P. K. FEYERABEND

Explanation, Reduction, and Empiricism

The main contention of the present paper is that a formal account of reduction and explanation is impossible for general theories, or noninstantial theories,[1] as they have also been called. More especially, it will be asserted and shown that wherever such theories play a decisive role both Nagel's theory of reduction[2] and the theory of explanation associated with Hempel and Oppenheim[3] cease to be in accordance with actual scientific practice and with a reasonable empiricism. It is to be admitted that these two "orthodox" accounts fairly adequately represent the relation between sentences of the 'All-ravens-are-black' type, which abound in the more pedestrian parts of the scientific enterprise.[4] But if the attempt is made to extend these accounts to such comprehensive structures of thought as the Aristotelian theory of motion, the impetus theory, Newton's celestial mechanics, Maxwell's electrodynamics, the theory of relativity, and the quantum theory, then complete failure is the result. What happens *here* when transition is made from a theory T′ to a wider theory T (which, we shall assume, is capable of covering all the phenomena

[1] In what follows, the usual distinction will be drawn between *empirical generalizations*, on the one side, and *theories*, on the other. Empirical generalizations are statements, such as 'All A's are B's' (the A's and B's are not necessarily observational entities), which are tested by inspection of instances (the A's). Universal theories, such as Newton's theory of gravitation, are not tested in this manner. Roughly speaking their test consists of two steps: (1) derivation, with the help of suitable boundary conditions, of empirical generalizations; and (2) tests, in the manner indicated above, of these generalizations. One should not be misled by the fact that universal theories, too, can be (and usually are) put in the form 'All A's are B's'; for, whereas, in the case of generalizations, this form reflects the test procedure in a very direct way, such an immediate relation between the form and the test procedures does not obtain in the case of theories. Many thinkers have been seduced by the similarity of form into thinking that the test procedures will be the same in both cases.

[2] Nagel has explained his theory in [60]. I shall quote from the reprint of the article in [20], pp. 288–312.

[3] For the theory of Hempel and Oppenheim see [47]. I shall quote from the reprint in [23], pp. 319–352.

[4] For important exceptions, see fn. 90.

that have been covered by T′) is something much more radical than incorporation of the *unchanged* theory T′ (unchanged, that is, with respect to the meanings of its main descriptive terms as well as to the meanings of the terms of its observation language) into the context of T. What does happen is, rather, a *complete replacement* of the ontology (and perhaps even of the formalism) of T′ by the ontology (and the formalism) of T and a corresponding change of the meanings of the descriptive elements of the formalism of T′ (provided these elements and this formalism are still used). This replacement affects not only the theoretical terms of T′ but also at least some of the observational terms which occurred in its test statements. That is, not only will description of things and processes in the domain in which so far T′ had been applied be infiltrated, either with the formalism and the *terms* of T, or if the terms of T′ are still in use, with the *meanings* of the terms of T, but the sentences expressing what is accessible to direct observation inside this domain will now mean something different. In short: introducing a new theory involves changes of outlook both with respect to the observable and with respect to the unobservable features of the world, and corresponding changes in the meanings of even the most "fundamental" terms of the language employed. So far this is the position which will be defended in the present paper.

This position may be said to consist of two ideas. The first idea is that the influence, upon our thinking, of a comprehensive scientific theory, or of some other general point of view, goes much deeper than is admitted by those who would regard it as a convenient scheme for the ordering of facts only. According to this first idea scientific theories are ways of looking at the world; and their adoption affects our general beliefs and expectations, and thereby also our experiences and our conception of reality. We may even say that what is regarded as "nature" at a particular time is *our own product* in the sense that all the features ascribed to it have first been invented by us and then used for bringing order into our surroundings. As is well known, it was Kant who most forcefully stated and investigated this all-pervasive character of theoretical assumptions. However, Kant also thought that the very generality of such assumptions and their omnipresence would forever prevent them from being refuted. As opposed to this, the second idea implicit in the position to be defended here demands that our theories be *testable* and that they be abandoned as soon as a test does not produce the predicted result. It is this second idea which makes sciences proceed to better and better theories and which creates the changes described in the introductory paragraphs of the present paper.

Now, it is easily seen that the mere statement of the second idea will not do. What we need is a guarantee that despite the all-pervasive character of a scientific theory as it is asserted in the first idea, it is still possible to specify facts that are inconsistent with it. Such a possibility has been denied by some philosophers. These philosophers started out by reacting against the claim that scientific theories are nothing but predictive devices; they recognized that their influence goes much deeper; however, they then doubted that it would be possible ever to get outside any such theory; and they therefore either became apriorists (Poincaré, Eddington), or they returned to instrumentalism. For these thinkers there seemed to exist only a choice between two evils—instrumentalism or apriorism.

Now, a closer look at the arguments leading up to this dilemma shows that they all proceed from a test model in which a *single* theory is confronted with the facts. As soon as this model is replaced by a model in which we make use of at least two factually adequate but mutually inconsistent theories, the first idea becomes compatible with the demand for testability which must now be interpreted as a demand for crucial tests either between two explicitly formulated theories, or between a theory and our "background knowledge." In this form, however, the test model turns out to be inconsistent with the "orthodox" theory of explanation and reduction. It is one of the aims of the present paper to exhibit this inconsistency.

It will be necessary, for this purpose, to discuss two principles which underlie the orthodox approach: (A) the principle of deducibility; and (B) the principle of meaning invariance. According to the principle of deducibility, explanation is achieved by deduction in the strict logical sense. This principle leads to the demand, which is incompatible with the test model just outlined, that all successful theories in a given domain must be mutually consistent. According to the principle of meaning invariance, an explanation must not change the meanings of the main descriptive terms of the explanandum. This principle, too, will be found to be inconsistent with empiricism.

It is interesting to note that (A) and (B) play a role both within modern empiricism and within some very influential "school philosophies." Thus it is one of the basic assumptions of Platonism that the key terms of sentences expressing knowledge (epistēmē) refer to unchangeable entities and must therefore possess a stable meaning. Similarly, the key terms of Cartesian physics—i.e., the terms 'matter,' 'space,' 'motion'—and the terms of Cartesian metaphysics—such as the terms 'god' and 'mind'—are supposed to remain unchanged in any explanation involving them. Compared with these similarities[5] between the school philosophies, on the one side, and modern empiricism, on the other, the differences are of very minor importance. These differences lie in the *terms* of which stability of meaning is required. A Platonist will direct his attention to numbers and other "ideas," and he will demand that words referring to these entities retain their (Platonic) meanings. Modern empiricism, on the other hand, regards empirical terms as fundamental and demands that their meanings remain unchanged.

Now, it will turn out, in the course of this essay, that any form of meaning invariance is bound to lead to difficulties when the task arises either to give a proper account of the growth of knowledge, and of discoveries contributing to this growth, or to establish correlations between entities which are described with the help of what we shall later call incommensurable concepts. It will also turn out that these are exactly the difficulties we encounter in trying to solve such age-old problems as the mind-body problem, the problem of the reality of the external world, and the problem of other minds. That is, it will usually turn out that a solution of these problems is deemed satisfactory only if it leaves unchanged the meanings of certain key terms and that it is exactly *this* condition, i.e., the condition of meaning invariance, which makes them insoluble. It will also be shown that the demand for meaning invariance is incompatible with empiricism. Taking all this into account, we may hope that once contemporary empiricism has been freed from the elements which it still shares with its more dogmatic opponents, it will be able to make swift progress in the solution of the above problems. It is the purpose of the present paper to develop and to defend the outlines of such a disinfected empiricism.[6]

Popper's admirable *Logic of Scientific Discovery* and his paper "The

[5] Concerning these similarities, see Popper's discussion of essentialism in [65], Ch. III and *passim*, as well as Dewey's very different account in [21], especially Ch. II.

[6] As will be shown in Sec. 2, the empiricism of the thirties *was* disinfected in the sense desired here. However, later on modern empiricism readopted some very undesirable principles of traditional philosophy.

Aim of Science"[7] have been both the starting point and the motive force of the investigations to follow. I have also profited a great deal from discussions with Professors Bohm (Bristol-Haifa), Feigl (Minneapolis), Körner (Bristol), Maxwell (Minneapolis), Putnam (Princeton), and Tranekjaer-Rasmussen (Copenhagen). Both Professor Körner and Professor Sellars (New Haven) seem to hold similar views with respect to the character of the observation language, and reading their publications has therefore been a great help.[8]

While the present paper was in progress I had an opportunity to consult various as yet unpublished papers by Professor T. S. Kuhn (Berkeley) in which the noncumulative character of scientific progress is illustrated very forcefully by historical examples. Despite some important and perhaps unalterable differences, the area of agreement between Professor Kuhn and myself seems to be quite considerable. One most important point of agreement is the emphasis which both of us put upon the need, in the process of the refutation of a theory, for at least another theory. As far as I am aware, this point has been made previously by K. R. Popper in his lectures on scientific method which I attended in 1948 and 1952. Popper has also pointed out[9] that the alternative theory used in the process of refutation need not be explicitly stated but can be part of our "background knowledge."

Bohm's theory of levels and Putnam's considerations in the present volume seem to lead in the same direction. What I regard as a most important feature of the situation—a feature, by the way, that has been emphasized by Bohm and Vigier—is that direct refutation of a fairly complicated theory may be impossible for *empirical* reasons. That this is so will be shown with the help of an example. Finally, I would like to thank Professor Popper and Mr. J. W. N. Watkins (London) for constructive criticism that has been utilized in the final version of the paper.

1. *Two Assumptions of Contemporary Empiricism.*

Nagel's theory of reduction is based upon two assumptions. The first assumption concerns the relation between the secondary science, i.e., the

[7] See [68] and [66]. The basic ideas of [66] can be found in [64], which was written earlier.

[8] I am referring here to Körner's [50] and to Sellars' admirable [70].

[9] In a lecture given at Stanford University in September 1960.

discipline to be reduced, on the one side, and the primary science, i.e., the discipline to which reduction is made, on the other. It is asserted that this relation is the relation of deducibility. Or, to quote Nagel,

(1) "The objective of the reduction is to show that the laws, or the general principles of the secondary science, are simply logical consequences of the assumptions of the primary science."[10]

The second assumption concerns the relation between the meanings of the primitive descriptive terms of the secondary science and the meanings of the primitive descriptive terms of the primary science. It is asserted that the former will not be affected by the process of reduction. Of course, this second assumption is an immediate consequence of (1), since a derivation is not supposed to influence the meanings of the statements derived. However, for reasons which will become clear later, it is advisable to formulate this invariance of meaning as a separate principle. This is also done by Nagel, who says: "It is of the utmost importance to observe that the expressions peculiar to a science will possess meanings that are fixed by its *own* procedures, and are therefore intelligible in terms of its own rules of usage, *whether or not the science has been, or will be, reduced to some other discipline*."[11] Or, to express it in a more concise manner:

(2) Meanings are invariant with respect to the process of reduction.

(1) and (2) admit of two different interpretations, just as does any theory of reduction and explanation: such a theory may be regarded either as a *description* of actual scientific practice, or as a *prescription* which must be followed if the scientific character of the whole enterprise is to be guaranteed. Similarly, (1) and (2) may be interpreted as *assertions* concerning actual scientific practice, or as *demands* to be satisfied by the theoretician who wants to follow the scientific method. Both of these interpretations will be scrutinized in the present paper.

Two very similar assumptions, or demands, play a decisive role in the orthodox theory of explanation, which may be regarded as an elaboration of suggestions that were first made, in a less definite form by Popper.[12] The first assumption (demand) concerns again the relation

[10] [20], p. 301. A more elaborate form of this condition is called the "condition of derivability" on p. 354 of [61].

[11] [20], p. 301. My italics. See also [61], p. 345, 352.

[12] [68], Sec. 12.

between the explanandum, or the laws, or the facts to be explained, on the one side, and the explanans, or the discipline which functions as the basis of explanation, on the other. It is again asserted (required) that this relation is (be) the relation of deducibility. Or, to quote Hempel and Oppenheim

(3) "The explanandum must be a logical consequence of the explanans; in other words, the explanandum must be logically deducible from the information contained in the explanans, for otherwise the explanans would not constitute adequate grounds for the explanation."[13]

Considering what has been said in the case of reduction one would expect the assumption (demand) concerning meanings to read as follows:

(4) Meanings are invariant with respect to the process of explanation.

However, despite the fact that (4) is a trivial consequence of (3), this assumption has never been expressed in as clear and explicit a way as (2).[14] There was even a time when a consequence of (4), viz., the assertion that *observational* meanings are invariant with respect to the process of explanation, seemed to be in doubt. It is for this reason that I have separated (2) from (1), and (4) from (3).

It is not difficult to show that, with respect to observational terms, (4) or its implications, is consistent with the earlier positivism of the Vienna Circle. Their main thesis that all descriptive terms of a scientific theory can be explicitly defined on the basis of observation terms guarantees the stability of the meanings of observational terms (unless one assumes that an explicit definition changes the meaning of the definiens, a possibility that to my knowledge has never been considered by empiricists). And as the chain of definitions leaves unchanged terms already defined, (4) turns out to be correct as well.

However, since these happy and carefree days of the *Aufbau*, logical

[13] [47], p. 321.

[14] An exception is Nagel who, in [61], p. 338, defines reduction as "the explanation of a theory or a set of experimental laws established in one area of inquiry by a theory usually, though not invariably, formulated for some other domain." This implies that the condition of meaning invariance formulated by him for the process of reduction is supposed to be valid in the case of explanation also. On pp. 86–87, meaning invariance for observational terms is stated quite explicitly: an experimental law "retains a meaning that can be formulated independently of [any] theory . . . [It] has . . . a life of its own, not contingent on the continued life of any particular theory that may explain the law."

empiricism has been greatly modified. The changes that took place were mainly of two kinds. On the one side, new ideas were introduced concerning the relation between observational terms and theoretical terms. On the other side, the assumptions made about the observational language itself were modified. In both cases the changes were quite drastic. For our present purpose a brief outline must suffice: The early positivists assumed that observational terms refer to subjective impressions, sensations, and perceptions of some sentient being. Physicalism for some time retained the idea that a scientific theory should be based upon experiences, and that the ultimate constituents of experience were sensations, impressions, and perceptions. Later, however, a behavioristic account was given of these perceptions to make them accessible to intersubjective testing. Such a theory was held, for some time, by Carnap and Neurath.[15] Soon afterwards the idea that it is *experiences* to which we must refer when trying to interpret our observation statements was altogether abandoned.[16] According to Popper, who has been responsible for this decisive turn, we must "*distinguish sharply between objective science on the one hand, and 'our knowledge' on the other.*" It is conceded "we can become aware of facts only by observation"; but it is denied that this implies an interpretation of observation sentences in terms of experiences, whether these experiences are explained subjectivistically or as features of objective behavior.[17] For example, we may admit that the sentence 'this is a raven' uttered by an observer who points at a bird in front of him is an observational sentence and that the observer has produced it because of the impressions, sensations, and perceptions he possesses. We may also admit that he would not have uttered the sentence had he not possessed these impressions. Yet, the sentence is not therefore about impressions; it is about a bird which is neither a sensation nor the behavior of some sentient being. Similarly, it may be admitted that the observation sentences which a scientific observer produces are prompted by his impressions. However, their content will again be determined, not by these impressions, but by the entities allegedly described. In the case of classical physics, therefore, "every basic statement must either be itself a statement about relative positions

[15] For this and the following, see Carnap's account in [13], especially the passages in small print on pp. 223–224.

[16] *Ibid.*, p. 223: "It is stipulated that under given circumstances *any* concrete statement may be regarded as a protocol statement."

[17] Popper [68], p. 98. Italics in the original.

of physical bodies . . . or it must be equivalent to some basic statement of this 'mechanistic' . . . kind."[18]

The descriptive terms of Carnap's "thing-language," too, no longer refer to experiences. They refer to properties of objects of medium size which are accessible to observation, i.e., which are such that a normal observer can quickly decide whether or not an object possesses such a property.[19] "What we have called observable predicates," says Carnap, "are predicates of the thing-language (they have to be clearly distinguished from what we have called perception terms . . . whether these are now interpreted subjectivistically, or behavioristically)."[20]

Now it is most important to realize that the characterization of observation statements implicit in the above quotations is a *causal* characterization, or if one wants to use more recent terminology, a *pragmatic* characterization:[21] an observation sentence is distinguished from other sentences of a theory, not, as was the case in earlier positivism, by its *content*; but by the *cause* of its production, or by the fact that its production conforms to certain *behavioral patterns*.[22] This being the case, the fact that a certain sentence belongs to the observation language does not allow us to infer anything about its content; more especially, it does not allow us to make any inference concerning the *kind* of entities described in it.

It is worthwhile to dwell a little longer on the features of this *pragmatic theory of observation*, as I shall call it. In the case of measuring instruments, the pragmatic theory degenerates into a triviality: nobody would ever dream of asserting that the way in which we interpret the movements of, say, the hand of a voltmeter is uniquely determined either by the character of this movement itself or by the processes inside the instrument; a person who can see and understand only these processes will be unable to infer that what is indicated is voltage, and he will be equally unable to understand what voltage is. Taken by themselves the indications of instruments do not mean anything unless we possess a *theory* which teaches us what situations we are to expect in

[18] Popper [68], p. 103. Popper himself does not restrict his characterization to the observation statements of classical physics.

[19] Carnap [14], p. 63, Explanation 1. Page references are to the reprint of this article in [22], pp. 47–92.

[20] *Ibid.*, p. 69.

[21] For this terminology see Morris [59], pp. 6ff.

