions belonging to the observational vocabulary, and yet the predictions which it renders possible neither start from nor lead to observation reports.

It is therefore a considerable oversimplification to say that scientific hypotheses and theories enable us to derive predictions of future experiences from descriptions of past ones. Unquestionably, scientific hypotheses do have a predictive function; but the way in which they perform this function, the manner in which they establish logical connections between observation reports, is logically more complex than a deductive inference. Thus, in the last illustration, the predictive use of H may assume the following form: On the basis of a number of individual tests, which show that a does stand in R_1 to three objects b, c, and d, we might accept the hypothesis that a stands in R_1 to all objects; or in terms of our formal mode of speech: In view of the observation report $\{R_1(a, b), R_1(a,c), R_1(a, d)\}$, the hypothesis that $(y)R_1(a, y)$ might be accepted as confirmed by,

34. To illustrate: a might be an iron object which possibly is a magnet; R_1 might be the relation of attracting; the objects under investigation might be iron objects. Then a finite number of observation reports to the effect that a did attract a particular piece of iron is insufficient to *infer* that a will attract every piece of iron.

though not logically inferable from, that report.[35] This process might be referred to as quasi-induction.[36] From the hypothesis thus established we can then proceed to derive, by means of H, the prediction that a stands in R_2 to at least one object. This again, as was pointed out above, is not an observation sentence; and indeed no observation sentence can be derived from it; but it can, in turn, be confirmed by a suitable observation sentence, such as '$R_2(a, b)$'. In other cases, the prediction of actual observation sentences may be possible; thus if the given hypothesis asserts that $(x)((y)R_1(x, y) \supset (z)R_2(x, z))$, then after quasi-inductively accepting, as above, that $(y)R_1(a, y)$, we can derive, by means of the given hypothesis, the sentence that a stands in R_2 to every object, and thence, we can deduce particular predictions such as '$R_2(a, b)$', which do have the form of observation sentences.

Thus, the chain of reasoning which leads from given observational findings to the "prediction" of new ones actually involves, besides deductive inferences, certain quasi-inductive steps each of which consists in the acceptance of an intermediate statement on the basis of confirming, but usually not logically conclusive, evidence. In most scientific predictions, this general pattern occurs in multiple reiteration; an analysis of the predictive use of the hypothesis mentioned above, concerning plane-polarized light, could serve as an illustration. In the present context, however, this general account of the structure of scientific prediction is sufficient. It shows that a general definition of confirmation by reference to successful prediction becomes circular; indeed, in order to make the original formulation of the prediction-criterion of confirmation sufficiently comprehensive, we should have to replace the phrase "can be logically deduced" by "can be obtained by a series of steps of deduction and quasi-induction"; and the definition of "quasi-induction" in the above sense presupposes the concept of confirmation.

Let us note, as a by-product of the preceding consideration, that an adequate analysis of scientific prediction (and analogously, of scientific explanation, and of the testing of empirical hypotheses) requires an analysis of the concept of confirmation. The reason may be restated in general terms as follows: Scientific

35. Thus, in the illustration given in the preceding footnote, the hypothesis that the object a will attract every piece of iron might be accepted as sufficiently well substantiated by, though by no means derivable from, an observation report to the effect that in tests a did attract the iron objects b, c, and d.

36. The prefix "quasi" is to contradistinguish the procedure in question from so-called induction, which is usually supposed to be a method of discovering, or inferring, general regularities on the basis of a finite number of instances. In quasi-induction, the hypothesis is not "discovered" but has to be *given* in addition to the observation report; the process consists in the acceptance of the hypothesis if it is deemed sufficiently confirmed by the observation report. *Cf.* also the discussion in section 1*c*, above.

laws and theories, as a rule, connect terms which lie on the level of abstract theoretical constructs rather than on that of direct observation; and from observation sentences, no merely deductive logical inference leads to statements about theoretical constructs, which can serve as starting points for scientific predictions; statements about theoretical constructs, such as 'This piece of iron is magnetic' or 'Here, a plane-polarized ray of light traverses a quartz crystal' can be confirmed, but not entailed, by observation reports. Thus, even though based on general scientific laws, the prediction of new observational findings by means of given ones is a process involving confirmation in addition to logical deduction.[37]

8. CONDITIONS OF ADEQUACY FOR ANY DEFINITION OF CONFIRMATION

The two most customary conceptions of confirmation, which were rendered explicit in Nicod's criterion and in the prediction criterion, have thus been found unsuitable for a general definition of confirmation. Besides this negative result, the preceding analysis has also exhibited certain logical characteristics of scientific prediction, explanation, and testing, and it has led to the establishment of certain standards which an adequate definition of confirmation has to satisfy. These standards include the equivalence condition and the requirement that the definition of confirmation be applicable to hypotheses of any degree of logical complexity, rather than to the simplest type of universal conditional only. An adequate definition of confirmation, however, has to satisfy several further logical requirements, to which we now turn.

First of all, it will be agreed that any sentence which is logically entailed by a given observation report has to be considered as confirmed by that report: entailment is a special case of confirmation. Thus, *e.g.*, we want to say that the observation report '*a* is black' confirms the sentence (hypothesis) '*a* is black or grey'; and—to refer to one of the illustrations given in the preceding section—the observation sentence '$R_2(a, b)$' should certainly be confirming evidence for the sentence '$(Ez)R_2(a, z)$'. We are therefore led to the stipulation that any adequate definition of confirmation must insure the fulfilment of the

37. In the above sketch of the structure of scientific prediction, we have disregarded the fact that in practically every case where a prediction is said to be obtained by means of a certain hypothesis or theory, a considerable body of auxiliary theories is used in addition. Thus, the prediction of observable effects of the deflection of light in the gravitational field of the sun on the basis of the general theory of relativity requires such auxiliary theories as mechanics and optics. But an explicit consideration of this fact would not affect our result that scientific predictions, even when based on hypotheses or theories of universal form, still are not purely deductive in character, but involve quasi-inductive steps as well.

(8.1) ENTAILMENT CONDITION. Any sentence which is entailed by an observation report is confirmed by it.[38]

This condition is suggested by the preceding consideration, but of course not proved by it. To make it a standard of adequacy for the definition of confirmation means to lay down the stipulation that a proposed definition of confirmation will be rejected as logically inadequate if it is not constructed in such a way that (8.1) is unconditionally satisfied. An analogous remark applies to the subsequently proposed further standards of adequacy.

Second, an observation report which confirms certain hypotheses would invariably be qualified as confirming any consequence of those hypotheses. Indeed: any such consequence is but an assertion of all or part of the combined content of the original hypotheses and has therefore to be regarded as confirmed by any evidence which confirms all of the latter. This suggests the following condition of adequacy:

(8.2) CONSEQUENCE CONDITION. If an observation report confirms every one of a class K of sentences, then it also confirms any sentence which is a logical consequence of K.

If (8.2) is satisfied, then the same is true of the following two more special conditions:

(8.21) SPECIAL CONSEQUENCE CONDITION. If an observation report confirms a hypothesis H, then it also confirms every consequence of H.

(8.22) EQUIVALENCE CONDITION. If an observation report confirms a hypothesis H, then it also confirms every hypothesis which is logically equivalent with H.

(8.22) follows from (8.21) in view of the fact that equivalent hypotheses are mutual consequences of each other. Thus, the satisfaction of the consequence condition entails that of our earlier equivalence condition, and the latter loses its status of an independent requirement.

