SYNTHESE LIBRARY / VOLUME 81

CAN THEORIES BE REFUTED?

Essays on the Duhem–Quine Thesis

Edited by Sandra G. Harding

D. REIDEL PUBLISHING COMPANY

DORDRECHT-HOLLAND / BOSTON-U.S.A.

CAN THEORIES BE REFUTED?

SYNTHESE LIBRARY

MONOGRAPHS ON EPISTEMOLOGY,

LOGIC, METHODOLOGY, PHILOSOPHY OF SCIENCE,

SOCIOLOGY OF SCIENCE AND OF KNOWLEDGE,

AND ON THE MATHEMATICAL METHODS OF

SOCIAL AND BEHAVIORAL SCIENCES

VOLUME 81

CAN THEORIES BE REFUTED?

Essays on the Duhem–Quine Thesis

Edited by

SANDRA G. HARDING
State University of New York at Albany

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND / BOSTON-U.S.A.

Library of Congress Cataloging in Publication Data

Main entry under title:

Can theories be refuted?

(Synthese library ; 81)
Includes bibliographies and index.
1. Science—Philosophy. 2. Science—Methodology.
3. Duhem, Pierre Maurice Marie, 1861–1916. 4. Quine, Willard Van Orman. I. Harding, Sandra G. II. Title: Duhem–Quine thesis.
Q175.C238 501 75–28339
ISBN-13: 978-90-277-0630-0 e-ISBN-13: 978-94-010-1863-0
DOI: 10.1007/978-94-010-1863-0

Published by D. Reidel Publishing Company,
P.O. Box 17, Dordrecht, Holland

Sold and distributed in the U.S.A., Canada, and Mexico
by D. Reidel Publishing Company, Inc.
Lincoln Building, 160 Old Derby Street, Hingham,
Mass. 02043, U.S.A.

To Dorian and Emily

TABLE OF CONTENTS

INTRODUCTION

According to a view assumed by many scientists and philosophers of science and standardly found in science textbooks, it is controlled experience which provides the basis for distinguishing between acceptable and unacceptable theories in science: acceptable theories are those which can pass empirical tests. It has often been thought that a certain sort of test is particularly significant: 'crucial experiments' provide supporting empirical evidence for one theory while providing conclusive evidence against another. However, in 1906 Pierre Duhem argued that the falsification of a theory is necessarily ambiguous and therefore that there are no crucial experiments; one can never be sure that it is a given theory rather than auxiliary or background hypotheses which experiment has falsified. W. V. Quine has concurred in this judgment, arguing that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body".

Some philosophers have thought that the Duhem–Quine thesis gratuitously raises perplexities. Others see it as doubly significant; these philosophers think that it provides a base for criticism of the foundational view of knowledge which has dominated much of western thought since Descartes, and they think that it opens the door to a new and fruitful way to conceive of scientific progress in particular and of the nature and growth of knowledge in general.

In this introductory essay, I shall indicate what considerations led Duhem and Quine to their views, and how some other leading philosophers and historians of science independently have arrived at similar conclusions. Then the major criticisms of the Duhem–Quine thesis will be presented. Finally I shall sketch the outlines of the rich and wide-ranging discussion of the implications of the Duhem–Quine thesis – a discussion which has occurred mainly in the last decade.

In his great book, *The Aim and Structure of Physical Theory*, Duhem was concerned with the way scientific theories were discussed by most scientists and philosophers of science of the late nineteenth century.

While these scientists and philosophers recognized that theories about nature could not be proved true, they did believe that by eliminating rival hypotheses through prescribed methods, science could finally reveal the residual, single, true description of nature. One kind of experiment was thought to be ideal for the purpose: 'crucial experiments' simultaneously refuted one hypothesis while verifying another hypothesis which was presumed to be the only logical alternative to the target hypothesis. Crucial experiments were thus thought to play a central role in science's project of searching for the truth.

In the selection presented here, Chapter VI of his book, Duhem argues against this view. He shows that two conditions must be satisfied if simultaneous falsification and verification are to take place, and that neither of these conditions can, as a matter of fact, be fulfilled. In the first place, an unambiguous falsification procedure must exist. *Modus tollens* arguments are usually taken to represent the appropriate falsification procedure[1], but Duhem argues that *modus tollens* is rarely, if ever, the structure of argument in the sciences since a scientist's predictions are in fact based not on any single hypothesis but, instead, on at least several assumptions and rules of inference, some of which are often only tacitly held. It is the target hypothesis plus a set of auxiliary hypotheses from which predictions are deduced. "The physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses". Thus there is no reason to single out any particular hypothesis as the guilty one for isolated hypotheses are immune from refutation: Duhem denies that unambiguous falsification procedures do exist in science.