[22] See again Explanation 1 of [14], as well as my elaboration of this explanation in [31].

the world, and which guarantees that there exists a reliable correlation between the indications of the instrument and such a particular situation. If a certain theory is replaced by a different theory with a different ontology, then we may have to revise the interpretation of *all* our measurements, however self-evident such a particular interpretation may have become in the course of time: according to the phlogiston theory, measurements of weight before and after combustion are measurements of the amount of phlogiston added or lost in the process. Today we must give a completely different interpretation of the results of these measurements. Again, Galileo's thermoscope was initially supposed to measure an intrinsic property of a heated body; however, with the discovery of the influence of atmospheric pressure, of the expansion of the substance of the thermoscope (which, of course, was known beforehand), and of other effects (nonideal character of the thermoscopic fluid), it was recognized that the property measured by the instrument was a very complicated function of such an intrinsic property, of the atmospheric pressure, of the properties of the particular enclosure used, of its shape, and so on.[23] Indeed, the point of view outlined in the beginning of the present paper gives an excellent account of the way in which results of measurement, or indications of instruments, are reinterpreted in the light of fresh theoretical insight. Nobody would dream of using the insight given by a new theory for the readjustment of some general beliefs only, leaving untouched the interpretation of the results of measurement. And nobody would dream of demanding that the meanings of observation statements *as obtained with the help of measuring instruments* remain invariant with respect to the change and progress of knowledge. Yet, precisely this is done when the measuring instrument is a human being, and the indication is the behavior of this human being, or the sensations he has, at a particular time.

It is not easy to set down in a few lines the reasons for this exceptional treatment of human observers. Nor is it possible to criticize them thoroughly and thereby fully pave the way for the acceptance of the pragmatic theory of observation. However, such a comprehensive criticism is not really necessary here. It was partly given by those very same philosophers who are responsible for the formulation of the pragmatic

[23] For historical references, see [18], especially the articles on the phlogiston theory (J. B. Conant) and on the early development of the concept of temperature (D. Roller).

theory[24] (which most of them dropped later on, their own excellent arguments in favor of it notwithstanding). I shall therefore content myself with giving an outline of the idea leading to the assumption that human observers are something special and cannot be treated in the same manner as physical measuring instruments.

These ideas are connected with the (very old) belief that (a) some states of the mind (sensations or abstract ideas) can be known with certainty; that (b) it is exactly this knowledge that constitutes the foundation of whatever assertion we make about the world; and that (c) meaning invariance is obtained in the following manner: if it is indeed the case that statements about, say, sensations are irrevocable once produced, then the same applies to the descriptive terms contained in them; their meaning is uniquely and irrevocably determined by the structure of the statements in which they occur as well as by the circumstances which lead to the certain production of these statements. (Similar considerations apply if we are dealing, not with sensations, but with the 'clear and distinct' appearance of ideas.)

The theories behind meaning invariance are, of course, a little more complicated than I have just indicated, and they should perhaps be outlined in greater detail in order that their force be duly realized. Nevertheless, their most fundamental assumptions—viz., (a), (b), and (c)—can be eliminated on the basis of some very simple and almost trivial considerations. These considerations, which cannot be found in the above-mentioned writings of the original defenders of the pragmatic theory, proceed from the remark that in the argument leading up to (c) the distinction is obliterated between (psychological and sociological) *facts* and (linguistic) *conventions*.[25] It is assumed that the urge we feel under certain circumstances to say 'I am in pain' and the peculiar character of this urge (it is different from the urge we feel when we say 'I am hungry') *already determine* the meaning of the main descriptive term of the sentence uttered, viz., the meaning of the term 'pain' or 'hunger').

Now, quite apart from leading to some very undesirable paradoxes[26] this procedure assumes that a *fact*, viz., the existence of either an urge

[24] Cf. Carnap [11] and [12].

[25] For a very clear presentation of this distinction, see Popper [65], Ch. V.

[26] For a more detailed discussion of these paradoxes, see my paper [31], especially Secs. 4 and 5.

to produce a sentence of a certain kind or the existence of psychological phenomenon, can without further ado transfer meaning upon a sentence, viz., the sentence 'I am in pain.' It is therefore unacceptable to any philosopher who takes seriously the distinction between facts and conventions. Conversely, the attempt to uphold this distinction leads at once to the separation, characteristic of the pragmatic theory, of the observational character of a statement from its meaning: according to the pragmatic theory, the fact that a statement belongs to the observational domain has no bearing upon its meaning. Even if its production is accompanied by very forceful sensations and related to them in a manner that makes substitution by a different sentence psychologically very difficult or perhaps even impossible, even then we are still free to interpret the sentence in whatever way we like. It is very important to point out that this freedom of interpretation obtains also in psychology, where our sentences are indeed about subjective events. Whatever restrictions of interpretation we accept are determined by the language we use, or by the theories or general points of view whose development has led to the formulation of this language.[27]

To repeat: strict adherence to the distinction between nature and convention at once eliminates the third of the three assumptions mentioned above and thereby introduces a very fundamental element of the pragmatic theory, namely, its emphasis upon the separation between observability and meaning. However, we cannot retain the first assumption either. The reason is that the sciences are the result of a decision to use only testable statements for the expression of laws and singular facts. This being the case we cannot admit any irrefutable statement,

[27] A pointed criticism of the idea that the interpretation of a statement is uniquely determined by the sensations that accompany its production has been given by Wittgenstein in his [74]. This book also emphasizes the dependence of interpretations upon the incorporation of the corresponding sentence into a language. It does not seem to me, however, that Wittgenstein possesses a clear idea of what we have called the pragmatic theory of observation. He fails to recognize that languages are not ends in themselves but are means for expressing *theories* and that they can and should be abandoned as soon as the corresponding theory has been refuted. Quite to the contrary—he dwells upon the difficulties one will encounter when trying to change a language in a very fundamental way, and he thereby insinuates that it may be altogether impossible to carry out decisive changes. The reason for this pessimism seems to be identical with the one I briefly discussed in the introductory part of this paper: it is assumed that the all-pervasive character of a language makes it impossible to specify grounds for abandoning it. For an application of this pessimism to more concrete problems, see Hanson [43], especially Chs. III and V. For a criticism, see my review of Hanson in [35].

however elevated and noble its source may seem to be.[28] Indeed, a whole theory may at some time turn out to be unsatisfactory, and the need may arise to replace it by a completely different idiom based upon a different point of view. Clearly, then, the interpretation of the observation sentences will have to follow suit, for again there is no way of conferring an interpretation upon them except by incorporating them into a new and better theory.

The pragmatic theory of observation thus turns out to be a presupposition of the feasibility of the point of view which has been outlined in my introductory remarks (and a consequence of the distinction between nature and convention). This position, this point of view, and especially the idea that our theories determine our entire conception of reality now emerge as a combination of (a) the demand to apply the terminology and the ontology of a given theory everywhere inside the domain of its validity with (b) the pragmatic theory of observation. It is in this form that I shall defend my position in the present paper.

The freedom of interpretation admitted by the pragmatic theory did not exist in the earlier positivism. Here sensations were thought to be the objects of observation. According to it, whether or not a statement is a sense-datum statement and, therefore, part of the observation language could be determined by logical analysis. Conversely, the assertion that a certain statement belongs to the observation language there implied an assertion about the kind of entities described (e.g., sense data). The ontology of the observational domain was therefore fixed independently of theorizing. This being the case, the demand for a unified ontology (which is still retained) could be met only by adopting the one or the other of the following two procedures: it could be met by either denying a descriptive function to the sentences of the theory and by declaring that these sentences are nothing but part of a complicated prediction machine (*instrumentalism*), or by conferring upon these sentences an interpretation that completely depends upon their connection with the observational language as well as upon the (fixed) interpretation of the latter (*reductionism*). It is important to realize that it is the clash between realism, on the one side, and the combination of the theory of sense data with the demand for a unified ontology, on the

[28] In [33] I discuss some of the consequences of the use of irrefutable statements of observation and thereby provide some reasons for their elimination from the body of the sciences and of knowledge in general.

other, which necessitates this transition to either instrumentalism or reductionism.

Now one of the most surprising features of the development of contemporary empiricism is that the very articulate formulation of the pragmatic account of observation was not at once followed by an equally articulate formulation of a realistic interpretation of scientific theories. After all, realism had been abandoned mainly because the theory of sense data had made it incompatible with the demand for a unified ontology. The arrival of the pragmatic theory of observation removed this incompatibility and thereby opened the way for a hypothetical realism of the kind outlined earlier. Yet, in spite of this possibility, the actual historical development was in a completely different direction. The pragmatic theory was retained for a while (and is still retained, in footnotes, by some empiricists[29]), but it was soon combined either with instrumentalism or with reductionism. As the reader can verify for himself, such a combination in effect amounts to abandoning the pragmatic theory, a more complicated language with a more complicated ontology now taking the place of the sense-data language of the earlier point of view. How close the most recent offspring of this development is to the old sense-data ideology may be seen in a recent paper by Professor Carnap.

In this paper Carnap analyzes scientific theories with the help of his well-known double-language model consisting of an observational language, L_O and a theoretical language, L_T, the latter containing a postulate system, T. The languages are connected to each other by correspondence rules, i.e., by sentences containing observational terms and theoretical terms. With respect to such a system, Carnap asserts that "there is no independent interpretation for L_T. The system T is itself an uninterpreted postulate system. The terms of [L_T] obtain only an indirect and incomplete interpretation by the fact that some of them are connected by correspondence rules with observational terms, and the remaining terms of [L_T] are connected with the first ones by the postulates of T."[30]

This procedure quite obviously presupposes that the meaning of the observational terms is fixed independently of their connection with theoretical systems. If the pragmatic theory of observation were still retained in this essay of Carnap's, then the interpretation of an observational

[29] See Hempel [46], especially fn. 10.
[30] See Carnap [15], p. 47.

statement would have to be independent of the behavioral pattern exhibited in the observational situation as well. It is not clear how, then, the observation sentence could be given any meaning at all. Now, Carnap is very emphatic about the fact that incorporation into a theoretical context is not sufficient for providing an interpretation, since no theoretical context possesses an "independent interpretation."[31] We must, therefore, suspect that, for Carnap, incorporation of a sentence into a complicated behavioral pattern has implications for its meaning, i.e., we must suspect that Carnap has silently dropped the pragmatic theory. This is indeed the case. He asserts that "a complete interpretation of L_O" is given since "L_O is used by a certain language community as a means of communication,"[32] adding in a later passage[33] that if people use a term in such a fashion that for some sentences containing the term "any possible observational result can never be absolutely conclusive evidence, but at best evidence yielding a high probability, then the appropriate place for [the term] in a dual language system . . . is in L_T rather than in L_O . . ." These two passages together seem to imply that the meaning of an observational statement is already fixed by the way in which the sentence expressing it is handled in the *immediate* observational situation (note the emphasis upon absolute confirmability for observational sentences!), i.e., they seem to imply the rejection of the pragmatic theory.

As I said above, this tacit withdrawal from the pragmatic theory of observation is one of the most surprising features of modern empiricism. It is responsible for the fact that this philosophy, despite the apparent progress that has been made since the thirties, is still in accordance with the assumption that observational meanings are invariant with respect to the process of explanation and perhaps even with full meaning invariance (if we only consider that the behavioristic criterion of observability will be satisfied by any language that has been used for a long time, a long history and observational plausibility brought about by it are the best preconditions for the petrification of meanings; this applies to Platonism as well as to modern empiricism).

This finishes a somewhat lengthy digression which started immediately after the pronouncement of (4). I will make only two points before

[31] For a detailed criticism of this assertion, cf. my [27] and [39].
[32] Carnap [15], p. 40.
[33] *Ibid.*, p. 69.

returning to the main argument of the present paper: first, that the unwitting and partial return to the ideology of sense data is responsible for many of the 'inner contradictions' which are so characteristic of contemporary empiricism as well as for the pronounced similarity of this philosophy to the "school philosophies" it has attacked; second, that (4) has been accepted, not only by philosophers, but also by many physicists who believe in the so-called Copenhagen interpretation of microphysics. It is one of Niels Bohr's most fundamental ideas that "however far the new phenomena" found on the microlevel "transcend the scope of classical physical *explanation*, the account of all evidence must be expressed in classical terms."[34] I shall not discuss, in the present section, the arguments which Bohr has developed in favor of this idea. Let me only say that this idea immediately leads to the invariance of the meanings of the descriptive terms of the observation language, the classical signs now playing the role of the observational vocabulary.

To sum up: two ideas which are common to both the modern empiricist's theory of reduction and to his theory of explanation are:

(A) reduction or explanation is (or should be) by derivation;
(B) the meanings of (observational) terms are invariant with respect to both reduction and explanation.

In the sections to follow it will be my task to scrutinize these two basic principles. I shall begin with (A).

2. *Criticism of Reduction or Explanation by Derivation.*

The task of science, so it is assumed by those who hold the theory about to be criticized, is the explanation, and the prediction, of known singular facts and regularities with the help of more general theories. In what follows we shall assume T' to be the totality of facts and regularities to be explained, D' the domain in which T' makes correct predictions, and T (domain $D' \subset D$) the theory which functions as the basis of explanation.[35] Considering (3) we shall have to demand that T be either strong enough to contain T' as a logical consequence, or at least

[34] [6], pp. 209ff. For a more detailed account of Bohr's philosophy of science, see [32].
[35] In what follows it will not be necessary explicitly to distinguish between " 'T' " and "T," and this distinction will therefore not be made. Also terms such as 'consistent,' 'incompatible,' and 'follows from' will be applied to pairs of theories, (T,T'), and they will then mean that T *taken together with the conditions of validity of T', or the boundary conditions characterizing D'*, is compatible with, consistent with, or sufficient to derive, T'.

compatible with T′ (inside D′, that is). Only theories which satisfy one or the other of the two demands just stated are admissible as explanatia. Or, taking the demand for explanation for granted,

(5) only such theories are admissible (for explanation and prediction) in a given domain which either *contain* the theories already used in this domain, or are at least *consistent* with them.

It is in this form that (A) will be discussed in the present section and in the sections to follow.

As has just been shown, condition (5) is an immediate consequence of the logical empiricist's theory of explanation and reduction, and it is therefore adopted—at least by implication—by all those who defend that theory. However, its correctness has been taken for granted by a much wider circle of thinkers, and it has also been adopted independently of the problem of explanation. Thus, in his essay "Studies in the Logic of Confirmation" C. G. Hempel demands that "every logically consistent observation report" be "logically compatible with the class of all the hypotheses which it confirms," and more especially, he has emphasized that observation reports do "not confirm any hypotheses which contradict each other."[36] If we adopt this principle, then a theory T (see the notation introduced at the beginning of the present section) will be confirmed by the observations confirming a more narrow theory T′ only if it is compatible with T′. Combining this with the principle that a theory is admissible only if it is confirmed to some degree by the evidence available, we at once arrive at (5).

Outside philosophy, (5) has been taken for granted by many physicists. Thus, in his *Waermelehre*, Ernst Mach makes the following remark: "Considering that there is, in a purely mechanical system of absolutely elastic atoms no real analogue for the *increase of entropy*, one can hardly suppress the idea that a violation of the second law . . . should be possible if such a mechanical system were the *real* basis of thermodynamic processes."[37] And he insinuates that, for this reason, the mechanical hypothesis must not be taken too seriously.[38] More recently, Max Born has based his arguments against the possibility of a return to

[36] [45], p. 105, condition (8.3). It was J. W. N. Watkins who drew my attention to this property of Hempel's theory.

[37] [53], p. 364.

[38] For a much more explicit statement of what appears in [53] only as an insinuation, see [54].

determinism upon (5) and the assumption, which we shall here take for granted,[39] that the theory of wave mechanics is incompatible with determinism. "If any future theory should be deterministic," says he, "it cannot be a modification of the present one, but must be entirely different. How this should be possible without sacrificing a whole treasure of well-established results I leave the determinist to worry about."[40]

The use of (5) is not restricted to such general remarks, however. A decisive part of the quantum theory itself, viz., the so-called quantum theory of measurement, is the immediate result of the postulate that the behavior of macroscopic objects, such as measuring instruments, must obey some classical laws *precisely* and not only *approximately*. For example, macroscopic objects must always dwell in a well-defined classical state, and this despite the fact that their microscopic constituents exhibit a very different behavior. It is this postulate which leads to the introduction of abrupt jumps in addition to the continuous changes that occur in accordance with Schrödinger's equation.[41] An account of measurement which very clearly exhibits this feature has been given by Landau and Lifshitz.[42] These authors point out that "the classical nature of the apparatus means that . . . the reading of the apparatus . . . has some definite value." "This," they continue, "enables us to say that the state of the system apparatus + electron after the measurement will in actual fact be described, not by the entire sum [$\Sigma A_n(q)\Phi_n(\xi)$ where q is the coordinate of the electron, ξ the apparatus coordinate] but by only the one term which corresponds to the 'reading' g_n of the apparatus, $A_n(q)\Phi_n(\xi)$." Moreover, most of the arguments against suggestions such as those put forth by Bohm, de Broglie, and Vigier make more or less explicit use of (5).[43] A discussion of this condition is therefore very

[39] Born believes that this assumption has been established by von Neumann's proof. In this he is mistaken; see [29]. However, there exist different and quite plausible arguments for the incompatibility of determinism and wave mechanics, and it is for this reason that I take the assumption for granted. An outline of these plausible arguments is given in [37]. It should be noted that von Neumann himself did not share Born's inductivism. See [62], p. 327.

[40] [7], p. 109. In his treatment of the relation between Kepler's laws and Newton's theory, which, he thinks, applies to all pairs of theories which overlap in a certain domain and are adequate in this domain, Born explicitly accepts (5). For an analysis of Born's inductivism, see Popper [67].