In view of the apparent obviousness of these conditions, it is interesting to note that the definition of confirmation in terms of successful prediction, while satisfying the equivalence condition, would violate the consequence condition. Consider, for example, the formulation of the prediction criterion given in the

38. As a consequence of this stimpulation, a contradictory observation report, such as [Black(a), $\sim$ Black(a)] confirms every sentence, because it has every sentence as a consequence. Of course, it is possible to exclude contradictory observation reports altogether by a slight restriction of the definition of 'observation report'. There is, however, no important reason to do so.

earlier part of the preceding section. Clearly, if the observational findings B_2 can be predicted on the basis of the findings B_1 by means of the hypothesis H, the same prediction is obtainable by means of any equivalent hypothesis, but not generally by means of a weaker one.

On the other hand, any prediction obtainable by means of H can obviously also be established by means of any hypothesis which is stronger than H, *i.e.* which logically entails H. Thus while the consequence condition stipulates in effect that whatever confirms a given hypothesis also confirms any weaker hypothesis, the relation of confirmation defined in terms of successful prediction would satisfy the condition that whatever confirms a given hypothesis also confirms every stronger one.

But is this "converse consequence condition," as it might be called, not reasonable enough, indeed should it not be included among our standards of adequacy for the definition of confirmation? The second of these two suggestions can be readily disposed of: The adoption of the new condition, in addition to (8.1) and (8.2), would have the consequence that any observation report B would confirm any hypothesis H whatsoever. Thus, *e.g.*, if B is the report '*a* is a raven' and H is Hooke's law, then, according to (8.1), B confirms the sentence '*a* is a raven'; hence B would, according to the converse consequence condition, confirm the stronger sentence '*a* is a raven, and Hooke's law holds'; and finally, by virtue of (8.2), B would confirm H, which is a consequence of the last sentence. Obviously, the same type of argument can be applied in all other cases.

But is it not true, after all, that very often observational data which confirm a hypothesis H are considered also as confirming a stronger hypothesis? Is it not true, for example, that those experimental findings which confirm Galileo's law, or Kepler's laws, are considered also as confirming Newton's law of gravitation?[39] This is indeed the case, but it does not justify the acceptance of the converse consequence condition as a general rule of the logic of confirmation; for in the cases just mentioned, the weaker hypothesis is connected with the stronger one by a logical bond of a particular kind: it is essentially a substitution instance of the stronger one; thus, *e.g.*, while the law of gravitation refers to the force obtaining between any two bodies, Galileo's law is a specialization referring to the case where one of the bodies is the earth, the other an object near its surface. In the preceding case, however, where Hooke's law was shown to be confirmed by the observation report that *a* is a raven, this situation does not prevail; and here, the rule that whatever confirms a given hypothesis also confirms any stronger

39. Strictly speaking, Galileo's law and Kepler's laws can be deduced from the law of gravitation only if certain additional hypotheses—including the laws of motion—are presupposed; but this does not affect the point under discussion.

one becomes an entirely absurd principle. Thus, the converse consequence condition does not provide a sound general condition of adequacy.[40]

A third condition remains to be stated:[41]

(8.3) CONSISTENCY CONDITION. Every logically consistent observation report is logically compatible with the class of all the hypotheses which it confirms.

The two most important implications of this requirement are the following:
(8.31) Unless an observation report is self-contradictory,[42] it does not confirm any hypothesis with which it is not logically compatible.
(8.32) Unless an observation report is self-contradictory, it does not confirm any hypotheses which contradict each other.

The first of these corollaries will readily be accepted; the second, however,—and consequently (8.3) itself—will perhaps be felt to embody a too severe restriction. It might be pointed out, for example, that a finite set of measurements concerning the changes of one physical magnitude, x, associated with those of another, y, may conform to, and thus be said to confirm, several different hypotheses as to the particular mathematical function in terms of which the relationship of x and y can be expressed; but such hypotheses are incompatible because to at least one value of x, they will assign different values of y.

No doubt it is possible to liberalize the formal standards of adequacy in line with these considerations. This would amount to dropping (8.3) and (8.32) and retaining only (8.31). One of the effects of this measure would be that when a logically consistent observation report B confirms each of two hypotheses, it

40. William Barrett, in a paper entitled "Discussion on Dewey's Logic" (*The Philosophical Review*, vol. 50, 1941, pp. 305 ff., esp. p. 312) raises some questions closely related to what we have called above the consequence condition and the converse consequence condition. In fact, he invokes the latter (without stating it explicitly) in an argument which is designed to show that "not every observation which confirms a sentence need also confirm all its consequences," in other words, that the special consequence condition (8.21) need not always be satisfied. He supports his point by reference to "the simplest case: the sentence 'C' is an abbreviation of 'A·B', and the observation O confirms 'A', *and so* 'C', but is irrelevant to 'B', which is a consequence of 'C'." (Italics mine).

For reasons contained in the above discussion of the consequence condition and the converse consequence condition, the application of the latter in the case under consideration seems to me unjustifiable, so that the illustration does not prove the author's point; and indeed, there seems to be every reason to preserve the unrestricted validity of the consequence condition. As a matter of fact, Barrett himself argues that "the degree of confirmation for the consequence of a sentence cannot be less than that of the sentence itself"; this is indeed quite sound; but it is hard to see how the recognition of this principle can be reconciled with a renunciation of the special consequence condition, which may be considered simply as its correlate for the nongraduated relation of confirmation.

41. For a fourth condition, see note 46.

42. A contradictory observation report confirms every hypothesis (*cf.* note 38) and is, of course, incompatible with every one of the hypotheses it confirms.

does not necessarily confirm their conjunction; for the hypotheses might be mutually incompatible, hence their conjunction self-contradictory; consequently, by (8.31), *B* could not confirm it. This consequence is intuitively rather awkward, and one might therefore feel inclined to suggest that while (8.3) should be dropped and (8.31) retained, (8.32) should be replaced by the requirement (8.33): If an observation sentence confirms each of two hypotheses, then it also confirms their conjunction. But it can readily be shown that by virtue of (8.2) this set of conditions entails the fulfilment of (8.32).

If, therefore, the condition (8.3) appears to be too rigorous, the most obvious alternative would seem to lie in replacing (8.3) and its corollaries by the much weaker condition (8.31) alone. [Added in 1970: But as G. L. Massey has pointed out to me, satisfaction of (8.1), (8.2), and (8.31) logically implies satisfaction of (8.3); hence, that alternative fails.] One of the advantages of a definition which satisfies (8.3) is that it sets a limit, so to speak, to the strength of the hypotheses which can be confirmed by given evidence.[43]

The remainder of the present study, therefore, will be concerned exclusively with the problem of establishing a definition of confirmation which satisfies the more severe formal conditions represented by (8.1), (8.2), and (8.3) together.