Secondly, even if it were possible to refute a particular hypothesis, one would not be justified in presuming that one had thereby shown any alternative hypothesis to be true, or as the claim has more recently been stated, shown any alternative hypothesis to be closer to the truth. In order to make this stronger truth claim – or truth-like claim – one must be able to show that *reductio ad absurdum* methods are applicable to scientific inference. First of all, Duhem points out that if it were possible to falsify any single hypothesis, then it might prove possible in the future to falsify any hypothesis to date unrefuted. But furthermore, in the future it might also be the case that some alternative explanation more satisfactory than any now known might be produced or discovered. We can see that

this second problem arises because the concepts involved in our hypotheses change as knowledge grows; the plausibility of a description of some characteristic of nature is as much a consequence of the adequacy of one's concepts as it is due to the truth of one's claims. Thus, even if one could falsify a given hypothesis, the only truth established by such a falsification would be the denial of the hypothesis. But the denial of the hypothesis is not itself a single hypothesis but, given conceptual creativity, a potentially infinite disjunction of hypotheses. Because the physicist – unlike the Greek geometer – cannot enumerate all the possible alternative hypotheses which would explain an event, *reductio* methods are not applicable to scientific inference: we can't "assimilate experimental contradiction to reduction to absurdity", as Duhem says. So neither condition required for experiments to be crucial can, in fact, be satisfied, according to Duhem.

In his well-known essay, 'Two Dogmas of Empiricism', Quine refers approvingly to Duhem when he argues that only science as a whole, including the laws of logic, is empirically testable. Many have seen this as a radical conventionalist thesis. A great deal of critical attention has been focused on Quine's attack on the analytic/synthetic distinction; but that may well turn out to be the less important claim Quine makes in this essay. In the first four sections of the essay, Quine criticizes several arguments which might be given in defense of the analytic/synthetic distinction. He then goes on to consider a way of defending the distinction which relies on the verification theory of meaning. Perhaps a statement can be taken to be analytic if it is confirmed by anything whatever that happens in the world. Surely, if we can take some particular statements to be verified by particular experiences, we can also take other statements to be verified 'come what may'. But to this line of argument Quine objects that in fact no individual statement can be verified. Well, one might think, perhaps a statement is analytic if it is disconfirmed by nothing whatever that happens in the world. But, Quine notes, "any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system". And, by the same token, "no statement is immune to revision". He says, "the unit of empirical significance is the whole of science". If Quine could provide the philosophic underpinnings to defend this point, he would have shown not only that the analytic/synthetic distinction is untenable, but, more importantly, that the true/false distinction is not

defensible except as applied to science as a whole, and that we cannot defend on epistemological grounds the distinction between physics and logic. However, Quine does not really provide the philosophic underpinnings needed to support these broad claims.

Quine's thesis is stronger than Duhem's, for where Duhem claimed that the physicist can never be sure that no saving set of auxiliary assumptions exists which, together with the target hypothesis, would entail the actual observational results, Quine seems to hold that saving hypotheses always exist: "Any statement can be held true come what may". Quine's thesis is also more general than Duhem's, for Quine extends Duhem's claim for conventionalism in physics to include the truths of logic as well as *all* of the laws of science.

Starting from somewhat different problems, both Carl Hempel and Thomas S. Kuhn have arrived at conclusions similar to Duhem's and Quine's. Hempel began with the problem of defining theoretical terms. He argued that the positivists were wrong to think that the theoretical terms of science can be explicitly or operationally defined using only observation terms. Instead, the theoretical terms must be introduced into science by the theories themselves, and this means that the theoretical term is, in effect, implicitly defined not by observation terms but by the theory. However, because statements are deducible from the theory which do not contain the theoretical terms in question, the theory as a whole can be said to have empirical significance, and, Hempel thinks, can be confirmed or falsified. He reminds that

> It is not correct to speak, as is often done, of 'the experiential meaning' of a term or a sentence in isolation. ... A single sentence in a scientific theory does not, as a rule, entail any observation sentences; consequences asserting the occurrence of certain observable phenomena can be derived from it only by conjoining it with a set of other, subsidiary, hypotheses. Of the latter, some will usually be observation sentences, others will be previously accepted theoretical statements.[2]

Thus for Hempel, the unit of empirical significance – the unit which is tested – must be only the theory as a whole, where this is evidently taken to include all the possible statements of any kind required for the derivation of observation sentences. Hempel has in effect given a defense of the Duhem–Quine thesis for that part of science which includes theories.