[41] See [30].

[42] [52], p. 22. See also von Neumann's treatment of the Compton effect in [62], pp. 211–215.

[43] Cf. [32], [36], [38].

topical and leads right into the center of contemporary arguments about microphysics.

This discussion will be conducted in three steps. It will first be argued that most of the cases which have been used as shining examples of scientific explanation *do not* satisfy (5) and that it is not possible to adapt them to the deductive scheme. It will then be shown that (5) *cannot* be defended on empirical grounds and that it leads to very unreasonable consequences. Finally, it will turn out that once we have left the domain of empirical generalizations, (5) *should not* be satisfied either. In connection with this last, methodological step, the elements of a positive methodology for theories will be developed, and the historical, psychological, and semantical aspects of such a methodology will be discussed. Altogether the three steps will show that (A) is in disagreement both with actual scientific practice and with reasonable methodological demands. I start now with the discussion of the *actual* inadequacy of (5).

3. *The First Example.*

A favorite example of both reduction and explanation is the reduction of what Nagel calls the Galilean science to the physics of Newton,[44] or the explanation of the laws of the Galilean physics on the basis of the laws of the physics of Newton. By the Galilean science (or the Galilean physics) is meant, in this connection, the body of theory dealing with the motion of material objects (falling stone, penduli, balls on an inclined plane) near the surface of the earth. A basic assumption here is that the vertical accelerations involved are constant over any finite (vertical) interval. Using T′ to express the laws of this theory, and T to express the laws of Newton's celestial mechanics, we may formulate Nagel's assertion to the effect that the one is reducible to the other (or explainable on the basis of the other) by saying that

(6) $\quad T \,\&\, d \vdash T'$

where d expresses, in terms of T, the conditions valid inside D′. In the case under discussion d will include description of the earth and its surroundings (supposed to be free from air; we shall also abstract from

[44] [20], p. 291. I am aware that, from a historical point of view, the discussion to follow is not adequate. However, I am here interested in the systematic aspect, and I have therefore allowed myself what could only be regarded as great liberties if the main interest were historical.

all those phenomena which are due to the rotation of the earth and whose inclusion would strengthen, rather than weaken our case), and reference will be made to the fact that the variation H of the height above ground level in the processes described is very small if compared with the radius R of the earth.

As is well known (6) cannot be correct: as long as H/R has some finite value, *however small*, T′ will not follow (logically) from T and d. What will follow will rather be a law, T″, which, while being experimentally indistinguishable from T′ (on the basis of the experiments which formed the inductive evidence for T′ in the first place), is yet inconsistent with T′. If, on the other hand, we want to derive T′ precisely, then we must replace d by a statement which is patently false, as it would have to describe the conditions in the close neighborhood of the earth as leading to a vertical acceleration that is constant over a finite interval of vertical distance. It is therefore impossible, *for quantitative reasons*, to establish a deductive relationship between T and T′, or even to make T and T′ compatible. This shows that the present example is not in agreement with (5) and is, therefore, also incompatible with (A), (1), and (3).

Now in this situation, we may adopt one or the other of the following two procedures. We may either declare that the Galilean science can neither be reduced to, nor explained in, terms of Newton's physics;[45] or we may admit that reduction and explanation are possible, but deny that deducibility, or even consistency (on the basis of suitable boundary conditions), is a necessary condition of either. It is clear that the question as to which of these two procedures is to be adopted is of subordinate importance (after all, it is purely a matter of terminology that is to be settled here!) if compared with the question whether newly invented theories should be consistent with, or contain, those of their predecessors with whom they overlap in empirical content. We shall therefore defer settlement of the terminological problem raised above and concentrate on the question of consistency, or derivability. And we shall use the terms 'explanation' and 'reduction' either in a vague and general sense, awaiting further explication, or in the manner suggested by Nagel and by Hempel and Oppenheim. The usage adopted should always be clear from the context.

[45] This suggestion was made to me by Professor Viktor Kraft.

The objection which has just been developed—so it is frequently pointed out—cannot be said to endanger the correct theory of explanation, since everybody would admit that explanation may be by approximation only. This is a curious remark indeed! It criticizes us for taking seriously, and objecting to, a criterion which has either been universally stated as a necessary condition of explanation, or which plays a central role in some theories of confirmation, viz., condition (3). Now dropping (3) means altogether giving up the orthodox theory, for (3) formed the very core of this theory.[46] On the other hand, the remark that we explain "by approximation" is much too vague and general to be regarded as the statement of an alternative theory. As a matter of fact, it will turn out that the idea of approximation cannot any more be incorporated into a formal theory, since it contains elements which are essentially subjective. However, before dealing with this aspect of explanation we shall inquire a little more closely into the reasons for the failure of (3). Such an inquiry will lead to the result not only that (3) is false, but it is also very unreasonable to assume that it could be true.

4. *Reasons for the Failure of (5) and (3).*

The basic argument is really very simple, and it is very surprising that it has not been used earlier. It is based upon the fact that *one and the same set of observational data is compatible with very different and mutually inconsistent theories.* This is possible for two reasons: first, because theories, which are universal, always go beyond any set of observations that might be available at any particular time; second, because the truth of an observation statement can always be asserted within a certain margin of error only.[47] The first reason allows for theories to differ in domains where experimental results are not yet available. The second reason allows for such differences even in those domains where observations have been made, provided the differences are restricted to the margin of error connected with the observations.[48] Both reasons taken together sometimes allow considerable freedom in the construction of our theories.

[46] This has been emphasized, in private communication, by Professors Kraft (Vienna) and Rynin (Berkeley).

[47] As J. W. N. Watkins has pointed out to me, this invalidates Hempel's conditions 9.1 and 9.2 (in [45]). An attempt to bring logical order into the relation between observation statements and the more precise statements derived from a theory has been made by Professor S. Körner [50], p. 140.

[48] Even this condition is too strong, as will be shown below.

Now, it is very important to realize that this freedom which experience grants the theoretician is nearly always restricted by conditions of an altogether different character. These additional conditions are neither universally valid, nor objective. They are connected partly with the tradition in which the scientist works, with the beliefs and the prejudices which are characteristic of that tradition; and they are partly connected with his own personal idiosyncrasies. The formal apparatus available and the structure of the language he speaks will also strongly influence the activity of the scientist. Whorff's assertion to the effect that the properties of the Hopi language are not very favorable for the development of a physics like the one with which we are acquainted may very well be correct.[49] Of course, it must not be overlooked[50] that man is capable not only of applying, but also of inventing, languages. Still, the influence of the language from which he starts should never be underestimated. Another factor which strongly influences theorizing is metaphysical beliefs. The Neoplatonism of Copernicus was at least a contributing factor in his acceptance of the system of Aristarchus.[51] Also, the contemporary issue between the followers of Niels Bohr and the realists, being still undecidable on the basis of contemporary experimentation, is mainly metaphysical in character.[52] That the choice of theories may be influenced even by aesthetic motives can be seen from Galileo's reluctance to accept Kepler's ellipses.[53]

Taking all this into account we see that the theory which is suggested by a scientist will also depend, apart from the *facts* at his disposal, on the *tradition* in which he participates, on the mathematical instruments he accidentally knows, on his preferences, on his aesthetic prejudices, on the suggestions of his friends, and on other elements which are rooted, not in facts, but in the mind of the theoretician and which are therefore subjective. This being the case it is to be expected that theoreticians working in different traditions, in different countries, will arrive at theories which, although in agreement with all the known facts, are yet mutually inconsistent. Indeed, any consistency over a long period

[49] See [73].

[50] As is done by Bohr, Heisenberg, and von Weizsaecker in their philosophical writings as well as by some Wittgensteinians. For the point of view of these physicists, see [34] and [38], as well as the end of Sec. 7 of the present paper.

[51] See T. S. Kuhn [51], pp. 128ff.

[52] See [36].

[53] See E. Panofsky [63].

of time would have to be regarded not, as is suggested by (3), (A), and (5), as a methodological virtue, but as an alarming sign that no new ideas are being produced and that the activity of theorizing has come to an end. Only the inductivistic doctrine that theories are uniquely determined by the facts could have persuaded people that lack of ideas is praiseworthy and that its consequences are an essential feature of the development of our knowledge.[54]

At this point it is worth mentioning what will be explained in great detail later: that the freedom of theorizing granted by the indeterminateness of facts is of great methodological importance. It will turn out that many test procedures presuppose the existence of a class of mutually incompatible, but factually adequate, theories. Any attempt to reduce this class to a single theory would result in a decisive decrease of the empirical content of this remaining theory and would therefore be undesirable from the point of view of empiricism. The freedom granted by the indeterminateness of facts is therefore not only psychologically important (it allows scientists of different temperament to follow their different inclinations and thereby gives them satisfaction which goes beyond the satisfaction derived from the exclusive consideration of facts); it is also needed for methodological reasons.

The gist of the argument developed so far is that because of the latitude which experience allows the theoretician, and because of the different way in which this latitude will be exercised by thinkers of different tradition, temperament, and interests, it is to be expected that two different theories, and especially two theories of a different degree of generality, will be inconsistent with each other even in those cases where both are confirmed by the set. In this argument it was assumed that the experimental evidence which inside D′ confirms T and T′ is the same in both cases. Although this may be so in the specific example discussed, it is certainly not true in general. Experimental evidence does

[54] This is true mainly of those more crude theories of induction which are held, by implication, by many physicists. It would seem to me that discussion and criticism of these theories is a much more effective way of advancing scientific knowledge than invention of highly technical theories of confirmation which are of no interest to the scientist because they cannot be applied to a single noninstantial theory. Unfortunately, many philosophers of science consider it below their dignity to discuss such crude but effective theories, and they prefer the construction of sophisticated theories which are totally ineffective and useless. The brief discussion of Hempel's paper seems to show that similar objections must be raised against some ideas held by contemporary empiricists.

not consist of facts pure and simple, but of facts analyzed, modeled, and manufactured according to some theory.

The first indication of this manufactured character of the evidence is seen in the corrections which we apply to the readings of our measuring instruments, and in the selection which is made among those readings. Both the corrections and the selection made depend upon the theories held, and they may be different for the theoretical complex containing T, and for the theoretical complex containing T′. Usually T will be more general, more sophisticated, than T′, and it will also be invented a considerable time after T′. New experimental techniques may have been introduced in the meantime. Hence, the 'facts,' within D′, which count as evidence for T will be different from the 'facts,' within D′, which counted as evidence for T′ when the latter theory was first introduced. An example is the very different manner in which the apparent brightness of stars was determined in the seventeenth century and is determined now. This is another important reason why T usually will not satisfy (5) with respect to T′: not only are T and T′ connected with different theoretical ideas leading to different predictions even in the domain where they overlap and are both confirmed, but the better experimental techniques and the improved theories of measurement will usually provide evidence for T which is different from the evidence for T′ even within the domain of common validity. In short: introducing T very often leads to recasting the evidence for T′. The demand that T should satisfy (5) with respect to T′ would in this case imply the demand that new and refined measurements not be used, which is clearly inconsistent with empiricism.

Against the argument in the last paragraph it might be pointed out that results of measurement which are capable of improvement, and which therefore change, do not belong to the observational domain, but must be formulated with the help of singular statements of the theoretical language.[55] Observational statements proper are such qualitative statements as "pointer A coincides with mark B," or "A is greater than B"—and these statements will not change, or be eliminated, whatever the development of the theory, or of the methods of measurement. This point will be dealt with, and refuted, in Section 7 of this paper.

A further indication of the "manufactured" character of the experi-

[55] For this move, cf. Carnap [15], p. 40.

mental evidence is seen in the fact that observable results, and indeed anything conveyed with the help of a language, are always expressed in some theory or other. Because this fact will also be of importance in connection with my criticism of (B), and because it leads to a further criticism of (A), I shall discuss at length the example I have chosen for its elucidation.

5. *Second Example: The Problem of Motion.*[56]

From its very beginning, rational cosmology, this creation of the Ionian "physiologists," was faced with the problem of change and motion (in the general sense in which it includes locomotion, qualitative alteration, quantitative augmentation and diminution, as well as generation and corruption). The problem arose in two forms. The first was the possibility of change and motion. This form of the problem had to be solved by the invention of a cosmology which allowed for change, i.e., which was not such that the occurrence of change was (unwittingly) excluded from it by the very nature of the assumptions upon which it was based. The second form of the problem which arose, once the first had been solved in a satisfactory manner, was the *cause* of change. As was shown by Parmenides, the early monistic theories of Thales, Anaximander, and others could not solve the first form of the problem. For Parmenides himself, this did not refute monism; it refuted the existence of change.

The majority of thinkers went a different path, however. They regarded monism as refuted and started with pluralistic theories. In the case of the atomic theory, which was one of these pluralistic theories, this relation between Parmenides' arguments and pluralism is very clear. Leucippus, who "had associated with Parmenides in philosophy,"[57] "thought he had a theory which was in harmony with the senses, and did not do away with coming into being and passing away, nor motion, nor with the multiplicity of things."[58] This is how the atomic theory arose, as an attempt to solve problems created by the empirical inadequacy of the early monism of the Ionians.

However, the theory which was most influential in the Middle Ages

[56] For a more detailed account of the theories mentioned in this section, see M. Clagett [17]. Concerning the first part of the present section, see J. Burnet [10], as well as Clagett [16] and Popper [67].

[57] Aristotle [2], A, 8 324b35.

[58] Theophrastus quoted from Burnet [10], p. 333.

and which also tried to solve what I have above called the second form of the problem was Aristotle's theory of motion as the actualization of potentiality. According to Aristotle

(7) "motion is a process arising from the continuous action of a source of motion, or a 'motor,' and a 'thing moving.' "[59]

This principle, according to which any motion (and not only accelerated motion) is due to the action of some kind of force, can be easily supported by such common observations as a cart drawn by a horse and a chair pushed around by an angry husband. It gets into difficulties when one considers the motion of things thrown: stones continue to move despite the fact that contact with the motor apparently ceases when they leave the hand. Various theories have been suggested to eliminate this difficulty. From the point of view of later developments, the most important one of these theories is the impetus theory. The impetus theory retains (7) and the general background of the Aristotelian theory of motion. Its distinction lies in the specific assumptions it makes concerning the causes that are responsible for the motion of the projectile. According to the impetus theory, the motor (for example the hand) transfers upon the projectile an inner moving force which is responsible for its continuation of path, and which is continually decreased by the resisting air and by the gravity of the projectile. A stone in empty space would therefore either remain at rest or move (along a straight line[60]) with constant speed, depending on whether its impetus is zero or possesses a finite value.

At this point a few words must be said about the characterization of locomotion. The question as to its proper characterization was a matter of dispute. To us it seems quite natural to characterize motion by space transversed, and, as a matter of fact, one of the suggested characterizations did just this: it defined motion kinematically by reference to space transversed. This apparently very simple characterization needs further specification if an account is to be given of nonuniform movements where the distinction becomes relevant between average velocity and instantaneous velocity. Compared with the actual space transversed by a given body, the instantaneous velocity is a rather abstract notion since

[59] Clagett [17], p. 425.

[60] The parentheses I have added because of the absence from the earlier forms of the impetus theory of an explicit consideration of direction.

it refers to the space that would be transversed if the velocity were to retain constancy over a finite interval of time.

Another characterization of motion is the dynamical. It defines motion in terms of the forces which bring it about in accordance with (7). Adopting the impetus theory the motion of a stone thrown would have to be characterized by its inherent impetus, which pushes it along until it is exhausted by the opposing forces of friction and gravity.

Which characterization is the better one to take? From an operationalistic point of view (and we shall adopt this point of view, since we want to follow the empiricist as far as possible), the dynamical characterization is definitely to be preferred: while it is fairly easy to observe the impetus enclosed in a moving body by bringing it to a stop in an appropriate medium (such as soft wax) and then noting the effect of such a maneuver, it is much more difficult, if not nearly impossible, to arrange matters in such a way that from a given moment on, a non-uniformly moving object assumes a constant speed with a value identical with the value of the instantaneous velocity of the object at that moment and then to watch the effect of this procedure.

With the use of the dynamical characterization, the "inertial law" pronounced above reads as follows:

(8) The impetus of a body in empty space which is not under the influence of any outer force remains constant.

Now, in the case of inertial motions (8) gives correct predictions about the behavior of material objects. According to (3), explanation of this fact will involve derivation of (8) from a theory and suitable initial conditions. Disregarding the demand for explanation, we can also say, on the basis of (5), that any theory of motion that is more general than (8) will be adequate only if it contains (8) which, after all, is a very basic law. According to (2), the meanings of the key terms of (8) will be unaffected by such a derivation. Assuming Newton's mechanics to be the primary theory, we shall therefore have to demand that (8) be derivable from it *salva significatione*. Can this demand be satisfied?

At first sight it would seem that it is much easier to derive (8) from Newton's theory than it is to establish the correctness of (6): as opposed to Galileo's law (8) is not in quantitative disagreement with anything asserted by Newton's theory. Even better: (8) seems to be identi-

cal with Newton's first law so that the process of derivation seems to degenerate into a triviality.[61]

In the remainder of the present section, it will be shown that this is not so and that it is impossible to establish a deductive relationship between (8) and Newton's theory. Later on this will be the starting point of our criticism of (B).

Let me repeat, before beginning the argument, that (8), *taken by itself*, cannot be attacked on empirical grounds. Indeed, we have indicated a primitive method of measurement of impetus, and the attempt to confirm (8) by using this method will certainly show that within the domain of error connected with such crude measurements, (8) is perfectly all right. It is, therefore, quite in order to ask for the explanation, or the reduction, of (8), and the failure to arrive at a satisfactory solution of this task cannot be blamed upon the empirical inadequacy of (8).