The fulfilment of these requirements, which may be regarded as general laws of the logic of confirmation, is of course only a necessary, not a sufficient, condition for the adequacy of any proposed definition of confirmation. Thus, *e.g.*, if '*B* confirms *H*' were defined as meaning 'B logically entails *H*', then the above three conditions would clearly be satisfied; but the definition would not be adequate because confirmation has to be a more comprehensive relation than entailment (the latter might be referred to as the special case of *conclusive* confirmation). Thus, a definition of confirmation, to be acceptable, also has to be materially adequate: it has to provide a reasonably close approximation to that conception of confirmation which is implicit in scientific procedure and methodological discussion. That conception is vague and to some extent quite unclear, as I have tried to show in earlier parts of this paper; therefore, it would be too much to expect full agreement as to whether a proposed definition of confirmation is materially adequate. On the other hand, there will be rather general agreement on certain points; thus, *e.g.*, the identification of confirmation with entailment, or the Nicod criterion of confirmation as analyzed above, or any definition of confirmation by reference to a "sense of evidence," will probably now be admitted not to be adequate approximations to that concept of confirmation which is relevant for the logic of science.

43. This was pointed out to me by Dr. Nelson Goodman. The definition later to be outlined in this essay, which satisfies conditions (8.1), (8.2) and (8.3), lends itself, however, to certain generalizations which satisfy only the more liberal conditions of adequacy just considered.

On the other hand, the soundness of the logical analysis (which, in a clear sense, always involves a logical reconstruction) of a theoretical concept cannot be gauged simply by our feelings of satisfaction at a certain proposed analysis; and if there are, say, two alternative proposals for defining a term on the basis of a logical analysis, and if both appear to come fairly close to the intended meaning, then the choice has to be made largely by reference to such features as the logical properties of the two reconstructions, and the comprehensiveness and simplicity of the theories to which they lead.

9. THE SATISFACTION CRITERION OF CONFIRMATION

As has been mentioned before, a precise definition of confirmation requires reference to some definite "language of science," in which all observation reports and all hypotheses under consideration are assumed to be formulated, and whose logical structure is supposed to be precisely determined. The more complex this language, and the richer its logical means of expression, the more difficult it will be, as a rule, to establish an adequate definition of confirmation for it. However, the problem has been solved at least for certain cases: With respect to languages of a comparatively simple logical structure, it has been possible to construct an explicit definition of confirmation which satisfies all of the above logical requirements, and which appears to be intuitively rather adequate. An exposition of the technical details of this definition has been published elsewhere;[44] in the present study, which is concerned with the general logical and methodological aspects of the problem of confirmation rather than with technical details, it will be

44. In my article referred to in note 1. The logical structure of the languages to which the definition in question is applicable is that of the lower functional calculus with individual constants, and with predicate constants of any degree. All sentences of the language are assumed to be formed exclusively by means of predicate constants, individual constants, individual variables, universal and existential quantifiers for individual variables, and the connective symbols of denial, conjunction, alternation, and implication. The use of predicate variables or of the identity sign is not permitted.

As to the predicate constants, they are all assumed to belong to the observational vocabulary, *i.e.* to denote properties or relations observable by means of the accepted techniques. ("Abstract" predicate terms are supposed to be defined by means of those of the observational vocabulary and then actually to be replaced by their definientia, so that they never occur explicitly.)

As a consequence of these stipulations, an observation report can be characterized simply as a conjunction of sentences of the kind illustrated by '$P(a)$', '$\sim P(b)$', '$R(c, d)$', '$\sim R(e,f)$', etc., where 'P', 'R', etc., belong to the observational vocabulary, and 'a', 'b', 'c', 'd', 'e', 'f', etc., are individual names, denoting specific objects. It is also possible to define an observation report more liberally as any sentence containing no quantifiers, which means that besides conjunctions also alternations and implication sentences formed out of the above kind of components are included among the observation reports.

attempted to characterize the definition of confirmation thus obtained as clearly as possible with a minimum of technicalities.

Consider the simple case of the hypothesis *H*: '(*x*)(Raven(*x*) ⊃ Black(*x*))', where 'Raven' and 'Black' are supposed to be terms of our observational vocabulary. Let *B* be an observation report to the effect that Raven(*a*) · Black(*a*) · ∼ Raven(*c*) · Black(*c*) · ∼ Raven(*d*) · ∼ Black(*d*). Then *B* may be said to confirm *H* in the following sense: There are three objects mentioned in *B*, namely *a*, *c*, and *d*; and as far as these are concerned, *B* informs us that all those which are ravens (*i.e.* just the object *a*) are also black.[45] In other words, from the information contained in *B* we can infer that the hypothesis *H* does hold true within the finite class of those objects which are mentioned in *B*.

Let us apply the same consideration to a hypothesis of a logically more complex structure. Let *H* be the hypothesis 'Everybody likes somebody'; in symbols: '(*x*)(*Ey*)Likes(*x*, *y*)', *i.e.* 'For every (person) *x*, there exists at least one (not necessarily different person) *y* such that *x* likes *y*'. (Here again, 'Likes' is supposed to be a relation term which occurs in our observational vocabulary.) Suppose now that we are given an observation report *B* in which the names of two persons, say '*e*' and '*f*', occur. Under what conditions shall we say that *B* confirms *H*? The previous illustration suggests the answer: If from *B* we can infer that *H* is satisfied within the finite class {*e*,*f*}; *i.e.*, that within {*e*,*f*} everybody likes somebody. This in turn means that *e* likes *e* or *f*, and *f* likes *e* or *f*. Thus, *B* would be said to confirm *H* if *B* entailed the statement '*e* likes *e* or *f*, and *f* likes *e* or *f*'. This latter statement will be called the development of *H* for the finite class {*e*,*f*}.

The concept of *development of a hypothesis, H, for a finite class of individuals, C*, can be defined precisely by recursion; here it will suffice to say that the development of *H* for *C* states what *H* would assert if there existed exclusively those objects which are elements of *C*. Thus, *e.g.*, the development of the hypothesis H_1='[(*x*)(*P*(*x*) v Q(*x*)]' (*i.e.* 'Every object has the property *P* or the property *Q*') for the class {*a*, *b*} is '[*P*(*a*) v Q(*a*)] · [*P*(*b*) v Q(*b*)]' (*i.e.* '*a* has the property *P* or the property Q, and *b* has the property *P* or the property *Q*'); the development of the existential hypothesis H_2 that at least one object has the property *P*, *i.e.* '(*Ex*)*P*(*x*)', for {*a*, *b*} is '*P*(*a*) v *P*(*b*)'; the development of a hypothesis which contains no quantifiers, such as H_3: '*P*(*c*) v *K*(*c*)' is defined as that hypothesis itself, no matter what the reference class of individuals is.

A more detailed formal analysis based on considerations of this type leads to the introduction of a general relation of confirmation in two steps; the first

45. I am indebted to Dr. Nelson Goodman for having suggested this idea; it initiated all those considerations which finally led to the definition to be outlined below.

consists in defining a special relation of direct confirmation along the lines just indicated; the second step then defines the general relation of confirmation by reference to direct confirmation.

Omitting minor details, we may summarize the two definitions as follows:

(9.1 Df). An observation report *B* directly confirms a hypothesis *H* if *B* entails the development of *H* for the class of those objects which are mentioned in *B*.

(9.2 Df.) An observation report *B* confirms a hypothesis *H* if *H* is entailed by a class of sentences each of which is directly confirmed by *B*.

The criterion expressed in these definitions might be called the *satisfaction criterion of confirmation* because its basic idea consists in construing a hypothesis as confirmed by a given observation report if the hypothesis is satisfied in the finite class of those individuals which are mentioned in the report.