Kuhn's project was to give an account of the nature of the scientific enterprise and the reasons for its special success – an account which would

fit the history and practice of science better than what he claims are the obviously inadequate accounts standardly found in science textbooks and in many discussions in the philosophy of science. Kuhn proposed a fundamental change in our perception and evaluation of familiar episodes from the history and practice of science. He suggested that the everyday practice of science – 'normal science' – consists for the most part of a very important kind of puzzle solving in which the nature of the puzzle and the terms set for its solution are not themselves regarded critically or subjected to test by scientists in any significant way. Normal science takes place within holistic 'paradigms' – extremely broad theoretical, metaphysical and methodological models of nature and of how to discover her secrets. In Chapter X of *The Structure of Scientific Revolutions*, the selection included here, Kuhn argues that the theories which are part of scientific paradigms are not refutable by observations at all. This is because, on Kuhn's view, sensory experience is not fixed and neutral and theories are not simply man-made interpretations of given data: "What occurs during a scientific revolution is not fully reducible to a reinterpretation of individual and stable data.... Paradigms are not corrigible by normal science at all". Instead, it seems that the paradigmatic theories in fact define what is to count as a relevant observation; they define the 'world' within which the scientist works. So, a theory is not refuted by experience but instead simply abandoned when there are a large number of problems with it and when a possibly more fruitful way of perceiving the world is at hand. Kuhn claims that his account of science directly challenges the Cartesian paradigm which has guided epistemology and accounts of science for three-hundred years.

The two strongest attacks on the Duhem–Quine thesis have come from Adolf Grünbaum explicitly, and, mainly implicitly, from Karl Popper. In his extraordinarily influential *The Logic of Scientific Discovery*, published in German in 1935, but translated into English only 24 years later, Popper begins by identifying what he takes to be the fundamental problems for philosophy of science and for epistemology. Virtually all problems in epistemology come down to two, Popper says: the problem of induction ('Hume's problem'), and the problem of finding an acceptable criterion of demarcation between science and metaphysics ('Kant's problem'). With respect to Hume's problem, Popper says that the inductivists have simply misconceived the nature of scientific method and of

human knowledge. Scientists do not try to prove universal statements true by demonstrating them to be reducible to true singular statements; what scientists do is deduce singular statements from a theory and compare these with the results of experiments. If the comparison is favorable, the theory has temporarily passed its test; if the conclusions of the deductions have been falsified, then their falsification also falsifies the theory from which they were deduced. Therefore, it is falsifiability, not verifiability, which is at the heart of scientific method. Popper claims to dissolve Hume's problem of 'inductivism' by substituting for it a theory of 'deductivism' – a theory of falsifiability.

And it is his theory of falsifiability which provides the demarcation line between science and metaphysics, Popper says, thereby solving Kant's problem. For the positivists at one time, both the universal laws of science and also metaphysical speculations were assigned the status of 'non-statements' – they were 'meaningless', unable to be reduced to elementary statements of experience. But for Popper, while the universal laws of science cannot be verified, they, unlike metaphysical speculations, can be falsified by experience. Thus a theory of the falsifiability of singular statements and of the universal laws from which they can be deduced lies at the heart of Popper's solutions to what he takes to be the two leading problems in epistemology.

In the sixth section of the first selection here – Chapter I of *The Logic of Scientific Discovery* – Popper considers the objection that no theoretical system is ever conclusively falsified for it can always be saved by the logically admissible procedures of "introducing *ad hoc* auxiliary hypotheses, by changing *ad hoc* a definition, or by simply refusing to acknowledge any falsifying experience whatever". Popper proposes that "the empirical method shall be characterized as a method that excludes precisely those ways of evading falsification which... are logically admissible". The method of science is precisely a method insuring that the singular statements deduced from theories can be falsified by experience, thereby transmitting their falsity back up to the theory from which they were deduced. Appended here to Chapter I of *The Logic of Scientific Discovery* are two brief selections from later pages of Popper's book which stress his reasons for arguing that empirical science should be characterized by its methods.