We now turn to an analysis of the main terms of (8). According to Nagel the meaning of these terms is to be regarded as "fixed" by the procedures and assumption of the impetus theory, and any one of them is "therefore intelligible in terms of its own rules of usage."[62] What are these meanings, and what are the rules which establish them?

Take the term 'impetus.' According to the theory of which (8) is a part, the impetus is the force responsible for the movement of the object that has ceased to be in direct contact, by push, or by pull, with the material mover. If this force did not act, i.e., if the impetus were destroyed, then the object would cease to move and fall to the ground (or simply remain where it is, in case the movement were on a frictionless horizontal plane). A moving object which is situated in empty space and which is influenced neither by gravity nor by friction is not outside the reach of any force. It is pushed along by the impetus, which may be pictured as a kind of inner principle of motion (similar, perhaps, to the vital force of an organism which is the inner principle of *its* motion).

We now turn to Newton's celestial mechanics and the description, in terms of this theory, of the movement of an object in empty space. (Newton's theory still retains the notion of absolute space and allows therefore for such a description to be formed.) Quantitatively, the same

[61] There existed theories, among them the theory of *mail* by Abu'l-Barakat, where quantitative disagreement with Newton's laws was to be expected: in these theories, the impetus decreased with time in the same manner in which a hot poker that is removed from the fire gradually loses the heat stored in it. Cf. Clagett [17], p. 513.

[62] See Nagel [20], p. 301.

movement results. But can we discover in the description of this movement, or in the explanation given for it, anything resembling the impetus of (8)? It has been suggested that the momentum of the moving object is the perfect analogue of the impetus. It is correct that the measure of this magnitude (viz., mv) is identical with the measure that has been suggested for the impetus.[63] However, it would be very mistaken if we were, on that account, to identify impetus and momentum. For whereas the impetus is supposed to be something that pushes the body along,[64] the momentum is the result rather than the cause of its motion. Moreover, the inertial motion of classical mechanics is a motion which is supposed to occur by itself, and without the influence of any causes. After all, it is this feature which according to most historians, radical empiricists included, constitutes one of the main differences between the Aristotelian theory and the celestial mechanics of the seventeenth, eighteenth, and nineteenth centuries: in the Aristotelian theory, the natural state in which an object remains without the assistance of any causes is the state of rest. A body at rest (in its natural place, we should add) is not under the influence of any forces. In the Newtonian physics it is the state of being at rest or in uniform motion which is regarded as the natural state. This means, of course, the explicit denial of a force such as the impetus is supposed to represent.

Now this denial need not mean that the concept of such a force cannot be formed within Newton's mechanics. After all, we deny the existence of unicorns and use in this denial the very concept of a unicorn. Is it then perhaps possible to define a concept such as impetus in terms of the theoretical primitives of Newton's theory? The surprising fact is that any attempt to arrive at such a definition leads to disappointment (which shows, by the way, that theories such as Newton's are expressed in a language that is much more tightly knit than is the language of everyday life). I have already pointed out that the momentum, which would give us the correct mathematical value, is not what we want. What we want is a *force* that acts upon the isolated object and is responsible for its motion. The concept of such a force can of course be formed within Newton's theory. But considering (a) that the movement under review (the inertial movement) occurs with constant ve-

[63] See Clagett [17], p. 523.

[64] For an elaborate discussion of the difference between momentum and impetus, see Anneliese Maier [58]. For what follows, see also M. Bunge [9], Ch. 4.4.

locity, and (b) Newton's second law, we obtain in all relevant cases zero for the *value* of this force which is not the measure we want. A positive measure is obtained only if it is assumed that the movement occurs in a resisting medium (which is, of course, the original Aristotelian assumption), an assumption which is inconsistent with another feature of the case considered, i.e., with the fact that the inertial movement is supposed by Newton's theory to occur in empty space. I conclude from this *that the concept of impetus, as fixed by the usage established in the impetus theory, cannot be defined in a reasonable way within Newton's theory*. And this is not further surprising. For this usage involves laws, such as (7), which are inconsistent with the Newtonian physics.

In the last argument, the assumption that the concept *force* is the same in both theories played an essential role. This assumption was used in the transition from the assertion, made by the impetus theory, that inertial motions occur under the influence of forces to the calculation of the magnitude of these forces on the basis of Newton's second law. Its legitimacy may be derived from the fact that both the impetus theory and Newton's theory apply the concept *force* under similar circumstances (paradigm-case argument!). Still, meaning and application are not the same thing, and it might well be objected that the transition performed is not legitimate, since the different contexts of the impetus theory, on the one hand, and of Newton's theory, on the other, confer different meanings upon one and the same word 'force.' This being the case, our last argument is based upon a *quaternio terminorum* and is, therefore, invalid. In order to meet this objection, we may repeat our argument using the word 'cause' instead of the word 'force' (the latter has a somewhat more specific meaning). But if someone again retorts that 'cause' has a different meaning in Newton's theory from what it has in the impetus theory, then all I can say is that a consistent continuation of that kind of objection will in the end establish what I wanted to show in a more simple manner, viz., the impossibility of defining the notion of an impetus in terms of the descriptive terms of Newton's theory. To sum up: the concept *impetus* is not "explicable in terms of the theoretical primitives of the primary science."[65] And this is exactly as it should be, considering the inconsistency between some very basic principles of these two theories.

[65] Nagel [20], p. 302.

However, explication in terms of the primitives of the primary science is not the only method which was considered by Nagel in his discussion of the process of reduction. Another way to achieve reduction, which he mentions immediately after the above quotation, "is to adopt a material, or physical hypothesis according to which the occurrence of the properties designated by some expression in the premises of the primary science is a sufficient, or a necessary and sufficient condition for the occurrence of the properties designated by the expressions of the secondary discipline." Both procedures are in accordance with (4), or with (2), or at least Nagel thinks that they are: ". . . in this case" he says, referring to the procedure just outlined, "the meaning of the expressions of the secondary science *as fixed by the established usage of the latter*, is not declared to be analytically related to the meanings of the corresponding expressions of the primary science."[66] Let us now see what this second method achieves in the present case.

To start with, this method amounts to introducing a hypothesis of the form

(9) impetus = momentum

where each side retains the meaning it possesses in its respective discipline. The hypothesis then simply asserts that wherever momentum is present, impetus will also be present (see the above quotation of Nagel's), and it also asserts that the measure will be the same in both cases. Now this hypothesis, although acceptable within the impetus theory (after all, this theory permits the incorporation of the concept of momentum), is incompatible with Newton's theory. It is therefore not possible to achieve reduction and explanation by the second method.

To sum up: a law such as (8) which, as I have argued, is empirically adequate, and in quantitative agreement with Newton's first law, is yet incapable of reduction to Newton's theory and therefore incapable of explanation in terms of the latter. Whereas the reasons we have so far found for irreducibility were of a quantitative nature, this time we met a qualitative reason, as it were, i.e., the incommensurable character of the conceptual apparatus of (8), on the one side, with that of Newton's theory, on the other.

Taking together the quantitative as well as the qualitative argument, we are now presented with the following situation: there exist pairs of

[66] *Ibid.* My italics.

theories, T and T′, which overlap in a domain D′ and which are incompatible (though experimentally indistinguishable) in this domain. Outside D′, T has been confirmed, and it is also more coherent, more general, and less *ad hoc* than T′. The conceptual apparatus of T and T′ is such that it is possible neither to define the primitive descriptive terms of T′ on the basis of the primitive descriptive terms of T nor to establish correct empirical relations involving both these terms (correct, that is, from the point of view of T). This being the case, explanation of T′ on the basis of T or reduction of T′ to T is clearly impossible if both explanation and reduction are to satisfy (A) and (B). Altogether, *the use of T will necessitate the elimination both of the conceptual apparatus of T′ and of the laws of T′*. The conceptual apparatus will have to be eliminated because its use involves principles, such as (7) in the example above, which are inconsistent with the principles of T; and the laws will have to be eliminated because they are inconsistent with what follows from T for events inside D′. (This would apply to the example above if the theory of *mail* had been used instead of the impetus theory.) This being the case the demand for explanation and reduction clearly cannot arise if this demand is interpreted as the demand for the explanation, or reduction, of T′, rather than of a set of laws that is in some respect similar to T′ but in other respects (meanings of fundamental terms included) very different from it. For such a demand would imply the demand to derive, from correct premises, what is false, and to incorporate what is incommensurable.

The effect of the transition from T′ and T is rather to be described in the manner indicated in the introductory remarks of the present paper: where I said: What happens when transition is made from a restricted theory T′ to a wider theory T (which is capable of covering all the phenomena which have been covered by T′) is something much more radical than incorporation of the *unchanged* theory T′ into the wider context of T. What happens is rather a *complete replacement* of the ontology of T′ by the ontology of T, and a corresponding change in the meanings of all descriptive terms of T′ (provided these terms are still employed). Let me add here that the not-too-well-known example of the impetus theory versus Newton's mechanical theory is not the only instance where this assertion holds. As I shall show a little later, more recent theories also correspond to it. Indeed, it will turn out that the principle correctly describes the relation between the elements of *any* pair of noninstantial theories satisfying the conditions which I have just enumerated.

This finishes step one of the argument against the assumption that reduction and explanation are by derivation. What I have shown (and shall show in later sections) is that some very important cases which have been, or could be used as examples of reduction (and explanation) are not in agreement with the condition of derivability. It will be left to the reader to verify that this holds in almost all cases of explanation by theories: assumption (A) does not give a correct account of actual scientific practice. It has also been shown that in this respect the thesis formulated in the beginning of this paper is much more adequate.

Now, as against this result it may be pointed out, with complete justification, that scientific method, as well as the rules for reduction and explanation connected with it, is not supposed to describe what scientists are actually doing. Rather, it is supposed to provide us with normative rules which should be followed, and to which actual scientific practice will correspond only more or less closely.[67] It is very important nowadays to defend such a normative interpretation of scientific method and to uphold reasonable demands even if actual scientific practice should proceed along completely different lines. It is important because many contemporary philosophers of science seem to see their task in a very different light. For them actual scientific practice is the material from which they start, and a methodology is considered reasonable only to the extent to which it mirrors such practice.[68] Looking at disciplines such as medicine, they discover (whether rightly or wrongly—this I do not want to discuss at the present moment) that what is called "explanation" here is not always the inverse of prediction, and they infer from this that the orthodox model which demands such an inverse relationship to hold between explanation and prediction[69] is unduly restrictive.

Two elements must be distinguished in this "discovery."[70] The first is of a purely linguistic character. It is connected with the problem as to what meaning should be given to the word 'explanation.' Clearly this

[67] This point has been made, most forcefully, by K. R. Popper [68], Sec. 10.

[68] If I understand him correctly, this is also the point of view held by my colleague, Professor T. S. Kuhn.

[69] Cf. Hempel and Oppenheim [47], p. 323.

[70] The "discovery" is due among others to Professor Barker. See his contribution to [24], my criticism in the same volume, and his reply.

element is without serious interest. It may be that the word 'explanation' sounds beautiful to some ears—but who would think it reasonable to start a war for, or against, its elimination? The second element, however, which usually remains hidden beneath the linguistic analysis, is much more serious. For what the suggested procedure amounts to is increased leniency with respect to questions of test: a certain medical hypothesis (which, let us say, is expressed by saying that a patient died *because* of tuberculosis) is accepted, and retained, despite the fact that independent tests (independent, that is, of the *past* histories of this case and of other cases) are not available, and its further use is defended by reference to the fact that it is in accordance with the "logic of medicine." Expressed in more pedestrian terms, this maneuver propagates the acceptance of unsatisfactory hypotheses on the grounds that this is what everybody is doing. It is conformism covered up with high-sounding language.[71] It is clear, however, that if this conformism had been propagated successfully in the Middle Ages, modern science with its so very different "logic" would never have come into existence. Modern science is the result of a conscious criticism of the theses propagated and the methods employed by the great majority of scholastic philosophers. For the thinker who demands that a subject be judged "according to its own standards," such criticism is of course impossible; he will be strongly inclined to reject any interference and to "leave everything as it

[71] It should be pointed out that almost all theses which terminate (if at all) the long-winded inquiries of linguistic philosophers possess this two-faced character. On the one side, they seem to be about the meanings of terms only, and therefore rather harmless and uninteresting (although there are enough enthusiastic buyers even for such products). However, on closer inspection, it often turns out that beneath these linguistic trappings there are hidden, and thereby removed from criticism, some highly questionable theories or methodological rules. Only consider the example discussed in the text: it may well be the case that in some disciplines the word 'explanation' is used in a manner that does not lead to the demand for additional predictions. This linguistic result may be expressed by saying that prediction is not essential to explanation (viz., to 'explanation' in the new sense that is characteristic for the disciplines under review). Now from this last result it is then very often inferred that the search for additional predictions is unnecessary and that all is well. Clearly this methodological consequence can be derived from the linguistic premise only if it is further assumed that all is well once an explanation has been given, and this regardless of the sense in which the word 'explanation' is being used in the discipline in question. This assumption, which, I submit, is the silent premise of many contemporary linguistic arguments about explanation, amounts to asserting that all is well as long as the *word* 'explanation' occurs somewhere in the description of the procedure to be analyzed. This is, of course, pure word magic. Curiously enough, it is the linguistic philosopher who is swayed by such word magic. Which only shows how little language is understood by some of its most verbose champions.

is."[72] It is somewhat puzzling to find that such demands are nowadays advertised under the title of philosophy of science.

Against such conformism it is of paramount importance to insist upon the normative character of scientific method. Adopting this point of view, one cannot regard the arguments of the last few sections as ultimately decisive. They are satisfactory insofar as they show that the "orthodox" are wrong when asserting that (A), (B), and (5) reflect actual scientific practice. But they do not dispose of these principles if they are interpreted as demands to be followed by the scientist (although, of course, they provide ample material for such disproof). I therefore proceed now to a methodological criticism of the demands of the orthodox. The first move in this criticism will be the examination of an argument which has sometimes been used to defend (5).

6. *Methodological Considerations.*

The argument runs as follows: (α) a good theory is a summary of facts; (β) the predictive success of T′ (I will continue to use the notation introduced in Section 2) has shown T′ to be good theory inside D′; hence (τ), if T, too, is to be successful inside D′, then it must either give us all the facts contained in T′, i.e., it must give us T′, or at least it must be compatible with T′.

It is easily seen that this very popular argument[73] will not do. We can show this by considering its premises. Premise (α) is acceptable if it is not taken in too strict a sense (for example, if it is not interpreted as implying an ontology of mutually independent 'facts' as has been suggested by Mach and the early Wittgenstein). Interpreted in such a loose manner (α) simply says that a good theory not only will be able to answer many questions, but will also answer them correctly. Now if this is to be the interpretation of (α), then (β) cannot possibly be correct: in (β) the predictive success of T′ is taken to indicate that T′ will give a correct account of *all* the facts inside its domain. However, one must remember that because of the general character of statements expressing laws and theories, their predictive success can be established

[72] Wittgenstein [74], Sec. 124.

[73] A sloppy version of this argument occurs frequently in arguments by physicists. It ought to be mentioned, by the way, that Hempel's condition 8 leads to the very same result, viz., to the demand that new theories be consistent with their confirmed predecessors. A justification for discussing "crude" arguments of the kind outlined in the text above is given in fn. 54.

only with respect to part of their content. Only part of a theory can at any time be known to be in agreement with observation. From this limited knowledge nothing can be inferred (logically!) with respect to the remainder.[74]

We have also to consider the margin of error involved in every single test. Hence, from a purely logical point of view, new theories will be restricted only to the extent to which their predecessors have been tested and confirmed.[75] Only to this extent will it be necessary for them to agree with their predecessors. In domains where tests have not yet been carried out, or where only very crude tests have been made, we have complete freedom on how to proceed, and this quite independently of which theories were originally used here for the purpose of prediction. Clearly this last condition, which is in agreement with empiricism, is much less restrictive than either (3) or (5).

One might hope to arrive at more restrictive conditions by adding inductive argument to logical reasoning. True, from a logical point of view we can only say that *part* of T′ has been found to be in agreement with observation and that T need agree only with that part and not, as is demanded in (5), with the whole of T′. However, if inductive reasoning is used as well, then we shall perhaps have to admit that this partial confirmation has established T′, and that therefore the whole of T′ should be covered by T. Does this help us to strengthen the condition mentioned at the end of the last paragraph and to demonstrate (5) after all?

It is clear that inductive reasoning cannot establish (5) either. For let us assume that T agrees with T′ only where T′ has been confirmed and is different from T′ in all other instances without having as yet been refuted. In this case T will satisfy our own condition of the last paragraph, and it will not satisfy any stronger condition (except accidentally). Can inductive reasoning prompt us to eliminate T? It is not easily seen how this could be the case, since T shares all its confirming instances with T′. Hence, if T′ is established by these instances, then so is T—unless we use formal considerations (which I shall discuss later). Again

[74] This point is indebted to Hume. That Hume's arguments are still not understood by many thinkers and are therefore still in need of repetition has been emphasized by Popper [68], Reichenbach [69], Goodman [42], and others.

[75] As was mentioned in Sec. 4, it is hardly ever the case that two theories which have been discussed in very different historical periods will be based upon exactly the same observations. The condition is therefore still too strict.

we arrive at the result that from the point of view of fact there is not much to choose between T and T′, and that (5) cannot be defended on empirical grounds.