Let us now apply the two definitions to our last examples: The observation report B_1: '*P*(*a*) . *Q*(*b*)' directly confirms (and therefore also confirms) the hypothesis H_1, because it entails the development of H_1 for the class {*a*, *b*}, which was given above. The hypothesis H_3 is not directly confirmed by *B*, because its development, *i.e.* H_3 itself, obviously is not entailed by B_1. However, H_3 is entailed by H_1, which is directly confirmed by B_1; hence, by virtue of (9.2), B_1 confirms H_3. Similarly, it can readily be seen that B_1 directly confirms H_2.

Finally, to refer to the first illustration in this section: The observation report 'Raven(*a*) · Black(*a*) · ∼ Raven(*c*) · Black(*c*) · ∼ Raven(*d*) · ∼ Black(*d*)' confirms (even directly) the hypothesis '(*x*)[Raven(*x*) ⊃ Black(*x*)]', for it entails the development of the latter for the class {*a*, *c*, *d*}, which can be written as follows: '[Raven(*a*) ⊃ Black(*a*)] · [Raven(*c*) ⊃ Black(*c*)] · [Raven (*d*) ⊃ Black(*d*)]'.

It is now easy to define disconfirmation and neutrality:

(9.3 Df.) An observation report *B* disconfirms a hypothesis *H* if it confirms the denial of *H*.

(9.4 Df.) An observation report *B* is neutral with respect to a hypothesis *H* if *B* neither confirms nor disconfirms *H*.

By virtue of the criteria laid down in (9.2), (9.3), (9.4), every consistent observation report *B* divides all possible hypotheses into three mutually exclusive classes: those confirmed by *B*, those disconfirmed by *B*, and those with respect to which *B* is neutral.

The definition of confirmation here proposed can be shown to satisfy all the formal conditions of adequacy embodied in (8.1), (8.2), and (8.3) and their consequences. For the condition (8.2) this is easy to see; for the other conditions the proof is more complicated.[46]

46. For these proofs, see the article referred to in note 1. I should like to take this opportunity to point out and to remedy a certain defect of the definition of confirmation which was

(*continued overleaf*)

Furthermore, the application of the above definition of confirmation is not restricted to hypotheses of universal conditional form (as Nicod's criterion is, for example), nor to universal hypotheses in general; it applies, in fact, to any hypothesis which can be expressed by means of property and relation terms of the observational vocabulary of the given language, individual names, the customary connective symbols for 'not', 'and', 'or', 'if-then', and any number of universal and existential quantifiers.

Finally, as is suggested by the preceding illustrations as well as by the general considerations which underlie the establishment of the above definition, it seems that we have obtained a definition of confirmation which is also materially

developed in that article, and which has been outlined above: this defect was brought to my attention by a discussion with Dr. Olaf Helmer.

It will be agreed that an acceptable definition of confirmation should satisfy the following further condition which might well have been included among the logical standards of adequacy set up in section 8 above: (8.4) If B_1 and B_2 are logically equivalent observation reports and B_1 confirms (disconfirms, is neutral with respect to) a hypothesis H, then B_2, too, confirms (disconfirms, is neutral with respect to) H. This condition is indeed satisfied if observation reports are construed, as they have been in this article, as classes or conjunctions of observation sentences. As was indicated at the end of note 44, however, this restriction of observation reports to a conjunctive form is not essential; in fact, it has been adopted here only for greater convenience of exposition, and all the preceding results, including especially the definitions and theorems of the present section, remain applicable without change if observation reports are defined as sentences containing no quantifiers. (In this case, if 'P' and 'Q' belong to the observational vocabulary, such sentences as '$P(a) \vee Q(a)$', '$P(a) \vee \sim Q(b)$', etc., would qualify as observation reports.) This broader conception of observation reports was therefore adopted in the article referred to in note 1; but it has turned out that in this case, the definition of confirmation summarized above does not generally satisfy the requirement (8.4). Thus, *e.g.*, the observation reports, $B_1 =$ '$P(a)$' and $B_2 =$ '$P(a) \cdot [Q(b) \vee \sim Q(b)]$' are logically equivalent, but while B_1 confirms (and even directly confirms) the hypothesis $H_1 =$ '$(x)P(x)$', the second report does not do so, essentially because it does not entail '$P(a) \,.\, P(b)$', which is the development of H_1 for the class of those objects mentioned in B_2. This deficiency can be remedied as follows: The fact that B_2 fails to confirm H_1 is obviously due to the circumstance that B_2 contains the individual constant 'b', without asserting anything about b: The object b is mentioned only in an analytic component of B_2. The atomic constituent '$Q(b)$' will therefore be said to occur (twice) *inessentially* in B_2. Generally, an atomic constituent A of a molecular sentence S will be said to occur inessentially in S if by virtue of the rules of the sentential calculus S is equivalent to a molecular sentence in which A does not occur at all. Now an object will be said to be mentioned inessentially in an observation report if it is mentioned only in such components of that report as occur inessentially in it. The sentential calculus provides mechanical procedures for deciding whether a given observation report mentions any object inessentially, and for establishing equivalent formulations of the same report in which no object is mentioned inessentially. Finally, let us say that an object is mentioned essentially in an observation report if it is mentioned, but not only mentioned inessentially, in that report. Now we replace 9.1 by the following definition:

(9.1*a*) An observation report B directly confirms a hypothesis H if B entails the development of H for the class of those objects which are mentioned essentially in B.

The concept of confirmation as defined by (9.1*a*) and (9.2) now satisfies (8.4 in addition to (8.1, (8.2), (8.3) even if observation reports are construed in the broader fashion characterized earlier in this footnote.

adequate in the sense of being a reasonable approximation to the intended meaning of confirmation.

A brief discussion of certain special cases of confirmation might serve to shed further light on this latter aspect of our analysis.

10. THE RELATIVE AND THE ABSOLUTE CONCEPTS OF VERIFICATION AND FALSIFICATION

If an observation report entails a hypothesis H, then, by virtue of (8.1), it confirms H. This is in good agreement with the customary conception of confirming evidence; in fact, we have here an extreme case of confirmation, the case where *B conclusively confirms H*; this case is realized if, and only if, B entails H. We shall then also say that *B verifies H*. Thus, verification is a special case of confirmation; it is a logical relation between sentences; more specifically, it is simply the relation of entailment with its domain restricted to observation sentences.

Analogously, we shall say that *B conclusively disconfirms H*, or *B falsifies H*, if and only if B is incompatible with H; in this case, B entails the denial of H and therefore, by virtue of (8.1) and (9.3), confirms the denial of H and disconfirms H. Hence, falsification is a special case of disconfirmation; it is the logical relation of incompatibility between sentences, with its domain restricted to observation sentences.

Clearly, the concepts of *verification and falsification* as here defined are *relative;* a hypothesis can be said to be verified or falsified only with respect to some observation report; and a hypothesis may be verified by one observation report and may not be verified by another. There are, however, hypotheses which cannot be verified and others which cannot be falsified by any observation report. This will be shown presently. We shall say that a given *hypothesis is verifiable* (*falsifiable*) if it is possible to construct an observation report which verifies (falsifies) the hypothesis. Whether a hypothesis is verifiable, or falsifiable, in this sense depends exclusively on its logical form. Briefly, the following cases may be distinguished:

(*a*) If a hypothesis does not contain the quantifier terms 'all' and 'some' or their symbolic equivalents, then it is both verifiable and falsifiable. Thus, *e.g.*, the hypothesis 'Object *a* turns blue or green' is entailed and thus verified by the report 'Object *a* turns blue'; and the same hypothesis is incompatible with, and thus falsified by, the report 'Object *a* turns neither blue nor green'.