More recently, in the short excerpt from *Conjectures and Refutations*

which is reprinted here, Popper argues specifically that the Quine–Duhem thesis, "the holistic view of tests", both "does not create a serious difficulty for the fallibilist and falsificationist" and also that "on the other hand... the holistic argument goes much too far". Evidently he thinks it does not create a difficulty since while the falsificationist does indeed take for granted a vast amount of traditional knowledge,

> He does not accept this background knowledge; neither as established nor as fairly certain, nor yet as probable. He knows that even its tentative acceptance is risky, and stresses that every bit of it is open to criticism even though only in a piecemeal way. We can never be certain that we shall challenge the right bit; but since our quest is not for certainty, this does not matter.... Now it has to be admitted that we can often test only a large chunk of a theoretical system, and sometimes perhaps only the whole system, and that, in these cases, it is sheer guesswork which of its ingredients should be held responsible for any falsification.[3]

However, the holistic argument goes too far since it is possible in "quite a few cases to find which hypothesis is responsible for the refutation; or in other words, which part or group of hypotheses was necessary for the derivation of the refuted prediction". Does the fact that in some cases scientists reach intersubjective agreement, at least temporarily, as to which part of their theories to revise save Popper's falsificationism from the Duhem–Quine thesis? The latter would appear to challenge not this uncontroversial sociological fact but the notion that it is tests which determine *which* part of our web of hypotheses and beliefs should be counted as refuted. We must still ask how Popper has succeeded in deflecting the challenge posed to his falsificationism by "the holistic view of tests".

In several publications Adolf Grünbaum has challenged Duhem's thesis that the falsifiability of an isolated empirical hypothesis as an explanans is unavoidably inconclusive. He regards the Duhemian thesis as a conventionalist ploy to be found also in Einstein, Poincaré, and Quine. In his well-known 1960 essay reprinted in this collection, Grünbaum argues that the Duhem–Quine thesis is both a logical non-sequitur and, furthermore, false. First he argues that conclusive falsifying experiments are possible. To deny this; Grünbaum says, Duhem would have to prove on general logical grounds that for any empirical finding whatever (e.g., ~ 0), there is a set of non-trivial auxiliary hypotheses from which, together with the target hypothesis in question, the findings could be deduced. But this Duhem cannot guarantee. Thus Duhem made a logical

error in taking the conventionalist thesis to follow from the occasional inconclusiveness of supposedly crucial experiments. But, furthermore, Grünbaum argues that as a matter of fact crucial falsifying experiments have actually occurred in physics, and he discusses what he thinks is an example of such a case. In his response, Quine himself suggests that he finds the Duhem–Quine thesis as challenged by Grünbaum tenable only if taken trivially. Grünbaum has since discussed this thesis in several places, most recently in the second of his essays reprinted here.

Recently a number of philosophers have thought further about the Duhem–Quine thesis, in large part due to Grünbaum's arguments. Laurens Laudan, Carlo Giannoni, and Gary Wedeking all point out that there are two versions of the Duhem–Quine thesis. There is a stronger one held not by Duhem but probably by Quine; the weaker one, actually held by Duhem, is untouched by Grünbaum's attack.

Grünbaum presumes, Laudan points out, that the burden of proof is on the scientist who refuses to call a refuted hypothesis false to show that his hypothesis can be saved by some suitable auxiliary hypothesis. But Duhem did not make this strong claim but only the weaker one that those who deny the target hypothesis must show that there does not exist an auxiliary hypothesis which would make the target hypothesis compatible with the unforeseen experimental results. Unless such a proof is forthcoming, Laudan points out, "a scientist is logically justified in seeking some rapprochement between his hypothesis and the uncooperative data". Duhem does not make the purported logical blunder ascribed to him by Grünbaum. Laudan goes on to argue that Grünbaum's counter-example purporting to make out a case for conclusive falsification does not hold up either. Grünbaum says that in the history of science one can find cases where a set of hypotheses has been falsified. But Duhem would not disagree with that holistic view, but only with a claim that isolated hypotheses are falsifiable. Furthermore, Grünbaum assumes that the fact that in a particular case the auxiliary hypotheses are highly probable forces a scientist to relinquish the target hypothesis. But 'highly probable' is not 'known to be true'; "the demands of prudence do not carry logical weight", Laudan notes.