It is worthwhile to inquire a little more closely into the effects which adoption of (5), and, incidentally, also of Hempel's condition 8,[76] would have upon the development of scientific knowledge. Such adoption would lead to the elimination of a theory, not because it is inconsistent with the facts, but because it is inconsistent with another, and as yet unrefuted, theory whose confirming instances it shares. This is a strange procedure to be adopted by thinkers who, above anything, claim to be empiricists! However, the situation becomes even worse when we inquire why the one theory is retained and the other rejected. The answer (which is, of course, not the answer given by the empiricist) can only be that the theory which is retained was there first. This shows that in practice the allegedly empirical procedure (5) leads to the preservation of the old theories and to the rejection of the new theories even before these new theories have been confronted with the facts. That is, it leads to the same result as transcendental deduction, intuitive argumentation, and other forms of a priori reasoning, the only difference being that now it is in the name of experience that such results are obtained. This is not the only instance where, on closer scrutiny, a rather close relation emerges between some versions of modern empiricism and the 'school philosophies' it attacks.

We have now to consider the argument that formal criteria may provide a principle of choice between T and T′ that is independent of fact. Such formal criteria can indeed be given.[77] However, while usually a more general and coherent theory is preferred to a less general collection of laws, and this on account of being less *ad hoc*, (5) tends to reverse this procedure. This is due to the fact that general theories of a high degree of coherence usually violate (5). Again this principle is seen to be incompatible with reasonable methodology.

Two things have been shown so far. First, the invalidity of an argument used for establishing (5). Second, the undesirability, from an empirical point of view, of some consequences of this argument. However, all this has little weight when compared with the following most important consideration.

[76] See [45], p. 105.
[77] See Popper [68], Ch. VI.

Within contemporary empiricism, discussions of test and of empirical content are usually carried out in the following manner: it is inquired how a theory is related to its empirical consequences and what these consequences are. True, in the derivation of these consequences reference will have to be made to principles or theorems which are borrowed from other disciplines and which then occur in the correspondence rules. However, these principles and these theorems play a subordinate role when compared with the theory under review; and it is, of course, also assumed that they are mutually consistent and consistent with the theory. One may therefore say that, for the orthodox procedure, the natural unit to which discussions of empirical content and of test methods are referred is always a single theory taken together with those of its consequences that belong to the observation language.

This manner of discussion does not allow us to give an adequate account of crucial experiments which involve more than one theory, none of which are expendable or of psychological importance only. A very good example of the structure of such crucial tests is provided by the more recent development of thermodynamics. As is well known, the Brownian particle is a perpetual motion machine of the second kind, and its existence refutes the (phenomenological) second law. However, could this fact have been discovered in a direct manner, i.e., by a direct investigation of the observational consequences of thermodynamics? Consider what such a refutation would have required! The proof that the Brownian particle is a perpetual motion machine of the second kind would have required (a) measurement of the exact motion of the particle in order to ascertain the changes of its kinetic energy plus the energy spent on the overcoming of the resistance of the fluid, and (b) precise measurements of temperature and heat transfer in the surrounding medium in order to ascertain that any loss occurring here was indeed compensated by the increase of the energy of the moving particle and the work done against the fluid as mentioned in (a). Such measurements, however, are beyond experimental possibilities.[78] Hence, a direct refutation of the second law, i.e., a refutation based upon an investigation of the testable consequences of thermodynamics alone, would have had to wait for one of those rare, not repeatable, and therefore, prima facie suspicious, large fluctuations in which the transferred heat is indeed

[78] Concerning the extreme difficulties of following the motion of the Brownian particle in all its details, see R. Fuerth [41].

accessible to measurement. This means that such a refutation would have never taken place, and, as is well known, the actual refutation of the second law was brought about in a very different manner. It was brought about via the kinetic theory and Einstein's utilization of it in the calculation of the statistical properties of the Brownian motion. In the course of this procedure the phenomenological theory (T′) was incorporated into the wider context of statistical physics (T) in such a manner that (5) was violated; and then a crucial experiment was staged (Perrin's investigations).

Now it seems to me that the more general our knowledge gets the more important it will be to carry out tests in the manner indicated, i.e., not by comparing a single theory with experience, but by staging crucial experiments between theories which, although in accordance with all the known facts, are mutually inconsistent and give widely different answers in unexplored domains. This suggests that outside the domain of empirical generalizations the methodological unit to which we refer when discussing questions of test and empirical content consists of a whole set of partly overlapping, factually adequate, but mutually inconsistent theories. To the extent to which utilization of such a set provides additional tests which, for empirical reasons, could not have been carried out in a direct manner, the use of a set of this kind is demanded by empiricism. For the basic principle of empiricism is, after all, to increase the empirical content of whatever knowledge we claim to possess.[79]

On the other hand, the fact that (5) does not allow for the formation of such sets now proves this principle to be inconsistent with empiricism. By excluding valuable tests it decreases the empirical content of the theories that are permited to remain (and which, as indicated above, will usually be the theories which were there first). This last result of a consistent application of (5) is of very topical interest: it may well be, as has been pointed out by Bohm and Vigier,[80] that the refutation of the quantum-mechanical uncertainties presupposes just an incorporation of the present theory into a wider context, which is not any more in

[79] With respect to this last demand, it might be objected that, given a certain theory, such an extension of content is not possible without changing the theory. This argument would be correct if it could be granted that the interpretation of a physical theory is wholly empirical. As I have shown elsewhere (see [39]), this is not the case. The demand to increase the empirical part of this interpretation is therefore a sensible demand.

[80] See the discussion remarks of these two physicists in [49], as well as those of Bohm [4].

accordance with the idea of complementarity and which therefore suggests new and decisive experiments. And it may also be that the insistence on part of the majority of contemporary physicists upon (5) will, if successful, forever protect these uncertainties from refutation. This is how modern empiricism may finally lead to a situation where a certain point of view petrifies into dogma by being, in the name of experience, completely removed from any conceivable criticism.

To sum up the arguments of the present section: it has been shown that neither (5) nor (A) can be defended on the basis of experience. Quite on the contrary, a strict empiricism will admit theories which are factually adequate and yet mutually inconsistent.[81] An analysis of the character of tests in the domain of theories has revealed, moreover, that the existence of sets of partly overlapping, mutually inconsistent, and yet empirically adequate theories is not only possible, but also required. I shall now conclude the present section by discussing a little more in detail the logical and psychological consequences of the use of such a set.

Increase of testability will not be the only result. The use of a set of theories with the properties indicated above will also improve our understanding of each of its members by making it very clear what is denied by the theory that happens to be accepted in the end. Thus, it seems to me that our understanding of Newton's somewhat obscure notion of absolute space and of its merits is greatly improved when we compare it with the relational ideas of Berkeley, Huyghens, Leibnitz, and Mach, and when we consider the failure of the latter ideas to give a satisfactory account of the phenomenon of inertial forces. Also, the study of general relativity will lead to a deeper understanding of this notion than could be obtained from a study of the *Principia* alone.[82] This is not meant to

[81] It is interesting to study in detail the dilemma of an inductivistic philosophy of science. To start with, a radical empiricist demands close adherence to the facts and is ultimately suspicious of any generalization; a summary of facts is all he will admit on empirical grounds. However, generalizations *do* play an important role in the sciences; scientific knowledge is, after all, a collection of theories. Hence, methods must be found to justify these theories on the basis of experience, and a "logic" must be constructed which allows us, again on the basis of experience, to confer some kind of certainty upon them. A "logic" of this kind is introduced, and theories are established with its help. But now the demand that the future theories should be consistent with the theories thus established turns out to be far too strict. The way out of this dilemma can only consist in abandoning the idea that theories can be "established" by experience and in admitting that insofar as they go beyond the facts we have no means whatever (except, perhaps, psychological ones) to guarantee their trustworthiness.

[82] This, by the way, is one of the reasons why an axiomatic exposition of physical

be understood in a psychological sense only. For just as the meaning of a term is not an intrinsic property of it but is dependent upon the way in which the term has been incorporated into a theory, in the very same manner the content of a whole theory (and thereby again the meaning of the descriptive terms which it contains) depends upon the way in which it is incorporated into both the set of its empirical consequences and the set of all the alternatives which are being discussed at a given time:[83] once the contextual theory of meaning has been adopted, there is no reason to confine its application to a single theory, or a single language, especially as the boundaries of such a theory or of such a language are almost never well defined. The considerations above have shown, moreover, that the unit involved in the test of a specific theory is not this theory taken together with its own consequences; they have shown

principles, such as Newton's, is inferior by far to a dialectical exposition where many ideas are considered, and the pros and cons discussed, until finally one theory is pronounced the most satisfactory one. Of course, if one holds that, concerning theories, the only relation of interest is the relation between a single theory and "the facts," and if one also believes that these facts single out a certain theory more or less uniquely, then one will be inclined to regard discussion of alternatives as a matter of history, or of psychology, and one will even wish to hide, with some embarrassment, the situation at the time when the clear message of the facts had not yet been grasped. However, as soon as it is recognized that the refutation (and thereby also the confirmation) of a theory necessitates its incorporation into a family of mutually inconsistent alternatives, in the very same moment, the discussion of these alternatives becomes of paramount importance for methodology and should be included in the presentation of the theory that is accepted in the end. For the same reason, adherence to either the distinction between context of discovery (where alternatives are considered, but given a psychological function only) and context of justification (where they are not mentioned any more), or strict adherence to the axiomatic approach must be regarded as an arbitrary and very misleading restriction of methodological discussion: much of what has been called "psychological," or "historical," in past discussions of method is a very relevant part of the theory of test procedures.

Considering all this, the increased attention paid to the historical aspects of a subject, and the attempts to break down the distinction between the synthetic and the analytic must be welcomed as steps in the right direction. However, even here there are drawbacks. Only very few of the enthusiastic proponents of an increased study of the history of a subject realize the methodological importance of their investigations. The justification they give for their interest is either sentimental or psychological ("it gives me ideas"), or based upon some very implausible (because Hegelian) notions concerning the "growth" of knowledge. What these thinkers need in order not to fall victims to all sorts of quasi-philosophies is a methodological backbone, and I hope that the theory of test which has been sketched above in its merest outlines will provide such a backbone.

[83] In the twentieth century, the contextual theory of meaning has been defended most forcefully by Wittgenstein; see [74] as well as my summary in [28]. However, it seems that Wittgenstein is inclined to restrict this theory to the inside of his language games: Platonism of concepts is replaced by Platonism of (theories or) games. For a brief criticism of this attitude see [35].

that this unit is a whole class of mutually incompatible and factually adequate theories. Hence, both consistency and methodological considerations suggest such a class as the context from which meanings are to be made clear.[84]

Also, the use of such a class rather than of a single theory is a most potent antidote against dogmatism. Psychologically speaking, dogmatism arises, among other things, from the inability to imagine alternatives to the point of view in which one believes. This inability may be due to the fact that such alternatives have been absent for a considerable time and that, therefore, certain ways of thinking have been left undeveloped; it may also be due to the conscious elimination of such alternatives. However that may be, persistence of a single point of view will lead to the gradual establishment of well-circumscribed methods of observation and measurement; it will lead to codification of the ways in which these results are interpreted; it will lead to a standardized terminology and to other developments of a similarly conservative kind. This being the case, the gradual acceptance of the theory by an ever-increasing number of people must finally bring about a transformation even of the most common idiom that is taught in very early youth. In the end, all the key terms will be fixed in an unambiguous manner, and the idea (which may have led to such a procedure in the first place) that they are copies of unchanging entities and that change of meaning, if it should happen, is due to human mistake—this idea will now be very plausible. Such plausibility reinforces all the maneuvers which may be used for the preservation of the theory (elimination of opponents included[85]).

The conceptual apparatus of the theory having penetrated nearly all means of communication, such methods as transcendental deduction and analysis of usage, which are further means of solidifying the theory, will be very successful. Altogether it will seem that at last an absolute and irrevocable truth has been arrived at. Disagreement with facts may of course occur, but, being now convinced of the truth of the existing

[84] Textbooks and historical presentations very often create the impression either that such classes never existed and that physicists (at least the "great" ones) at once arrived at the one and good theory or that their existence must not be taken too seriously. This is quite understandable. After all, historians have been just as much under the influence of inductivistic ideas as the physicists and the philosophers.

[85] Today, of course, the "elimination" takes the more "refined" form of a refusal to publish (or to read) what is not in agreement with the accepted doctrine. However, this "liberalism" applies to physical theories only. It does not seem to apply to political theories.

point of view, its proponents will try to save them with the help of *ad hoc* hypotheses. Experimental results that cannot be accommodated, even with the greatest ingenuity, will be put aside for later consideration. The result will be absolute truth, but, at the same time, it will decrease in empirical content to such an extent that all that remains will be no more than a verbal machinery which enables us to accompany any kind of event with noises (or written symbols) which are considered true statements by the theory.[86]

The picture painted above is by no means exaggerated. The way in which, for example, the theory of witchcraft and daemonic influence crept into the most common way of thinking, and could be preserved for quite a considerable time, offers a vivid illustration of each point mentioned in the last paragraph. Moreover, the story of its overthrow furnishes another illustration of our thesis that comprehensive theories cannot be eliminated by a direct confrontation with "the facts."

Now let us compare such a dogmatic procedure with the effects of the use of a class of theories rather than a single theory. First of all such a procedure will encourage the building of a great variety of measuring instruments. There will be no *one* way of interpreting the results, and the theoretician will be trained to switch quickly from one interpretation to another.[87] Intuitive appeal will lose its paralyzing effect, transcendental deduction which, after all, presupposes uniformity of usage, will be impossible; and the question of agreement with the facts will assume a very prominent position. Experimental results which are inconsistent with one theory may be consistent with a different theory; this eliminates the motives for *ad hoc* hypotheses, or at least reduces them considerably. Nor will it be necessary to use instrumentalism as a means of getting out of trouble, since a coherent account may be provided by an alternative to the theory considered. The likelihood that empirical results will be left lying around will also be smaller; if they do not fit one theory, they will fit another. It is not at all superfluous to mention the tremendous development of human capabilities encouraged by such a procedure and the antidotes it contains against the wish to set up, and to obey, all-powerful regimes, be they political, religious, or scientific. Taking all this into account, we are inclined to say that *whereas unanimity of opinion may be fitting for a church, or for the willing followers of a tyrant, or some other kind of 'great man,' variety of opinion is a methodological necessity for the sciences and, a fortiori, for philosophy*. Neither (A), nor (B), nor (5) allows for such variety. It follows that, to the extent to which both principles (and the philosophy behind them) delimit variety and demand future theories to be consistent with theories already in existence, they contain a theological element (which lies, of course, in the worship of "facts" so characteristic for nearly all empiricism).

The paralyzing effect of familiarity and intuitive appeal upon social reconstruction and the progress of knowledge has been understood and described in a most excellent manner by Bertolt Brecht. It is true that Brecht was mainly concerned with the function of the theater in the process of making familiar, and thereby creating the impression of the unchangeability of, social relations. But in his analysis, which was to establish the need for a new kind of theater, he hit upon a very important fact of a much more general character, namely, the paralyzing effect of familiarity and of the methods used to bring about such familiarity. In the theater these methods consist in trying to represent what is accidental and changeable (a particular social situation, for example) as essential to man or nature, and therefore unchangeable. "The theatre with which we are confronted shows the structure of society (represented on the stage) as being removed from the influence of society (which is situated in the audience). Oedipus, who has sinned against some principles which were supported by the society of his time, is executed; the gods take care of that—they cannot be criticised" ([8], p. 146). "What we need," Brecht continues, "is a theatre which does not only allow for those sensations, insights, and impulses that are permitted by

[86] The fact that all the features of absolute knowledge can be manufactured by exercising an absolute conformism I regard as the most important objection to any claim to finality: it would seem to show that whereas hypothetical knowledge, being the result of repeated corrections in the light of criticism, is at least in partial contact with the world, absolute knowledge is entirely man made and can therefore not raise any claim to factual content. Even the stability of a testable hypothesis cannot be regarded as a sign of its truth because it may be due to the fact that, owing to some particular astigmatism on our part, we have as yet overlooked some very decisive tests. There is no sign by which factual truth may be recognized.

It is interesting to note that the development toward dogmatism as it has been described in the text occurred both in the "school philosophies" and in empiricism, and, in the latter case, notably in physics: there is no indication that empiricism provides special protection against dogmatic petrification. The only difference is that its defense will be in terms of "experience" rather than in terms of "intuition" or "revelation." For this see Dewey [21], esp. Ch. II.

[87] As Professor Agassi has pointed out to me, this method was consciously used by Faraday in order to escape the influence of prejudice. Concerning its role in modern discussions about the microlevel, see [37].

the field of human relations where the action takes place; what we need is a theatre which uses and creates thoughts and feelings that play a role in the change of the field itself" ([8], p. 147). *It is exactly the same thing that is needed in the domain of epistemology*. What is needed is a method which does not—in the name of either "universal principles," "revelation," or "experience"—put fetters on the scientist's imagination but which enables him to use alternatives to the point of view which is the one commonly accepted. What is needed is a method that also enables him to take a critical attitude with respect to any element of this point of view, be it a law, or a so-called empirical fact. I am afraid that only very few scientists have ever been aware of the need for such a method, and that most of them are to be compared to the transfixed audience of one of the familiar pieces of the "classical" repertoire which has been described so vividly by Brecht—textbooks and scientific journals replacing the images projected by the actors.

At all times, the existence, within a certain tradition, of a variety of opinions (or a variety of theories) has been regarded as proof of the unsoundness of the method adopted by the members of this tradition. It was assumed, as being nearly self-evident, that the proper method must lead to the truth, that the truth is one, and that the proper method must therefore result in the establishment of a single theory and the perennial elimination of all alternatives. Conversely, the existence of various points of view and of a community where discussion of alternatives was regarded as fundamental was always regarded as a sign of confusion. Curiously enough, this attitude is found in thinkers who otherwise have very little in common. This can be seen from an examination of various criticisms of the pre-Socratic philosophers that have been proffered in the course of history.