(*b*) A purely existential hypothesis (*i.e.* one which can be symbolized by a formula consisting of one or more existential quantifiers followed by a sentential function containing no quantifiers) is verifiable, but not falsifiable, if—as is usually assumed—the universe of discourse contains an infinite number of objects. Thus, *e.g.*, the hypothesis 'There are blue roses' is verified by the observation report 'Object *a* is a blue rose', but no finite observation report can ever contradict and thus falsify the hypothesis.

(*c*) Conversely, a purely universal hypothesis (symbolized by a formula consisting of one or more universal quantifiers followed by a sentential function containing no quantifiers) is falsifiable but not verifiable for an infinite universe of discourse. Thus, *e.g.*, the hypothesis '$(x)[\text{Swan}(x) \supset \text{White}(x)]$' is completely falsified by the observation report $\{\text{Swan}(a), \sim \text{White}(a)\}$; but no finite observation report can entail and thus verify the hypothesis in question.

(*d*) Hypotheses which cannot be expressed by sentences of one of the three types mentioned so far, and which in this sense require both universal and existential quantifiers for their formulation, are as a rule neither verifiable nor falsifiable.[47] Thus, *e.g.*, the hypothesis 'Every substance is soluble in some solvent'—symbolically '$(x)(Ey)\text{Soluble}(x, y)$'—is neither entailed by nor incompatible with any observation report, no matter how many cases of solubility or non-solubility of particular substances in particular solvents the report may list. An analogous remark applies to the hypothesis 'You can fool some of the people all of the time', whose symbolic formulation '$(Ex)(t)\text{Fl}(x,t)$' contains one existential and one universal quantifier. But of course, all of the hypotheses belonging to this fourth class are capable of being confirmed or disconfirmed by suitable observation reports; this was illustrated early in section 9 by reference to the hypothesis '$(x)(Ey)\text{Likes}(x, y)$'.

This rather detailed account of verification and falsification has been presented not only in the hope of further elucidating the meaning of confirmation and disconfirmation as defined above, but also in order to provide a basis for a sharp differentiation of two meanings of verification (and similarly of falsification) which have not always been clearly separated in recent discussions of the character of empirical knowledge. One of the two meanings of verification which we wish to distinguish here is the relative concept just explained; for greater clarity we shall sometimes refer to it as *relative verification*. The other meaning is what may be called *absolute or definitive verification*. This latter concept of verification does not belong to formal logic, but rather to pragmatics: it refers to the acceptance of hypotheses by observers or scientists, etc., on the basis of relevant evidence. Generally speaking, we may distinguish three phases in the scientific test of a given hypothesis (which do not necessarily occur in the order in which they are listed here). The first phase consists in the performance of suitable

47. A more precise study of the conditions of nonverifiability and nonfalsifiability would involve technicalities which are unnecessary for the purposes of the present study. Not all hypotheses of the type described in (*d*) are neither verifiable nor falsifiable; thus, *e.g.*, the hypothesis '$(x)(Ey)[P(x) \vee Q(y)]$' is verified by the report '$Q(a)$', and the hypothesis '$(x)(Ey)[(P(x) \cdot Q(y)]$' is falsified by '$\sim P(a)$'.

experiments or observations and the ensuing acceptance of observation reports stating the results obtained; the next phase consists in confronting the given hypothesis with the accepted observation reports, *i.e.* in ascertaining whether the latter constitute confirming, disconfirming or irrelevant evidence with respect to the hypothesis; the final phase consists either in accepting or rejecting the hypothesis on the strength of the confirming or disconfirming evidence constituted by the accepted observation reports, or in suspending judgment, awaiting the establishment of further relevant evidence.

The present study has been concerned almost exclusively with the second phase. As we have seen, this phase is of a purely logical character; the standards of evaluation here invoked—namely the criteria of confirmation, disconfirmation and neutrality—can be completely formulated in terms of concepts belonging to pure logic.

The first phase, on the other hand, is of a pragmatic character; it involves no logical confrontation of sentences with other sentences. It consists in performing certain experiments or systematic observations and noting the results. The latter are expressed in sentences which have the form of observation reports, and their acceptance by the scientist is connected (by causal, not by logical relations) with experiences occurring in those tests. Of course, a sentence which has the form of an observation report may in certain cases be accepted, not on the basis of direct observation, but because it is confirmed by other observation reports which were previously established; but this process is illustrative of the second phase, which was discussed before. Here we are considering the case where a sentence is accepted directly "on the basis of experiential findings" rather than because it is supported by previously established statements.

The third phase, too, can be construed as pragmatic, namely as consisting in a decision on the part of a scientist or a group of scientists to accept (or reject, or leave in suspense, as the case may be) a given hypothesis after ascertaining what amount of confirming or of disconfirming evidence for the hypothesis is contained in the totality of the accepted observation sentences. However, it may well be attempted to give a reconstruction of this phase in purely logical terms. This would require the establishment of general "rules of acceptance". Roughly speaking, these rules would state how well a given hypothesis has to be confirmed by the accepted observation reports to be scientifically acceptable itself;[48] *i.e.* the rules would formulate criteria for the acceptance or rejection of a hypothesis by reference to the kind and amount of confirming or disconfirming evidence for it embodied in the totality of accepted observation reports. Possibly, these

48. A stimulating discussion of some aspects of what we have called rules of acceptance is contained in an article by Felix Kaufmann, "The Logical Rules of Scientific Procedure", *Philosophy and Phenomenological Research*, June, 1942.

criteria would also refer to such additional factors as the simplicity of the hypothesis in question, the manner in which it fits into the system of previously accepted theories, etc. It is at present an open question to what extent a satisfactory system of such rules can be formulated in purely logical terms.[49]

At any rate, the acceptance of a hypothesis on the basis of a sufficient body of confirming evidence will as a rule be tentative, and will hold only "until further notice," *i.e.* with the proviso that if new and unfavorable evidence should turn up (in other words, if new observation reports should be accepted which disconfirm the hypothesis in question) the hypothesis will be abandoned again.

Are there any exceptions to this rule? Are there any empirical hypotheses which are capable of being established definitively, hypotheses such that we can

49. The preceding division of the test of an empirical hypothesis into three phases may prove useful for the clarification of the question whether or to what extent an empiricist conception of confirmation implies a "coherence theory of truth." This issue has recently been raised by Bertrand Russell, who, in chap. x of his *Inquiry into Meaning and Truth*, has levelled a number of objections against the views of Otto Neurath on this subject (*cf.* the articles mentioned in the next footnote), and against statements made by myself in articles published in *Analysis* in 1935 and 1936. I should like to add here a few, necessarily brief, comments on this issue.

(1) While, in the articles in *Analysis*, I argued in effect that the only possible interpretation of the phrase 'Sentence *S* is true' is '*S* is highly confirmed by accepted observation reports', I should now reject this view. As the work of A. Tarski, R. Carnap, and others has shown, it is possible to define a semantical concept of truth which is not synonymous with that of strong confirmation, and which corresponds much more closely to what has customarily been referred to as truth, especially in logic, but also in other contexts. Thus, *e.g.*, if *S* is any empirical sentence, then either *S* or its denial is true in the semantical sense, but clearly it is possible that neither *S* nor its denial is highly confirmed by available evidence. To assert that a hypothesis is true is equivalent to asserting the hypothesis itself; therefore the truth of an empircal hypothesis can be ascertained only in the sense in which the hypothesis itself can be established: *i.e.* the hypothesis—and thereby *ipso facto* its truth—can be more or less well confirmed by empirical evidence; there is no other access to the question of the truth of a hypothesis.