After criticizing the Grünbaum objections to the Duhemian thesis, Giannoni goes on to defend that balance of the Quinean thesis which is both the same as the original Duhemian thesis and which is not defended

by Hempel. Giannoni defends the Duhem–Quine thesis for the part of science depending on the use of operational definitions which refer to measuring instruments, and, in particular, where the measurement is derivative rather than fundamental. In the last part of his essay, Giannoni points to some broader implications of the Duhemian thesis for our conception of scientific knowledge. The Duhemian thesis is not an epistemological thesis regarding our knowledge of the world, he says, but a semantical thesis regarding the meaning of scientific words and of scientific language. But to say this is not to trivialize the thesis, he thinks, for the thesis is required by the very notion of scientific discovery. This and other considerations lead Giannoni to suggest that the question about the nature of Duhemian conventionalism in physics comes down to a choice between a realistic or a nominalistic approach to the symbols of physics. And this in turn leads him to conclude that the Duhem–Quine thesis shows that the distinction between descriptive simplicity and inductive simplicity is not a logical distinction but an ontological one. Where a realist searches for inductive simplicity, a nominalist searches for descriptive simplicity.

Wedeking argues that Popper as well as Grünbaum are examples of the "empirically-minded philosophers" who think that there are empirical theories which are 'falsifiable' in the sense that each such theory unambiguously separates out those basic statements which it prohibits and those which it permits. After discussing Grünbaum's position, Wedeking tries to formulate the Quinean version of the Duhem–Quine thesis in such a way that its validity can be tested. But, he can turn up only the pragmatic formulation that the new theoretical system in which the target hypothesis is held true in light of the unforeseen observations must be one which is 'adequate' to the facts of experience. Thus Wedeking doubts the truth of the Quinean thesis in any but its most trivial form. Does this mean, he asks, that Quine's theory is worthless? No, for it leads us to observe that those cases in which Quine's thesis holds only trivially and those in which we are confronted with a real and significant alternative as to which sentences we should reject as false and which we should retain as true "differ only in a matter of degree".

The next two selections are extraordinarily rich and complex essays which, along with Feyerabend's essay, set the Duhem–Quine thesis and its criticisms in the broadest philosophic context. Hesse shows how

Duhem introduces two important modifications into classical empiricism: a new theory of correspondence and a new theory of coherence. She argues that Quine has taken up both aspects of Duhem's new empiricism; and, after countering both Popper's and Grünbaum's objections to the Duhem–Quine thesis, she points to what she thinks are two residual and significant problems with Duhem's and Quine's 'network theory'. First, some theory of relative empirical confirmation must be given, and this must be done without identifying any statement of the system which expresses the evidence incorrigibly. Secondly, a way must be provided for analysing stability and change of meaning in the network. Hesse proposes solutions to these problems, and then marks five features which distinguish the new Duhem–Quine empiricism, as she has now developed it, from the older, classical empiricism. First, she points out how the difference between theoretical and observational aspects of science is not espistemological but pragmatic and causal. Second, on the Duhem–Quine theory, empirical applications of observation predicates are not incorrigible, and the empirical laws taken to hold between them are not infallible. Third, in order to avoid the meaning variance paradoxes, the majority of descriptive predicates must be intersubjectively stable. However, we cannot know which observation predicates will retain stability of meaning in subsequent theories, just as we cannot know ahead of time which observation statements will be retained as true in later theories. Fourth, she thinks that we must assume some prior principles of selection for picking well-confirmed theories, and some criteria for shifts of applicability of some observation predicates if we are to avoid complete arbitrariness in adoption of a new 'best' theory. Finally, while some aspects of the epistemological problem are systematically conflated with causal mechanisms in the new empiricism, this does not result in a fatal circularity for the network model. She concludes that while the new Duhem–Quine view of science can provide no guarantee that our knowledge of this world is firmly based, that demand of the old empiricists was evidently unreasonable anyway.