As has been shown by Popper [67], these early philosophers were not only the inventors of a theoretical science (as opposed to a science which is content with assembling empirical generalizations as was the physics, the mathematics, and the astronomy of the Egyptians), but they also invented the method characteristic for this kind of science, i.e., the method of test within a class of mutually inconsistent, partly overlapping, and to that extent empirically adequate theories. All this was thoroughly misunderstood. Thus the sophists are reported to have ridiculed the Ionians by pointing out that their motto seemed to be "To every philosopher his own principle." Plato made full use of this popular

sentiment (*Sophist* 242ff), and so did the church fathers later on: "As of canonical authors," writes St. Augustine (quoted from [16], p. 132), "God forbid that they should differ . . . [But] let one look amongst all the multitude of philosophers' writings, and if he finds two that tell both one tale in all respects, it may be registered for a rarity." Soon after the rise of Baconian empiricism with its so very different message, the variety of the theories discussed by the pre-Socratics was used as an example of where one gets when leaving the solid ground, not of revelation, but of *experience*. The following quotation is very characteristic: "As to the particular tenets of Thales, and his successors of the *Ionian* school, the sum of what we learn from the imperfect accounts we have of them is that each overthrew what his predecessor had advanced; and met with the same treatment himself from his successor . . . So early did the passion for systems begin."[88] In fact, nearly all inventors of new methods in philosophy and in the sciences have been inspired by the hope that they would be able to put an end to the quarrel of the schools and to establish the one true body of knowledge (this inspiration is present even today in some of the defenders of the Copenhagen interpretation of the quantum theory). What emerges from our own considerations is that *once dispute has been made empirically testable* (and this was already the case with the early Ionians—see again Popper's article [67]), *it becomes an essential element of the development of knowledge*; and it also emerges that cessation of dispute is not to be regarded as a sign that now we have finally arrived at the truth, but rather as a sign of fatigue in those originally participating in the dispute (just as the more recent return to religious beliefs is a sign of fatigue and of despair in the capabilities of reason).

According to Popper this procedure of testing a theory by comparing it with experience, as seen in the light of alternatives, is identical with the scientific method. Professor Matson, who also emphasizes that "the key" to the method of the pre-Socratics lies in the fact that "they were not dogmatic about their predecessors, but rather . . . criticized them most acutely,"[89] differs from Popper insofar as he regards this method as relevant either to part of the sciences only (he regards it as relevant to contemporary cosmology), or even as a nondogmatic alternative to

[88] McLaurin [55], p. 28.
[89] [57], p. 445.

empiricism (for Popper the test procedure outlined *is* the empirical method). That the pre-Socratics were inventors of *method* as well as *theories* is emphasized in Professor Matson's admirable paper on Anaximander [56]. As far as I can make out, Popper's views on the matter were already developed at the time of the first publication of his *Open Society*. Both these thinkers represent a minority view which, in my opinion, is of the greatest importance for the understanding of the history of early Greek philosophy.

This finishes my criticism of (A) and (5). (A) has been shown to be in disagreement not only with actual scientific practice but also with the principles of a sound empiricism. The account of theorizing given in the introduction has been shown to be superior to the hierarchy of axioms and theorems which seems to be the favorite model of contemporary empiricism. The use of a set of mutually inconsistent and partially overlapping theories has been found to be of fundamental importance for methodology. The desideratum mentioned in connection with what has here been called the second idea has thereby been fulfilled. Serious doubt has been thrown upon the correctness and the desirability of (B). I now turn to the refutation of (B).

7. *Criticism of the Assumption of Meaning Invariance.*

In Section 5 it was shown that the "inertial law" (8) of the impetus theory is incommensurable with Newtonian physics in the sense that the main concept of the former, viz., the concept of impetus, can neither be defined on the basis of the primitive descriptive terms of the latter, nor related to them via a correct empirical statement. The reason for this incommensurability was also exhibited: although (8), *taken by itself*, is in quantitative agreement both with experience and with Newton's theory, the "rules of usage" to which we must refer in order to explain the meanings of its main descriptive terms contain the law (7) and, more especially, the law that constant forces bring about constant velocities. Both of these laws are inconsistent with Newton's theory. Seen from the point of view of this theory, any concept of a force whose content is dependent upon the two laws just mentioned will possess zero magnitude, or zero denotation, and will therefore be incapable of expressing features of actually existing situations. Conversely, it will be capable of being used in such a manner only if all connections with Newton's theory have first been severed. It is clear that this example refutes (B) if we interpret that thesis as the description of how science actually proceeds.

We may generalize this result in the following fashion: consider two theories, T′ and T, which are both empirically adequate inside D′, but which differ widely outside D′. In this case the demand may arise to explain T′ on the basis of T, i.e., to derive T′ from T and suitable initial conditions (for D′). Assuming T and T′ to be in quantitative agreement inside D′, such derivation will still be impossible if T′ is part of a theoretical context whose "rules of usage" involve laws inconsistent with T.[90]

It is my contention that the conditions just enumerated apply to many pairs of theories which have been used as instances of explanation and reduction. Many (if not all) of such pairs on closer inspection turn out to consist of elements which are incommensurable and therefore incapable of mutual reduction and explanation. However, the above conditions admit of still wider application and then lead to very important consequences with regard to the structure and development both of our knowledge and of the language used for the expression of it. After all, the principles of the context of which T′ is a part need not be explicitly formulated, and as a matter of fact they rarely are. To bring about the situation described above (sets of mutually incommensurable concepts), it is sufficient that they govern the use of the main terms of T′. In such a case T′ is formulated in an idiom some of whose implicit rules of usage are inconsistent with T (or with some consequences of T in the domain where T′ is successful). Such inconsistency will not be obvious at a glance; it will take considerable time before the incommensurability of T and T′ can be demonstrated. However, as soon as this demonstration has been carried out, in the very same moment, the idiom of T′ must be given up and must be replaced by the idiom of T. Of course, one need not go through the laborious and very uninteresting task of analyzing the context of which T′ is part.[91] All that is needed is the adoption of the terminology and the 'grammar' of the most detailed and most

[90] Since this difficulty can arise even in the domain of empirical generalizations, the orthodox account may be inappropriate for them as well.

[91] There are many philosophers (including my friends in the Minnesota Center) who would admit that the importance of linguistic analysis is very limited. However, they would still hold that its application is necessary in order to find out to what extent the advent of a new theory modifies the customary idiom. The considerations above would show that even this is granting too much and that one travels best without any linguistic ballast.

successful theory throughout the domain of its application.[92] This automatically takes care of whatever incommensurabilities may arise, and it does so without any linguistic detective work (which therefore turns out to be entirely unnecessary for the progress of knowledge).

What has just been said applies most emphatically to the relation between (theories formulated in) some commonly understood language and more abstract theories. That is, I assert that languages such as the "everyday language," this notorious abstraction of contemporary linguistic philosophy, frequently contain (not explicitly formulated, that is, but implicit in the way in which its terms are used) principles which are inconsistent with newly introduced theories, and that they must therefore be either abandoned and replaced by the language of the new and better theories even in the most common situations, or they must be completely separated from these theories (which would lead to a situation where it is possible to believe in various kinds of "truth"): it is far from correct to assume that the everyday languages are so widely conceived, so tolerant, indefinite, and vague that they will be compatible with any scientific theory, that science can at most fill in details, and that a scientific theory will never run against the principles implicitly contained in them. *The very opposite is the case.* As will be shown later, even everyday languages, like languages of highly theoretical systems, have been introduced in order to give expression to some theory or point of view, and they therefore contain a well-developed and sometimes very abstract ontology. It is very surprising that the champions of the "ordinary language" should have such a low opinion of its descriptive power.

However, before turning to this part of the argument, I shall briefly discuss another example where the questionable principles of T′ have been explicitly formulated, or can at least be easily unearthed.

The example which is dealt with by Nagel is the relation between phenomenological thermodynamics and the kinetic theory. Employing his own theory of reduction and, more especially, the condition I have quoted in the text adjacent to my footnote 11, Nagel claims that the

[92] One hears frequently that a *complete* replacement of the grammar and the terminology of the "old language" is impossible because this old language will be needed for introducing the new language and will, therefore, infect at least part of the new language. This is curious reasoning indeed if we consider that children learn languages without the help of a previously known idiom. Is it really asserted that what is possible for a small child will be impossible for a philosopher, a linguistic philosopher at that?

terms of the statements which have been derived from the kinetic theory (with the help of correlating hypotheses similar to (9)) will have the meanings they originally possessed within the phenomenological theory, and he repeatedly emphasizes that these meanings are fixed by "its own procedures" (i.e., by the procedures of the phenomenological theory) "whether or not [this theory] has been, or will be, reduced to some other discipline."[93]

As in the case of the impetus theory, we shall begin our study of the correctness of this assertion with an examination of these "procedures" and "usages"; more especially, we shall start with an examination of the usage of the term 'temperature,' "as fixed by the established procedures" of thermodynamics.

Within thermodynamics proper,[94] temperature ratios are defined by reference to reversible processes of operating between two levels, L′ and L″, each of these levels being characterized by one and the same temperature throughout. The definition, viz.,

(10) $T':T'' = Q':Q''$

identifies (after a certain arbitrary choice of units) the ratio of the temperature with the ratio between the amount of heat absorbed at the higher level and the amount of heat rejected at the lower level. Closer inspection of the "established usage" of the temperature thus defined shows that it is supposed to be

(11) independent of the material of the substance chosen for the cycle, and unique.

This property can be inferred from the extension of the concept of temperature thus defined to radiation fields and from the fact that the constants of the main laws in this domain are universal, rather than dependent upon either the thermometric substance or the substance of the system investigated.

Now, it can be shown by an argument not to be presented here that (10) and (11) taken together imply the second law of thermodynamics in its strict (phenomenological) form: the concept of temperature as "fixed by the established usages" of thermodynamics is such that its application to concrete situations entails the strict (i.e., nonstatistical) second law.

[93] Nagel [20], p. 301.
[94] See Fermi [22], Sec. 9.

Now whatever procedure is adopted, the kinetic theory does *not* give us such a concept. First of all, there does not exist any dynamical concept that possesses the required property.[95] The statistical account, on the other hand, allows for fluctuations of heat back and forth between two levels of temperature and, therefore, again contradicts one of the laws implicit in the "established usage" of the thermodynamic temperature. The relation between the thermodynamic concept of temperature and what can be defined in the kinetic theory, therefore, can be seen to conform to the pattern that has been described at the beginning of the present section: we are again dealing with two incommensurable concepts. The same applies to the relation between the purely thermodynamic entropy and its statistical counterpart; whereas the latter admits of very general application, the former can be measured by infinitely slow reversible processes only. Taking all this into consideration we must admit that it is impossible to relate the kinetic theory and the phenomenological theory in the manner described by Nagel, or to explain all the laws of the phenomenological theory in the manner demanded by Hempel and Oppenheim on the basis of the statistical theory. Again replacement rather than incorporation, or derivation (with the help, perhaps, of premises containing statistical as well as phenomenological concepts), is seen to be the process that characterizes the transition from a less general theory to a more general one.

It ought to be pointed out that the discussion is very idealized. The reason is that a purely kinetic account of the phenomena of heat does not yet seem to exist. What exists is a curious mixture of phenomenological and statistical elements, and it is this mixture which has received the name 'statistical thermodynamics.' However, even if this is admitted, it remains that the concept of temperature as it is used in this new and mixed theory is different from the original, purely phenomenological concept. To our point of view, according to which terms change their meanings with the progress of science, Nagel raises the following objection: "The redefinition of expressions with the development of inquiry [so it is noted], is a recurrent feature in the history of science. Accordingly, though it must be admitted that in an earlier use the word 'temperature' had a meaning specified exclusively by the rules and procedures of thermometry and classical thermodynamics, it is *now* so used

[95] I shall not discuss, in the present paper, the somewhat different situation with respect to the first law.

that temperature is 'identical by definition' with molecular energy. The deduction of Boyle-Charles' law does not therefore require the introduction of a further postulate, whether in the form of a coordinating definition or a special empirical hypothesis, but simply makes use of this definitional identity. This objection illustrates the unwitting double talk into which it is so easy to fall. It is certainly possible to redefine the word 'temperature' so that it becomes synonymous with 'mean kinetic energy.' But it is equally certain that on this redefined usage the word has a different meaning from the one associated with it in the classical science of heat, and therefore a meaning different from the one associated with the word in the statement of the Boyle-Charles law. However, if thermodynamics is to be reduced to mechanics, it is temperature in the sense of the term in the classical science of heat which must be asserted to be proportional to the mean kinetic energy of gas molecules. Accordingly, if the word 'temperature' is redefined as suggested by the objection, the hypothesis must be invoked that the state of bodies described as 'temperature' (in the classical thermodynamic sense) is also characterized by 'temperature' in the redefined sense of the term. This hypothesis, however, will then be one that does not hold as a matter of definition . . . Unless this hypothesis is adopted, it is not the Boyle-Charles law which can be derived from the assumptions of the kinetic theory of gases. What is derivable without the hypothesis is a sentence similar in syntactical structure to the standard formulation of the law, but possessing a sense that is unmistakably different from what the law asserts."[96] So far Nagel.

Commencing my criticism, I shall at once admit the correctness of the last assertion. After all, it has been my contention all through this paper that extension of knowledge leads to a decisive modification of the previous theories both as regards the quantitative assertions made and as regards the meanings of the main descriptive terms used. Applying this to the present case I shall therefore at once admit that incorporation into the context of the statistical theory is bound to change the meanings of the main descriptive terms of the phenomenological theory. The difference between Nagel and myself lies in the following. For me, such a change to new meanings and new quantitative assertions is a natural occurrence which is also desirable for methodological reasons

[96] [61], pp. 357–358.

(the last point will be established later in the present section). For Nagel such a change is an indication that reduction has not been achieved, for reduction in Nagel's sense is supposed to leave untouched the meanings of the main descriptive terms of the discipline to be reduced (cf. his "if thermodynamics is to be reduced to mechanics, it is temperature in the sense of the term in the classical science of heat which must be asserted to be proportional to the mean kinetic energy of gas-molecules"). "Accordingly," he continues, quite obviously assuming that reduction in his sense can be carried through, "if the word 'temperature' is redefined as suggested by the objection, the hypothesis must be invoked that the state of bodies described as 'temperature' (in the classical thermodynamic sense) is also characterized by 'temperature' in the redefined sense of the term. *This* hypothesis . . . will then be one that does not hold as a matter of definition." It will also be a false hypothesis because the conditions for the definition of the phenomenological temperature are *never* satisfied in nature (see the arguments above in the text and compare also the arguments in connection with formula (9)), which is only another sign of the fact that reduction, in the sense of Nagel, of the phenomenological theory to the statistical theory is not possible (obviously the additional premises used in the reduction are not supposed to be false). Once more arguments of meaning have led to quite unnecessary complications.

Further examples exhibiting the same features can be easily provided. Thus in classical, prerelativistic physics the concept of mass (and, for that matter, the concept of length and the concept of time duration) was absolute in the sense that the mass of a system was not influenced (except, perhaps, causally) by its motion in the coordinate system chosen. Within relativity, however, mass has become a relational concept whose specification is incomplete without indication of the coordinate system to which the spatiotemporal descriptions are all to be referred. Of course, the values obtained on measurement of the classical mass and of the relativistic mass will agree in the domain D′, in which the classical concepts were first found to be useful. This does not mean that what is measured is the same in both cases: what is measured in the classical case is an *intrinsic property* of the system under consideration; what is measured in the case of relativity is a *relation* between the system and certain characteristics of D′. It is also impossible to define the exact classical concepts in relativistic terms or to relate them with the help of an

empirical generalization. Any such procedure would imply the false assertion that the velocity of light is infinitely large. It is therefore again necessary to abandon completely the classical conceptual scheme once the theory of relativity has been introduced; and this means that it is imperative to use relativity in the theoretical considerations put forth for the explanation of a certain phenomenon as well as in the observation language in which tests for these considerations are to be formulated; after all, the empirically untenable consequences of the attempts above to give a reduction of classical terms to relativistic terms emerges whether or not the elements of the definition belong to the observation language.

Many more examples can be added to those discussed in the present paper (viz., the impetus theory, phenomenological thermodynamics, and the classical conception of mass). All these examples show that the postulate of meaning invariance is incompatible with actual scientific practice. That is, it has been shown that in most cases it is impossible to relate successive scientific theories in such a manner that the key terms they provide for the description of a domain D′, where they overlap and are empirically adequate, either possess the same meanings or can at least be connected by empirical generalizations. It is also clear that the methodological arguments against meaning invariance will be the same as the arguments against the derivability condition and the consistency condition. After all, the demand for meaning invariance implies the demand that the laws of later theories be compatible with the principles of the context of which the earlier theories are part, and this demand is, therefore, seen to be a special case of condition (5). Using our earlier arguments against (5) we may now infer the untenability, on methodological grounds, of meaning invariance as well. And as our argument is quite general we may also infer that it is undesirable that the "ordinary" usage of terms be preserved in the course of the progress of knowledge. Wherever such preservation is observed, we shall feel inclined to think that the suggested new theories are not as revolutionary as they perhaps ought to be, and we shall have the suspicion that some *ad hoc* procedures have perhaps been adopted. Violation of ordinary usage, and of other "established" usages, on the other hand, is a sign that real progress has been made, and it is to be welcomed by anybody interested in such progress (provided of course that this violation is connected with the suggestion of a new point of view or a new theory and is not just the result of linguistic arbitrariness).