In the light of these considerations, it seems advisable to me to reserve the term 'truth' for the semantical concept ; I should now phrase the statements in the *Analysis* articles as dealing with confirmation. (For a brief and illuminating survey of the distinctive characteristics of truth and confirmation, see R. Carnap, "Wahrheit and Bewährung," *Actes Ier Congrès Internat. de Philosophie Scientifique* 1935, vol. 4; Paris, 1936).

(2) It is now clear also in what sense the test of a hypothesis is a matter of confronting sentences with sentences rather than with "facts", or a matter of the "coherence" of the hypothesis and the accepted basic sentences: All the logical aspects of scientific testing, *i.e.* all the criteria governing the second and third of the three phases distinguished above, are indeed concerned only with certain relationships between the hypotheses under test and certain other sentences (namely the accepted observation reports); no reference to extra-linguistic "facts" is needed. On the other hand, the first phase, the acceptance of certain basic sentences in connection with certain experiments or observations, involves, of course, extra-linguistic procedures; but this had been explicitly stated by the author in the articles referred to before. The claim that the views concerning truth and confirmation which are held by contemporary logical empiricism involve a coherence theory of truth is therefore mistaken.

be sure that once accepted on the basis of experiential evidence, they will never have to be revoked? Hypotheses of this kind will be called absolutely or definitively verifiable; and the concept of absolute or definitive falsifiability will be construed analogously.

While the existence of hypotheses which are relatively verifiable or relatively falsifiable is a simple logical fact, which was illustrated in the beginning of this section, the question of the existence of absolutely verifiable, or absolutely falsifiable, hypotheses is a highly controversial issue which has received a great deal of attention in recent empiricist writings.[50] As the problem is only loosely connected with the subject of this essay, I shall restrict myself here to a few general observations.

Let it be assumed that the language of science has the general structure characterized and presupposed in the previous discussions, especially in section 9. Then it is reasonable to expect that only such hypotheses can possibly be absolutely verifiable as are relatively verifiable by suitable observation reports; hypotheses of universal form, for example, which are not even capable of relative verification, certainly cannot be expected to be absolutely verifiable. In however many instances such a hypothesis may have been borne out by experiential findings, it is always possible that new evidence will be obtained which disconfirms the hypothesis. Let us, therefore, restrict our search for absolutely verifiable hypotheses to the class of those hypotheses which are relatively verifiable.

Suppose now that *H* is a hypothesis of this latter type, and that it is relatively verified, *i.e.* logically entailed, by an observation report *B*, and that the latter is accepted in science as an account of the outcome of some experiment or observation. Can we then say that *H* is absolutely verified; that it will never be revoked? Clearly, that depends on whether the report *B* has been accepted irrevocably, or whether it may conceivably suffer the fate of being disavowed later. Thus the question as to the existence of absolutely verifiable hypotheses leads back to the question of whether all, or at least some, observation reports become irrevocable parts of the system of science once they have been accepted in connection with certain observations or experiments. This question is not

50. *Cf.* especially A. Ayer, *The Foundations of Empirical Knowledge* (New York, 1940); see also the same author's article, "Verification and Experience," *Proceedings of the Aristotelian Society* for 1937. R. Carnap, "Ueber Protokollsätze," *Erkenntnis*, vol. 3 (1932), and § 82 of the same author's *The Logical Syntax of Language* (New York and London, 1937). O. Neurath, "Protokollsätze," *Erkenntnis*, vol. 3 (1932); "Radikaler Physikalismus und "wirkliche Welt," *Erkenntnis*, vol. 4 (1934); "Pseudorationalismus der Falsifikation," *Erkenntnis*, vol. 5 (1935). K. Popper, *Logik der Forschung* (see note 3). H. Reichenbach, *Experience and Prediction* (Chicago, 1938), chap. iii. Bertrand Russell, *An Inquiry into Meaning and Truth* (New York, 1940), especially chaps. x and xi. M. Schlick, "Ueber das Fundament der Erkenntnis," *Erkenntnis*, vol. 4 (1934).

simply one of fact; it cannot adequately be answered by a descriptive account of the research behavior of scientists. Here, as in all other cases of logical analysis of science, the problem calls for a rational reconstruction of scientific procedure, *i.e.* for the construction of a consistent and comprehensive theoretical model of scientific inquiry, which is then to serve as a system of reference, or a standard, in the examination of any particular scientific research. The construction of the theoretical model has, of course, to take account of the characteristics of actual scientific procedure, but it is not determined by the latter in the sense in which a descriptive account of some scientific study would be. Indeed, it is generally agreed that scientists sometimes infringe the standards of sound scientific procedure; besides, for the sake of theoretical comprehensiveness and systematization, the abstract model will have to contain certain idealized elements which cannot possibly be determined in detail by a study of how scientists actually work. This is true especially of observation reports. A study of the way in which laboratory reports, or descriptions of other types of observational findings, are formulated in the practice of scientific research is of interest for the choice of assumptions concerning the form and the status of observation sentences in the model of a language of science; but clearly, such a study cannot completely determine what form observation sentences are to have in the theoretical model, nor whether they are to be considered as irrevocable once they are accepted.

Perhaps an analogy may further elucidate this view concerning the character of logical analysis: Suppose that we observe two persons whose language we do not understand playing a game on some kind of chess board; and suppose that we want to "reconstruct" the rules of the game. A mere descriptive account of the playing behavior of the individuals will not suffice to do this; indeed, we should not even necessarily reject a theoretical reconstruction of the game which did not always characterize accurately the actual moves of the players: we should allow for the possibility of occasional violations of the rules. Our reconstruction would rather be guided by the objective of obtaining a consistent and comprehensive system of rules which are as simple as possible, and to which the observed playing behavior conforms at least to a large extent. In terms of the standard thus obtained, we may then describe and critically analyze any concrete performance of the game.

The parallel is obvious; and it appears to be clear, too, that in both cases the decision about various features of the theoretical model will have the character of a convention, which is influenced by considerations of simplicity, consistency, and comprehensiveness, and not only by a study of the actual procedure of scientists at work.[51]

51. A clear account of the sense in which the results of logical analysis represent conventions can be found in §§ 9-11 and 25-30 of K. Popper's *Logik der Forschung*.

This remark applies in particular to the question here under consideration, namely whether "there are" in science any irrevocably accepted observation reports (all of whose consequences would then be absolutely verified empirical hypotheses). The situation becomes clearer when we put the question into this form: Shall we allow, in our rational reconstruction of science, for the possibility that certain observation reports may be accepted as irrevocable, or shall the acceptance of all observation reports be subject to the "until further notice" clause? In comparing the merits of the alternative stipulations, we would have to investigate the extent to which each of them is capable of elucidating the structure of scientific inquiry in terms of a simple, consistent theory. We do not propose to enter into a discussion of this question here except for mentioning that various considerations militate in favor of the convention that no observation report is to be accepted definitively and irrevocably.[52] If this alternative is chosen, then not even those hypotheses which are entailed by accepted observation reports are absolutely verified, nor are those hypotheses which are found incompatible with accepted observation reports thereby absolutely falsified: in fact, in this case, no hypothesis whatsoever would be absolutely verifiable or absolutely falsifiable. If, on the other hand, some—or even all—observation sentences are declared irrevocable once they have been accepted, then those hypotheses entailed by or incompatible with irrevocable observation sentences will be absolutely verified, or absolutely falsified, respectively.

It should now be clear that the concepts of absolute and of relative verifiability (and falsifiability) differ fundamentally from each other. Failure to distinguish them has caused considerable misunderstanding in recent discussions on the nature of scientific knowledge. Thus, *e.g.*, K. Popper's proposal to admit as scientific hypotheses exclusively sentences which are (relatively) falsifiable by suitable observation reports has been criticized by means of arguments which, in effect, support the claim that scientific hypotheses should not be construed as being absolutely falsifiable—a point that Popper had not denied. As can be seen from our earlier discussion of relative falsifiability, however, Popper's proposal to limit scientific hypotheses to the form of (relatively) falsifiable sentences involves a very severe restriction of the possible forms of scientific hypotheses.[53] In particular, it rules out all purely existential hypotheses as well as most hypotheses whose formulation requires both universal and existential quantification; and

52. *Cf.* especially the publications by Carnap, Neurath, and Popper mentioned in note 50; also Reichenbach, *loc. cit.*, section 9.

53. This was pointed out by R. Carnap; *cf.* his review of Popper's book in *Erkenntnis*, vol. 5 (1935), and "Testability and Meaning," §§ 25, 26. For a discussion of Popper's falsifiability criterion, see for example H. Reichenbach, "Ueber Induktion and Wahrscheinlichkeit," *Erkenntnis*, vol. 5 (1935); O. Neurath, "Pseudorationalismus der Falsifikation," *Erkenntnis*, vol. 5 (1935).

it may be criticized on this account, for in terms of this theoretical reconstruction of science it seems difficult or altogether impossible to give an adequate account of the status and function of the more complex scientific hypotheses and theories.

What has been said above about the nature of the logical analysis of science in general, applies to the present analysis of confirmation in particular: It is a specific proposal for a systematic and comprehensive logical reconstruction of a concept which is basic for the methodology of empirical science as well as for epistemology. The need for a theoretical clarification of that concept was evidenced by the fact that no general theoretical account of confirmation has been available so far, and that certain widely accepted conceptions of confirmation involve difficulties so serious that it might be doubted whether a satisfactory theory of the concept is at all attainable.

It was found, however, that the problem can be solved: A general definition of confirmation, couched in purely logical terms, was developed for scientific languages of a specified, relatively simple, logical character. The logical model thus obtained appeared to be satisfactory in the sense of the formal and material standards of adequacy that had been set up previously.

I have tried to state the essential features of the proposed analysis and reconstruction of confirmation as explicitly as possible in the hope of stimulating a critical discussion and of facilitating further inquiries into the various issues pertinent to this problem area. Among the open questions which seem to deserve careful consideration, I should like to mention the exploration of concepts of confirmation which fail to satisfy the general consistency condition; the extension of the definition of confirmation to the case where even observation sentences containing quantifiers are permitted; and finally the development of a definition of confirmation for languages of a more complex logical structure than that incorporated in our model.[54] Languages of this kind would provide a greater variety of means of expression and would thus come closer to the high logical complexity of the language of empirical science.

54. The languages to which our definition is applicable have the structure of the lower functional calculus without identity sign; it would be highly desirable so to broaden the general theory of confirmation as to make it applicable to the lower functional calculus with identity, or even to higher functional calculi; for it seems hardly possible to give a precise formulation of more complex scientific theories without the logical means of expression provided by the higher functional calculi.

POSTSCRIPT (1964) *ON*

CONFIRMATION

1. ON THE PARADOXES

The views expressed in my essay in regard to the paradoxes still seem sound to me: the "paradoxical" cases have to be counted as confirmatory, or positive, instances; impressions to the contrary may be attributable to factors such as those suggested in section 5.2 Several writers[1] have concurred with this estimate either fully or to a large extent.

A number of commentators[2] have argued, in a manner more or less akin to that of Mrs. Hosiasson-Lindenbaum[3], that on certain assumptions, objective logical differences can be established between paradoxical and nonparadoxical

1. Among them, H. G. Alexander, "The Paradozes of Confirmation," *The British Journal for the Philosophy of Science*, vol. 9 (1958–59), 227–33; R. Carnap. *Logical Foundations of Probability* (Chicago, 1950), 469; I. J. Good,"The Paradox of Confirmation,"Parts I and II, *The British Journal for the Philosophy of Science*, vol. 11 (1960), 145–48; vol. 12 (1961), 63–64; N. Goodman, *Fact, Fiction, and Forecast* (Cambridge, Mass., 1955), pp. 69–73; J. L. Mackie, "The Paradoxes of Confirmation," *The British Journal for the Philosophy of Science*, vol. 13 (1963), 265–77; I. Scheffier, *The Anatomy of Inquiry* (New York, 1963), Part III. Critical questions have been raised, in the name of Popper's anti-inductivism, for example by J. W. N. Watkins, "Between Analytic and Empirical," *Philosophy*, vol. 32 (1957), 112-31, and "A rejoinder to Professor Hempel's Reply," *Philosophy*, vol. 33 (1958), 349-55; J. Agassi, "Corroboration versus Induction," *The British Journal for the Philosophy of Science*, vol. 9 (1959), 311-17. For a discussion of these and other strictures see Alexander, *loc. cit.*; Hempel, "A Note on the Paradoxes of Confirmation," *Mind*, vol. 55 (1946), 79-82 and "Empirical Statements and Falsifiability," *Philosophy*, vol. 33 (1958), 342-48; Mackie, *loc. cit.*; Scheffler, *loc. cit.*; R. H. Vincent, "The Paradoxes of Confirmation," *Mind*, vol. 73 (1964), 273-79.

2. Among them, Alexander, *loc. cit.*; Good, *loc. cit.*; D. Pears, "Hypotheticals," *Analysis*, vol. 10 (1950), 49-63; G. H. von Wright, *The Logical Problem of Induction* (Oxford, 1957), pp. 122-27.

3. See note 25 of the preceding essay.

instances of generalizations of the form 'All *P*'s are *Q*'s'. The principal requisite assumption is to the effect that there are many more non-*Q*'s than *P*'s (or alternatively, that the probability of an object being a non-*Q* is much greater than that of its being a *P*). Several writers presuppose in addition a suitable theory of degrees of confirmation or inductive probabilities, and some also assume that the generalization has a positive initial probability. On such assumptions it is then argued that, for example, examining a nonblack thing for nonravenhood involves much less risk of refuting the generalization 'All ravens are black' than does examining a raven for blackness, and that a positive outcome of the former kind of test has therefore much less importance or weight than a positive outcome of the latter (thus Pears, who does not invoke a theory of degrees of confirmation); or that an instance of a paradoxical kind will increase the prior probability of the generalization by much less than a nonparadoxical one.

Some of these arguments seem to me open to questions such as those suggested in note 25 of my essay. But—and this is the essential point—even if satisfactorily established, such differences in degree between paradoxical and nonparadoxical instances clearly do not refute my diagnosis of the paradoxical cases as confirmatory. My essay is concerned exclusively with the classificatory or qualitative concept of confirmation, and it does not claim that the different kinds of positive instance are all confirmatory to the same degree or that they carry the same weight in testing a generalization.