Our argument against meaning invariance is simple and clear. It proceeds from the fact that usually some of the principles involved in the determination of the meanings of older theories or points of view are inconsistent with the new, and better, theories. It points out that it is natural to resolve this contradiction by eliminating the troublesome and unsatisfactory older principles and to replace them by principles, or theorems, of the new and better theory. And it concludes by showing that such a procedure will also lead to the elimination of the old meanings and thereby to the violation of meaning invariance.

The most important method used for escaping the force of this clear and simple argument is the transition to instrumentalism. Instrumentalism maintains that the new theory must not be interpreted as a series of statements but that it is rather to be understood as a predictive machine whose elements are tools rather than statements and therefore cannot be incompatible with any principle already in existence. This very popular move (popular, that is, because used also by scientists) admittedly cuts the ground from beneath our argument and makes it inapplicable. However, it has never been explained why a new and satisfactory theory should be interpreted as an instrument, whereas the principles behind the established usage, which can easily be shown to be empirically inadequate, are not so interpreted. After all, the only advantage of the latter is that they are *familiar*—an advantage which is a psychological and historical accident and which should therefore not have any influence upon questions of interpretation and of reality. One may try to answer this criticism by ascribing an instrumental function to all principles, old or new, and not only to those contained in the most recent theory. Such a procedure means acceptance of a sense-data account of knowledge. Having shown elsewhere[97] that such an account is impossible, I can now say that this consequence of a universal instrumentalism is tantamount to its refutation. Result: neither a restricted nor a universal instrumentalism can be carried through in a satisfactory manner. This disposes of the instrumentalistic move.

While instrumentalism possesses at least a semblance of plausibility, the arguments to be discussed now are devoid even of this feature. Indeed, I am very hesitant to apply the word 'arguments' to these expressions of confused thinking, their wide acceptance and asserted self-evi-

[97] [38].

dence notwithstanding. Consider for example the following question (which is supposed to be a criticism of our suggestion that after the acceptance of the kinetic theory the word 'temperature' will be in need of reinterpretation):[98] "If the meaning of 'temperature' is [now] the same as that of 'mean kinetic energy of molecular motion,' what are we talking about when milk is said to have a temperature of 10° Cels? Surely not the kinetic energy of the molecular constituents of the liquid, for the uninstructed layman is able to understand what is here said without possessing any notion about the molecular composition of the milk."

Now it may be quite correct that the "uninstructed layman"[99] does not think of molecules when speaking about the temperature of his milk and that he has not the slightest notion of the molecular constitution of the liquid either. However, what has the reference to him got to do with our argument according to which a person who has already accepted *and understood* the theory of the molecular constitution of gases, liquids, and solids cannot at the same time demand that the premolecular concept of temperature be retained? It is not at all denied by our argument that the "uninstructed layman" may possess a concept of temperature that is very different from the one connected with the molecular theory (after all, some "uninstructed laymen," intelligent clergymen included, still believe in ghosts and in the devil). What is denied is that anybody can consistently continue using this more primitive concept and at the same time believe in the molecular theory. Again, this does not mean that a person may not, on different occasions, use concepts which belong to different and incommensurable frameworks. The only thing that is forbidden for him is the use of both kinds of concepts *in the same argument*; for example, he may not use the one kind of concept in his observation language and the other kind in his theoretical language. Any such combination—and this is the gist of our considerations in the pres-

[98] The argument in connection with this question can be found in Nagel [20], p. 293. It is not clear to me whether or not Nagel would be prepared to support the argument.

[99] By the way, who *is* this uninstructed layman? From the purpose for which he is being employed in many arguments, it would seem to emerge that he is not supposed to know much science, or much politics, or much religion, or much of anything. This means that in these times of mass communication and mass education he must be very careful not to read the wrong parts of his newspaper, he must be careful not to leave his television set on for too long a time, and he must also not allow himself to converse too much with his friends, his children, etc. That is, he must be either a savage or an idiot. I really wonder what are the motives which lead to a philosophy where the most interesting language is the language of savages or of idiots.

ent section—would introduce principles which are mutually inconsistent and thereby destroy the argument in which it is supposed to occur. It is evident that this position is not at all endangered by the objection implied by the question above.

However, quite apart from being so obviously irrelevant to our thesis, the objection reflects an attitude that must appear quite incredible to anybody who possesses even the slightest acquaintance with the history of knowledge. The question insinuates that the layman's ability to handle the word 'temperature' according to the rules prescribed for it in some simple idiom indicates his understanding of the thermal properties of bodies. It insinuates that the existence of an idiom allows us to infer the truth of the principles which underlie this idiom. Or, to be more specific, it insinuates that *what is being used is, on that account alone, already exhibited as adequate, useful, and perhaps irreplaceable*. After all, the reference to the layman's understanding of the word 'temperature' is not made without purpose. It is made with the purpose of preserving the common meaning of this word since, it is alleged, this common meaning can be understood and is not in need of replacement. The discussion of a specific example will at once show the detrimental effect of any such procedure.

The example chosen now brings us to the second part of the present section where the relation is investigated, not between explicitly formulated theories, but between a theory and the implicit principles that govern the usage of the descriptive terms of some idiom. As has been said a little earlier, it is our conviction that "everyday languages," far from being so widely and generally conceived that they can be made compatible with any scientific theory, contain principles that may be inconsistent with some very basic laws. It was also pointed out that these principles are rarely expressed in an explicit manner (except, perhaps, in those cases where there is an attempt to defend the corresponding idiom against replacement or change) but that they are implicit in the rules that govern the use of its main descriptive terms. And our point was that, once these principles are found to be empirically inadequate, they must be given up and with them the concepts that are obtained by using terms in accordance with them. Conversely, the attempt to retain these concepts will lead to the conservation of false laws and to a situation where every connection between concepts or facts is severed.

The example which I have chosen to show this involves the pair 'up-down.' There existed a time when this pair was used in an absolute fashion, i.e., without reference to a specified center, such as the center of the earth. That it was used in such a manner can be easily seen from the "vulgar" remark that the antipodes would "fall off" the earth if the earth were spherical,[100] as well as from the more sophisticated attempts of Thales, Xenophanes, and others to find support for the earth as a whole, assuming that it would otherwise fall "down."[101] These attempts, as well as that remark about the antipodes employ two assumptions: first, that any material object is under the influence of a force; second, that this force acts in a privileged direction in space and must therefore be regarded as anisotropic. It is this privileged direction to which the pair 'up-down' refers. The second assumption is not explicitly made; it can only be derived from the way in which the pair 'up-down' is used in arguments such as those mentioned above.[102] We have here an example of a cosmological assumption (anisotropic character of space) implicit in the common idiom.

This example refutes the thesis which has been defended by some philosophers that "everyday languages" are fairly free from hypothetical elements and therefore ideally suited as observational languages.[103] It refutes the thesis by showing that even the most harmless part of a common idiom may rest upon very far-reaching hypotheses and must therefore be regarded as hypothetical to a very high degree.

Another remark concerns the changes of meaning needed once the Newtonian (or perhaps even the Aristotelian) explanation of the fall of heavy bodies is adopted. Newtonian space is isotropic and homogeneous. Hence, accepting this theory, one cannot anymore use the pair 'up-down' in the previous fashion and at the same time assume that one is describing actual features of physical situations. More especially, one cannot retain the absolute use of this pair for the description of observable features, since such features are quite obviously assumed to exist. Any per-

[100] For a discussion of this remark and of a related "vulgar" remark concerning the shape and arrangement of the terrestrial waters, see Pliny, *Natural History*, II, 161–166, quoted in Cohen and Drabkin [19], pp. 159–161.

[101] For a description and criticism of these attempts, see [1], 294a12ff; also quoted in [19], pp. 143–148.

[102] For the atomist's conception of space, which, at least since Epicurus, seems to be influenced by the popular ideas discussed above, cf. M. Jammer [48], p. 11.

[103] This thesis was introduced by Professor Herbert Feigl in discussions with me. For my own position, see also Philipp Frank [40].

son accepting Newton's physics and the conception of space it contains must, therefore, give a new meaning even to such a familiar pair of terms as is the pair 'up-down,' and he must now interpret it as a relation between the direction of a motion and a center that has been fixed in advance. And as Newton's theory is preferable, on empirical grounds, to the older and "absolute" cosmology, it follows that the relational usage of the pair 'up-down' will be preferable, too. Conversely, the attempt to retain the old usage amounts to retaining the old cosmology, and this despite the discoveries which have shown it to be obsolete.

To this argument it may be, and has been, objected that the "vulgar" usage of the pair 'up-down' was never supposed to be so general as to be applicable to the universe as a whole. This may be the case (although I do not see any reason for assuming that "ordinary" people are so very cautious as to apply the pair to the surface of the earth only; all the passages referred to in the above quotations contradict this assumption and so does the fact that at all times *real* ordinary people—and not only their Oxford substitutes—were very much interested in celestial phenomena[104]). However, even such a restriction would not invalidate our argument. It would rather show that the pair was used for singling out an absolute direction near the surface of the earth and that it did not assume such a direction to exist throughout the universe. It is clear that even this modest position is incompatible with the ideas implicit in the Newtonian point of view, which does not allow for local anisotropies either.

Consider now, after this example, the following argument in favor of the thesis that what is being used is, on that account alone, already exhibited as adequate, useful, and perhaps irreplaceable. The argument is the late Professor Austin's, and it has been repeated by G. J. Warnock.[105] "Language," writes Warnock, "is to be used for a vast number of highly important purposes; and it is at the very least unlikely that it should contain either much more, or much less, than those purposes require. If so, the existence of a number of different ways of speaking is very likely indeed to be an indication that there is a number of different things to be said . . . Where the topic at issue really is one that does

[104] The reason why Oxford philosophers so rarely discuss the influence of astronomy upon everyday languages may perhaps be found in the weather of their favorite discussion place. However, this reason unfortunately does not explain their ignorance in physics, theology, mythology, biology, and even linguistics.

[105] [71], pp. 150–151.

constantly concern most people in some practical way—as for example perception, the ascription of responsibility, or the assessment of human character or conduct—then it is certain that everyday language is as it is for some extremely good reasons; its verbal variety is certain to provide clues to important distinctions."[106]

If I understand this passage correctly, it means that the existence of certain distinctions in a language may be taken as an indication of similar distinctions in the nature of things, situations, and the like. And the reason for this is that people who are in constant contact with things and situations will soon develop the correct linguistic means for describing their properties. In short: human beings are good inductive machines in domains of concentrated interest, and their inductive ability will be the better the greater their concern, or the greater the practical value of the topic treated. Consequently, languages containing distinctions of practical interest are very likely to be adequate and irreplaceable.

There are many objections against this train of reasoning. First of all, it would seem to be somewhat arbitrary to restrict interests to those which can be derived from the immediate necessities of the physical life of the human race. From history we learn that the motives emerging from abstract considerations such as those found in a myth, or in a theo-astronomical system, are at least as strong as the more pedestrian motives connected with the immediate fulfillment of material needs (after all, people have died for their convictions!). Now if a language can be trusted because of the commitment of those who use it and if commitment is found to range over a much wider area than has first been imagined, if it is found to range over physics, astronomy (think of Giordano Bruno!), and biology, then the result will be that the principle we are discussing at the present moment (viz., the principle that what is being used for a purpose is on that account alone already useful and irreplaceable) must be applied to any language and any theory that has ever been developed and seriously tested. However—and this is the second point—there exist many theories and languages which have been found to be inadequate, and this despite their usefulness and despite the zeal of those who had developed them. This applies to the language of

[106] Astronomy is again omitted. It would seem to me that problems of astronomy had a much greater influence upon the formation of our language than problems of perception, which are of a very ephemeral nature and also are very technical. The skies and the stars (which, after all, were assumed to be gods) were everyone's concern.

the Aristotelian physics, which had to be introduced into medieval thinking under very great difficulties and whose influence went much further than is sometimes realized; it applies to the language of the physics of Newton (mechanicism); and it applies to many other languages. Of course, this is the result one would expect: success under even very severe tests does not guarantee infallibility; no amount of commitment and no amount of success can guarantee the perennial reliability of inductions.

The principle which we have been discussing just now does not occur only in philosophy. Bohr's contention[107] that the account of all quantum mechanical evidence must forever "be expressed in classical terms" has been defended in a very similar manner. According to Bohr, we need our classical concepts not only if we want to give a summary of facts; without these concepts the facts to be summarized could not even be stated. As Kant before him, Bohr observes that our experimental statements are always formulated with the help of certain theoretical terms and that the elimination of these terms would lead, not to the "foundations of knowledge" as a positivist would have it, but to complete chaos. "Any experience," he asserts, "makes its appearance within the frame of our customary points of view and forms of perception"[108]—and at the present moment the forms of perception are those of classical physics.

Now does it follow, as is asserted by Bohr, that we can *never* go beyond the classical framework and that therefore all our future microscopic theories will have to use the notion of complementarity as a fundamental notion?

It is quite obvious that the actual use of classical concepts for the description of experiments within contemporary physics can never justify such an assumption, even if these concepts happen to have been very successful in the past (Hume's problem). For a theory may be found whose conceptual apparatus, when applied to the domain of validity of classical physics, would be just as comprehensive and useful as the classical apparatus without coinciding with it. Such a situation is by no means uncommon. The behavior of the planets, of the sun, and of the satellites can be described both by the Newtonian concepts and by the concepts

[107] See p. 43 above.

[108] [5], p. 1. For a more detailed account of what follows, see [32], [34], [36], [37], [38].

of general relativity. The order introduced into our experiences by Newton's theory is retained and improved upon by relativity. This means that the concepts of relativity are sufficiently rich for the formulation of all the facts which were stated before with the help of Newtonian physics. Yet the two sets of concepts are completely different and bear no logical relation to each other.

Other examples of the same kind can be provided very easily. What we are dealing with here is, of course, again the old problem of induction. No number of examples of usefulness of an idiom is ever sufficient to show that the idiom will have to be retained forever. And if it is objected, as it has been in the case of the quantum theory, that the language of classical physics is the only actual language in existence for the description of experiments,[109] then the reply must be that man is not only capable of using theories and languages but that he is also capable of inventing them.[110] How else could it have been possible, to mention only one example, to replace the Aristotelian physics and the Aristotelian cosmology with the new physics of Galileo and Newton? The only conceptual apparatus then available was the Aristotelian theory of change with its opposition of actual and potential properties, the four causes, and the like. Within this conceptual scheme, which was also used for the description of experimental results, Galileo's (or rather Descartes') law of inertia does not make sense, nor can it be formulated. Should, then, Galileo have followed Heisenberg's advice and have tried to get along with the Aristotelian concepts as well as possible, since his "actual situation . . . [was] such that [he did] use the Aristotelian concepts"[111] and since "there is no use discussing what could be done if we were other beings than we are"? By no means. What was needed was not improvement or delimitation of the Aristotelian concepts; what was needed was an entirely new theory. This concludes our argument against the principle that a useful language is to be regarded as adequate and irreplaceable and, thereby, fully restores the force of our attack against meaning invariance, as well as reinforces the positive suggestions made in connection with this attack and especially the idea that conceptual changes may occur anywhere in the system that is employed at a certain time for the explanation of the properties of the world we live in.

[109] See Heisenberg [44], p. 56, and von Weizsaecker [72], p. 110.

[110] See also fn. 92.

[111] This is a paraphrase of a passage in Heisenberg [44], p. 56.

As I indicated in the introductory discussion, this transition from a point of view which demands that certain 'basic' terms retain their meaning, come what may, to a more liberal point of view which allows for changes anywhere in the system employed is bound to influence profoundly our attitude with respect to many philosophical problems and will also facilitate their solution. Let me take the mind-body problem as an example. It seems to me that the difficulties of this problem are to be sought precisely in the fact that meaning invariance is regarded as a necessary condition of its satisfactory solution. That is, it is assumed, or even demanded, that the meanings of at least some terms of the problem must remain constant throughout the discussion of the problem and further that these terms must retain their meanings in the solution as well.

Of course, different schools will apply the demand for meaning invariance to different concepts. A Platonist will demand that terms such as 'mind' and 'matter' remain unchanged, whereas an empiricist will require that some observational terms, such as the term 'pain,' or the more abstract term 'sensation,' retain their (common) meaning. Now a closer analysis of these key terms will, I think, reveal that they are incommensurable in exactly the sense in which this term has been defined at the beginning of the present section. This being the case, it is of course completely impossible either to reduce them to each other, or to relate them to each other with the help of an empirical hypothesis, or to find entities which belong to the extension of both kinds of terms. That is, the conditions under which the mind-body problem has been set up as well as the particular character of its key terms are such that a solution is forever impossible: a solution of the problem would require relating what is incommensurable without allowing for a modification of meanings which would eliminate this incommensurability.

All these difficulties disappear if we are prepared to admit that, in the course of the progress of knowledge, we may have to abandon altogether a certain point of view and the meanings connected with it—for example, if we are prepared to admit that the mental connotation of mental terms may be spurious and in need of replacement by a physical connotation according to which mental events, such as pains, states of awareness, and thoughts, are complex physical states of either the brain or the central nervous system, or perhaps the whole organism. I personally happen to

favor this idea that at some time sensations will turn out to be fairly complex central states which therefore possess a definite location inside the human body (which need not coincide with the place where the sensation is *felt* to be). I also hope that it will be possible to carry out a similar analysis of all so-called mental states.