As for the pragmatic question of why paradoxical cases appear to be nonconfirmatory, Pears[4] may well be right in suggesting that those descriptive words (e.g. 'raven', 'black') which we normally use to formulate our generalizations pick out classes that satisfy (perhaps, better, that are commonly believed to satisfy) the crucial assumption about relative size, and that this in turn explains, in virtue of the kind of argument mentioned before, why paradoxical instances "are thought to provide less confirmation" than nonparadoxical ones. Indeed, as Mackie[5] suggests, it might even explain why to some persons the finding of a nonblack thing that is not a raven seems not to be evidentially relevant at all. This may well constitute a further factor, different from those suggested in section 5.2 of my article, that partly contributes to the impression of paradoxicality.[6]

2. ON THE GENERAL DEFINITION OF CONFIRMATION

My general formal definition of qualitative confirmation now seems to me

4. Pears, *loc. cit.*, pp. 51-52.—This was suggested also by Miss Hosiasson-Lindenbaum in footnote 11 of her article.

5. Mackie, *loc. cit.*, pp. 266-67.

6. *Cf.* also the lucid discussion of these issues by Scheffler, *loc. cit.*

rather too restrictive. Here are some of the reasons for this appraisal, in order of increasing importance:

(a) Some hypotheses of the kind covered by my definition, though logically consistent, are not capable of confirmation by any logically consistent observation report. For example, a hypothesis of the form

$$(x)\,(\exists y)Sxy \cdot (x)(y)(z)\,[(Sxy \cdot Syz) \supset Sxz] \cdot (x) \sim Sxx$$

can be satisfied only in an infinite domain; its development for any finite class of objects is self-contradictory. Generally, no scientific hypothesis that implies the existence of infinitely many objects can, on my definition, be confirmed by any observation report. This seems worth noting, but it surely constitutes no serious shortcoming of the definition.

(b) My definition qualifies as neutral certain kinds of evidence that would normally be regarded as confirmatory. Thus, as Canfield[7] has pointed out, no finite set of sentences of the type

$$Rab,\ Rbc,\ Rcd,\ Rde, \ldots$$

qualifies as confirming the hypothesis

$$H_1: (x)(y)Rxy$$

A report that mentions just the individuals *a* and *b*, for example, confirms H_1 only if it implies the development of H_1 for the class $\{a,b\}$, *i.e.*, the sentence

$$Raa \cdot Rab \cdot Rba \cdot Rbb$$

And as the number of individuals mentioned in an observation report increases, the condition the report has to meet if it is to confirm H_1 becomes increasingly stringent. Analogous remarks apply to the case of disconfirmation.

(c) Some writers[8] have argued that the consistency condition for confirmation is too strong, for a reason I had considered, but then set aside, in my comments on that condition in section 8: One and the same observable phenomenon may well be accounted for by each of two incompatible hypotheses, and the observation report describing its occurrence would then normally be regarded as confirmatory for either hypothesis. This point does seem to me to carry considerable weight; but if it is granted, then the consequence condition has to be given up along with the consistency condition. Otherwise, a report confirming each of two incompatible hypotheses would count as confirming any consequence of the two, and thus any hypothesis whatsoever.

7. J. Canfield, "On the Paradox of Confirmation," *Metrika*, vol. 5 (1962), 105-18.

8. Particularly Carnap in his detailed exposition and critical analysis of my essay, in sections 87, 88 of *Logical Foundations of Probability* (*cf.* especially pp. 476-78). See also the comment in K. Popper, *The Logic of Scientific Discovery* (London, 1959), p. 374.

For the reasons here briefly surveyed, I believe Carnap is right in his estimate that the concept of confirmation defined in my essay "is not clearly too wide but is clearly too narrow."[9] Accordingly, I think that the criteria specified in my definition may be sufficient, but are not necessary for the confirmation of a hypothesis H by an observation report B.

Perhaps the problem of formulating adequate criteria of qualitative confirmation had best be tackled, after all, by means of the quantitative concept of confirmation. This has been suggested especially by Carnap, who holds that "any adequate explicatum for the classificatory concept of confirmation must be in accord with at least one adequate explicatum for the quantitative concept of confirmation"; *i.e.*, there must be at least one function c that is a suitable explicatum for the concept of logical probability such that whenever B qualitatively confirms H, then $c(H,B) > c(H,t)$, where t is the tautological, or null, evidence.[10] In other words: on some suitable definition of logical probability, the probability of H on B should exceed the *a priori* probability of H whenever B qualitatively confirms H.[11] This general principle leads Carnap also to reject the consequence condition for qualitative confirmation and to restrict the entailment condition to the case where H is not a logical truth.

Finally, I shall discuss quite a different aspect of the problem. In accordance with the objective stated toward the end of section 6, my definition of confirmation is purely syntactical, since for the formalized languages in question the concept of logical consequence, which occurs in the definiens, is characterizable in purely syntactical terms, as are all other concepts used in the definition. But confirmation—whether in its qualitative or in its quantitative form—cannot be adequately defined by syntactical means alone. That has been made clear especially by Goodman,[12] who has shown that some hypotheses of the form '$(x)(Px \supset Qx)$' can obtain no confirmation at all even from evidence sentences of the form '$Pa \cdot Qa$'. To illustrate this, I will adapt Goodman's example to my ornithological paradigm. Let 'x is P' stand for 'x is a raven' and 'x is Q' for 'x is blite', where an object is said to be blite if it has been examined before a certain time t and is black or has not been examined before t and is white. Then any raven observed before t and found to be black affords a formally confirming instance, in the sense of Nicod's criterion, of the hypothesis 'All ravens are blite'.

9. Carnap, *loc. cit.*, p. 479.

10. Carnap, *loc. cit.*, p. 472.

11. As noted by Mackie, several other writers construe confirmation rather in accordance with "the *Inverse Principle*, that a hypothesis h is confirmed by an observation-report b in relation to background knowledge if and only if the observation-report is made more probable by the adding of the hypothesis to the background knowledge" (*loc. cit.*, p. 267; author's italics).

12. Goodman, *loc. cit.*, chapters III and IV.

Yet no matter how many such instances may have been collected, they lend no support or confirmation to the hypothesis; for the latter implies that all ravens not examined before t—hence in particular all those that might be examined after t—are white, and this consequence must surely count as disconfirmed rather than as confirmed. Whether a universal conditional hypothesis is capable of being confirmed by its positive instances, whether it can be "projected," as Goodman says, from examined cases to unexamined ones, will depend on the character of its constituent predicates; use of the predicate 'blite', for example, precludes projectibility. Goodman traces the difference between predicates that can occur in projectible hypotheses and those that cannot to their "entrenchment," *i.e.*, the extent to which they (or predicates coextensive with them) have been used in previously projected generalizations; 'blite,' for example, never having been so used, is much less well entrenched than such terms as 'black', 'white', and 'raven'. By reference to the comparative entrenchment of the constituent predicates, Goodmen formulates criteria for the comparative projectibility of universal conditional hypotheses, and thus also for their susceptibility to confirmation by formally positive instances.

Thus the search for purely syntactical criteria of qualitative or quantitative confirmation presupposes that the hypotheses in question are formulated in terms that permit projection; and such terms cannot be singled out by syntactical means alone. Indeed, the notion of entrenchment that Goodman uses for this purpose is clearly pragmatic in character.