Now whatever the merit of this belief of mine, it cannot be refuted by reference to the fact that what we "mean" by a sensation, or by a thought, is nothing that could have a location,[112] an internal structure, or physical ingredients. For if my belief is correct, and if it is indeed possible to develop a "materialistic" theory of human beings, then we shall of course be forced to abandon the "mental" connotations of the mental terms, and we shall have to replace them by physical connotations. According to the point of view which I am defending in the present paper, the only legitimate way of criticizing such a procedure would be to criticize this new materialistic theory by either showing that it is not in agreement with experimental findings or pointing out that it possesses some undesirable formal features (for example, by pointing out that it is *ad hoc*). Linguistic counter arguments have, I hope, been shown to be completely irrelevant.

The considerations in these last paragraphs are of course very sketchy. Still I hope that they give the reader an indication of the tremendous changes implied by the renunciation of the principle of meaning invariance as well as of the nefarious influence this principle has had upon traditional philosophy (modern empiricism included).

8. *Summary and Conclusion.*

Two basic assumptions of the orthodox theory of reduction and explanation have been found to be in disagreement with actual scientific practice and with reasonable methodology. The first assumption was that the explanandum is *derivable* from the explanans. The second assumption was that *meanings are invariant* with respect to the process of reduction and explanation. We may sum up the results of our investigation in the following manner:

Let us assume that T and T′ are two theories satisfying the conditions outlined at the beginning of Section 3. Then, from the point of view of scientific method, T will be most satisfactory if it is (*a*) *inconsistent*

[112] See Wittgenstein's investigation of the "grammar" of mental terms as presented in [74].

with T′ in the domain where they both overlap;[113] and if it is (β) *incommensurable* with T′.

Now it is clear that a theory which is satisfactory according to the criterion just pronounced will not be capable of functioning as an explanans in any explanation or reduction that satisfies the principles put forth by Hempel and Oppenheim or Nagel. Paradoxically speaking: *Hempel-Oppenheim explanations cannot use satisfactory theories as explanantia. And satisfactory theories cannot function as explanantia in Hempel-Oppenheim explanations.* How is the theory of explanation and reduction to be changed in order to eliminate this very undesirable paradox?

It seems to me that the changes that are necessary will make it impossible to retain a formal theory of explanation, because these changes will introduce pragmatic or "subjective" considerations into the theory of explanation. This being the case, it seems perhaps advisable to eliminate altogether considerations of explanation from the domain of scientific method and to concentrate upon those rules which enable us to compare two theories with respect to their formal character and their predictive success and which guarantee the constant modification of our theories in the direction of greater generality, coherence, and comprehensiveness. I shall now give a more detailed outline of the reasons which have prompted me to adopt this pragmatic point of view.

Consider again T and T′ as described above. Under these circumstances, the set of laws T″ following from T inside D′ will either be inconsistent with T′ or incommensurable with it. In what sense, then, can T be said to explain T′? This question has been answered by Popper for the case of the inconsistency of T′ and T″. "Newton's theory," he says, "unifies Galileo's and Kepler's. But far from being a mere conjunction of these two theories—which play the part of *explicanda* for Newton—*it corrects them while explaining them*. The original explanatory task was the deduction of the earlier results. It is solved, not by deducing them, but by deducing something better in their place: new results which, under the special conditions of the older results, come numerically very close to these older results, and at the same time correct them. Thus the empirical success of the old theory may be said to corroborate the new theory; and in addition, the corrections may be tested in their turn . . . What is brought out strongly by [this] . . . situation . . . is the fact that the new theory cannot possibly be *ad hoc* . . . Far from repeating its *explicandum*, the new theory contradicts it and corrects it. In this way, even the evidence of the *explicandum* itself becomes independent evidence for the new theory."[114]

In a letter to me, J. W. N. Watkins has suggested that this theory may be summarized as follows: Explanation consists of two steps. The first step is derivation, from T, of those laws which obtain under the conditions characterizing D′. The second step is comparison of T″ and T′ and realization that both are empirically adequate, i.e., fall within the domain of uncertainty of the observational results. Or, to express it in a more concise manner: T explains T′ satisfactorily only if T is true and there exists a consequence T″ of T for the conditions of validity of T′ such that T″ and T′ are at least equally strong and also experimentally indistinguishable.

The first question that arises in connection with Dr. Watkins' formulation is this: experimentally indistinguishable on the basis of which observations? T′ and T″ may be indistinguishable by the crude methods used at the time when T was first suggested, but they may well be distinguishable on the basis of later and more refined methods. Reference to a certain observational method will therefore have to be included in the clause of experimental indistinguishability. The notion of explanation will be relative to this observational material. It will not make sense any longer to ask whether or not T explains T′. The proper question will be whether T explains T′ *given the observational material, or the observational methods* O. Using this new mode of speech we are forced to deny that Kepler's laws are explained by Newton's theory *relative to the present observations*—and this is perfectly in order; for these present observations in fact refute Kepler's laws and thereby eliminate the demand for explanation. It seems to me that this theory can well deal with all the problems that arise when T and T′ are commensurable, but inconsistent inside D′. It does not seem to me that it can deal with the case where T′ and T are incommensurable. The reason is as follows.

As soon as reference to certain observational material has been included in the characterization of what counts as a satisfactory explanation, in the very same moment the question arises as to how this observational

[113] This condition has been discussed with great clarity in [66]. It was this discussion (as well as dissatisfaction with [60]) that was the starting point of the present analysis of the problem of explanation.

[114] Popper [66], p. 33.

material is to be presented. If it is correct, as has been argued all the way through the present paper, that the meanings of observational terms depend on the theory on behalf of which the observations have been made, then the observational material referred to in this modified sketch of explanation must be presented in terms of this theory also. Now incommensurable theories may not possess any comparable consequences, observational or otherwise. Hence, there may not exist any possibility of finding a characterization of the observations which are supposed to confirm two incommensurable theories. How, then, is the above account of explanation to be modified to cover the case of incommensurable theories also?[115]

It seems to me that the only possible way lies in closest adherence to the pragmatic theory of observation. According to this theory, it will be remembered, we must carefully distinguish between the *causes* of the production of a certain observational sentence, or the features of the process of production, on the one side, and the *meaning* of the sentence produced in this manner on the other. More especially, a sentient being must distinguish between the fact that he possesses a certain sensation, or disposition to verbal behavior, and the interpretation of the sentence being uttered in the presence of this sensation, or terminating this verbal behavior. Now our theories, apart from being pictures of the world, are also instruments of prediction. And they are good instruments if the information they provide, taken together with information about initial conditions characterizing a certain observational domain D_o, would enable a robot, who has no sense organs, but who has this information built into himself (or herself), to react in this domain in exactly the same manner as sentient beings who, without knowledge of the theory, have been trained to find their way about D_o and who are able to answer, 'on the basis of observation,' many questions concerning their surroundings.[116] This is the criterion of predictive success, and it is seen not at all to involve reference to the *meanings* of the reactions carried out either by the robot or by the sentient beings (which latter need not be humans, but can also be other robots). All it involves is agreement of behavior.

Now this criterion involves "subjective" elements. Agreement is demanded between the behavior of (nonsentient, but theory-fed) robots and that of sentient beings, and it is thereby assumed that the latter possesses a privileged position. Considering that perceptions are influenced by belief in theories and that behavior, too, is influenced by belief in theories, this criterion would seem to be somewhat arbitrary. It is easily seen, however, that it cannot be replaced by a less arbitrary and more "objective" criterion. What would such an objective criterion be? It would be a criterion which is either based upon behavior that is not connected with any theoretical element—and this is impossible (cf. my criticism of the theory of sense data above)—or it would be behavior that is tied up with an irrefutable and firmly established theory—which is equally impossible. We have to conclude, therefore, that a formal and "objective" account of explanation cannot be given.

[115] As Professor Feigl has pointed out to me, this difficulty also arises in the case of crucial experiments.

[116] Of course, the *motivations* of the robot and of the sentient being must also be the same.

REFERENCES

1. Aristotle. *De Coelo.*
2. Aristotle. *De Generatione et Corruptione.*
3. Barker, S. "The Role of Simplicity in Explanation," in *Current Issues in the Philosophy of Science*, H. Feigl and G. Maxwell, eds. New York: Holt, Rinehart, and Winston, 1961. Pp. 265–274.
4. Bohm, D. *Causality and Chance in Modern Physics.* London: Routledge and Kegan Paul, 1957.
5. Bohr, N. *Atomic Theory and the Description of Nature.* Cambridge: Cambridge University Press, 1932.
6. Bohr, N. "Discussions with Einstein," in *Albert Einstein, Philosopher-Scientist*, P. A. Schilpp, ed. Evanston, Ill.: Library of Living Philosophers, 1948. Pp. 201–241.
7. Born, M. *Natural Philosophy of Cause and Chance.* Oxford: Oxford University Press, 1948.
8. Brecht, B. *Schriften zum Theater.* Berlin and Frankfurt/Main: Suhrkamp Verlag, 1957.
9. Bunge, M. *Causality.* Cambridge, Mass.: Harvard University Press, 1959.
10. Burnet, J. *Early Greek Philosophy.* London: Adam and Charles Black, 1930.
11. Carnap, R. "Die Physikalische Sprache als Universalsprache der Wissenschaft," *Erkenntnis*, 2:432–465 (1932).
12. Carnap, R. "Psychologie in Physikalischer Sprache," *Erkenntnis*, 3:107–142 (1933).
13. Carnap, R. "Über Protokollsaetze," *Erkenntnis*, 3:215–228 (1933).
14. Carnap, R. "Testability and Meaning," *Philosophy of Science*, 3:419–471 (1936) and 4:1–40 (1937).
15. Carnap, R. "The Methodological Character of Theoretical Concepts," in *Minnesota Studies in the Philosophy of Science*, Vol. I, H. Feigl and M. Scriven, eds. Minneapolis: University of Minnesota Press, 1956. Pp. 38–76.
16. Clagett, M. *Greek Science in Antiquity.* London: Abelard-Schuman, 1957.
17. Clagett, M. *The Science of Mechanics in the Middle Ages.* Madison: University of Wisconsin Press, 1959.
18. Conant, J. B. *Case Histories in the Experimental Sciences*, Vol. I. Cambridge, Mass.: Harvard University Press, 1957.

19. Cohen, M. R., and I. E. Drabkin, eds. *A Source Book in Greek Science*. New York: McGraw-Hill, 1948.
20. Danto, A., and S. Morgenbesser, eds. *Philosophy of Science*. New York: Meridian Books, 1960.
21. Dewey, John. *The Quest for Certainty*. New York: Capricorn Books, 1960.
22. Fermi, E. *Thermodynamics*. New York: Dover Publications, 1956.
23. Feigl, H., and M. Brodbeck, eds. *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts, 1953.
24. Feigl, H., and G. Maxwell, eds. *Current Issues in the Philosophy of Science*. New York: Holt, Rinehart, and Winston, 1961.
25. Feigl, H., and M. Scriven, eds. *Minnesota Studies in the Philosophy of Science*, Vol. I. Minneapolis: University of Minnesota Press, 1956.
26. Feigl, H., M. Scriven, and G. Maxwell, eds. *Minnesota Studies in the Philosophy of Science*, Vol. II. Minneapolis: University of Minnesota Press, 1958.
27. Feyerabend, P. K. "Carnap's Theorie der Interpretation Theoretischer Systeme," *Theoria*, 21:55–62 (1955).
28. Feyerabend, P. K. "Wittgenstein's 'Philosophical Investigations,'" *Philosophical Review*, 54:449–483 (1955).
29. Feyerabend, P. K. "Eine Bemerkung zum Neumannschen Beweis," *Zeitschrift für Physik*, 145:421–423 (1956).
30. Feyerabend, P. K. "On the Quantum Theory of Measurement," in *Observation and Interpretation*, S. Körner, ed. London: Butterworth, 1957. Pp. 121–130.
31. Feyerabend, P. K. "An Attempt at a Realistic Interpretation of Experience," *Proceedings of the Aristotelian Society*, New Series, 58:143–170 (1958).
32. Feyerabend, P. K. "Complementarity," *Proceedings of the Aristotelian Society*, Supplementary Vol., 32:75–104 (1958).
33. Feyerabend, P. K. "Das Problem der Existenz Theoretischer Entitaeten," *Probleme der Wissenschaftstheorie*. Vienna: Springer, 1960. Pp. 35–72.
34. Feyerabend, P. K. "O Interpretacji Relacyj Nieokreslonosci," *Studia Filozoficzne*, 19:23–78 (1960).
35. Feyerabend, P. K. "Patterns of Discovery," *Philosophical Review*, 59:247–252 1960).
36. Feyerabend, P. K. "Professor Bohm's Philosophy of Nature," *British Journal for the Philosophy of Science*, 10:321–338 (1960).
37. Feyerabend, P. K. "Bohr's Interpretation of the Quantum Theory," in *Current Issues in the Philosophy of Science*, H. Feigl and G. Maxwell, eds. New York: Holt, Rinehart, and Winston, 1961. Pp. 371–390.
38. Feyerabend, P. K. "On the Interpretation of Microphysical Theories," to appear in *Minnesota Studies in the Philosophy of Science*, Vol. IV, H. Feigl and G. Maxwell, eds.
39. Feyerabend, P. K. "On the Interpretation of Scientific Theories," to appear in *Proceedings of the XIIth International Congress of Philosophy*, Milan.
40. Frank, P. *Relativity, a Richer Truth*. Boston: Beacon Press. 1950.
41. Fuerth, R. "Über einige Beziehungen Zwischen Klassischer Statistik und Quantenmechanik," *Zeitschrift für Physik*, 81:143–162 (1933).
42. Goodman, N. *Fact, Fiction, and Forecast*. Cambridge, Mass: Harvard University Press, 1955.
43. Hanson, N. R. *Patterns of Discovery*. Cambridge: Cambridge University Press, 1958.
44. Heisenberg, W. *Physics and Philosophy*. New York: Harper and Brothers, 1958.
45. Hempel, C. G. "Studies in the Logic of Confirmation," *Mind*, 54:1–26, 97–121 (1945).
46. Hempel, C. G. "A Logical Appraisal of Operationism," in *Validation of Scientific Theories*, P. Frank, ed. Boston: Beacon Press, 1954. Pp. 52–67.

47. Hempel, C. G., and P. Oppenheim. "Studies in the Logic of Explanation," *Philosophy of Science*, 15:135–175 (1948).
48. Jammer, M. *Concepts of Space*. Cambridge, Mass.: Harvard University Press, 1957.
49. Körner, S., ed. *Observation and Interpretation*. London: Butterworth, 1957.
50. Körner, S. *Conceptual Thinking*. New York: Dover Publications, 1960.
51. Kuhn, T. S. *The Copernican Revolution*. New York: Random House, 1959.
52. Landau, L. D., and E. M. Lifschitz. *Quantum Mechanics*. Reading, Mass.: Addison-Wesley, 1958.
53. Mach, E. *Waermelehre*. Leipzig: Johann Ambrosius Barth, 1897.
54. Mach, E. *Zwei Aufsaetze*. Leipzig: Johann Ambrosius Barth, 1912.
55. McLaurin, C. *An Account of Sir Isaak Newton's Philosophical Discoveries*. London: Buchanan's Head, 1750.
56. Matson, W. I. "The Naturalism of Anaximander," *Review of Metaphysics*, 6: 387–395 (1953).
57. Matson, W. I. "Cornford and the Birth of Metaphysics," *Review of Metaphysics*, 8:443–454 (1955).
58. Maier, A. *Die Vorlaeufer Galilei's im 14, Jahrhundert*. Rome: Edizioni di Storiae Litteratura, 1949.
59. Morris, E. "Foundation of the Theory of Signs," *International Encyclopaedia of Unified Science*, Sec. II/7. Chicago: University of Chicago Press, 1942.
60. Nagel, E. "The Meaning of Reduction in the Natural Sciences," in *Science and Civilization*, R. C. Stauffer, ed. Madison: University of Wisconsin Press, 1949. Pp. 99–145.
61. Nagel, E. *The Structure of Science*. New York: Harcourt, Brace, and Company, 1961.
62. Neumann, J. von. *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press, 1957.
63. Panofsky, E. "Galileo as a Critic of the Arts," *Isis*, 47:3–15 (1956).
64. Popper, K. R. "Naturgesetze und Theoretische Systeme," in *Gesetz und Wirklichkeit*, S. Moser, ed. Innsbruck: Hochschulverlag, 1948. Pp. 65–84.
65. Popper, K. R. *The Open Society and Its Enemies*. Princeton: Princeton University Press, 1950.
66. Popper, K. R. "The Aim of Science," *Ratio*, 1:24–35 (1957).
67. Popper, K. R. "Back to the Pre-Socratics," *Proceedings of the Aristotelian Society*, New Series, 54:1–24 (1959).
68. Popper, K. R. *The Logic of Scientific Discovery*. New York: Basic Books, 1959.
69. Reichenbach, H. *Experience and Prediction*. Chicago: University of Chicago Press, 1948.
70. Sellars, W. "The Language of Theories," in *Current Issues in the Philosophy of Science*, H. Feigl and G. Maxwell, eds. New York: Holt, Rinehart, and Winston, 1961. Pp. 57–77.
71. Warnock, J. *British Philosophy in 1900*. Oxford: Oxford University Press, 1956.
72. Weizsaecker, C. F. von. *Zum Weltbild der Physik*. Leipzig: Verlag Hirzel, 1954.
73. Whorff, B. L. *Language, Thought, and Reality: Selected Writings*, John B. Carroll, ed. Cambridge, Mass.: Technology Press of Massachusetts Institute of Technology, 1956.
74. Wittgenstein, L. *Philosophical Investigations*. Oxford: Basil Blackwell, 1953.