Lakatos thinks that Kuhn, in particular, faced with the problems with both the older empiricist views and with what he calls 'naive falsificationism', abandons efforts to give a rational explanation of the success of science. Kuhn settles, he says, for merely trying to explain changes in paradigms in terms of social psychology. Lakatos proposes, instead, to

follow Popper's lead and find a workable criterion of rational progress in the growth of knowledge; he presents a theory of 'progressive problem shifts'. On Lakatos' theory, the history of science is not the history of scientific theories but the history of series of theories – the history of 'research programmes'. A research programme consists of methodological rules: a negative heuristic tells us what paths of research to avoid, a positive heuristic tells us what paths to pursue. In particular, the negative heuristic forbids us to direct our *modus tollens* at the 'hard core' of a series of theories. The positive heuristic consists of a partially articulated set of suggestions or hints on how to develop the 'refutable variants' of the research programme, how to modify and sophisticate the refutable protective belt of auxiliary hypotheses. Lakatos claims that his demarcation between progressive and degenerating problem shifts is almost identical with Popper's celebrated demarcation criterion between science and metaphysics.

Like Popper, Lakatos does not think that the 'weak form' of the Duhem–Quine thesis conflicts with methodological falsificationism (though it does with 'naive falsificationism'). But a strong form, which he thinks Quine probably holds, is inconsistent with all forms of methodological falsificationism because, as he sees it, it excludes any rational selection rules among the alternative theories. Do Lakatos and Popper take the conventionalism of Duhem and Quine to involve a greater component of arbitrariness in the construction and selection of theories than would be espoused by either Duhem or Quine?

In the next section, Grünbaum in part modifies his original critique of the Duhem–Quine thesis. At least in some cases, he now holds, we can ascertain the falsity of a component hypothesis to all scientific intents and purposes, although we cannot falsify it beyond any and all possibility of subsequent rehabilitation. Grünbaum here replies to the discussions of his earlier position by Laudan, Giannoni, Hesse, Lakatos and others.

The essays above have shown how the Duhem–Quine thesis provides the cutting edge for an attack on the empiricist reliance on the authority of the senses and, more generally, for a criticism of Cartesian foundationalism, or authoritarianism, in epistemology. Many would take Feyerabend to be the most extreme anti-Cartesian today, and in the last essay in this collection, one can detect attacks on two forms of what might be called the 'methodological authoritarianism' which characterizes much of

epistemology since Descartes. In its historically extremely influential form, the 'naive inductivist' holds that there are indeed rational principles for the discovery of nature's regularities, and that these are identical with the rational principles for justifying the accepted theories of science at any particular time. The naive inductivist has the idea that there exist definite principles and rules of a quasi-logical sort for inducing theories from phenomena. But recent philosophers of science as different as Hempel and Popper have retreated to a second position: science definitely can be reconstructed as a rational enterprise, but there may well be no principles for a prospective methodology of science. However, Feyerabend holds that there are no rationally defensible rules for either a prospective methodology of science or for a retrospective assessment of science.[4] In the selection presented here, the last five sections of his long (110 pp) essay 'Against Method', Feyerabend presents four objections to the first two positions, and suggests that the persistent attempt of epistemologists to identify a rational 'authority' – a firm source of knowledge or a sure method for discriminating between reliable and unreliable hypotheses – is not entirely separable from authoritarianism in values.

Against the naive inductivist, Feyerabend argues that since the history of science reveals no rationally justifiable rules for a prospective methodology, it should be granted that there are no rationally justifiable rules for retrospective assessments of science either. Second, Feyerabend points out that any proposed reconstructions of science which seem to us to be reasonably rational actually end up being useless, for they only succeed in eliminating all of science as irrational. Third, Feyerabend repeats his well-known criticisms of the purported asymmetry between observation statements and theoretical statements. It is not the case that observations are irrefutable while theories are refutable. While we often do test our theories by experience, we equally often assess experience in light of new or newer theories, changing our assessment of experience accordingly. Thus, for Feyerabend, research is an interaction between new theories explicitly stated and older beliefs which lie deep in the observation language. Finally, Feyerabend presents his well-known arguments for the incommensurability of theories.

Throughout Feyerabend's writing, and here in particular, can be found a conscious and conscientious attempt to push philosophy of science into an ethical dimension. Feyerabend sees epistemological and ethical prefer-

ences as tightly linked. For one thing, someone, or a culture, which prefers or is led to prefer a secure, routine, entirely predictable way of life will naturally be led to accept an epistemology characterized by unchanging, 'objective', guaranteed facts and by a theory of how, using just the right methods, we can apprehend and collect them. Ways of life and theories of knowledge are mutually supporting, Feyerabend suggests. And, the choice between alternative theories of knowledge becomes a normative act, the conscious selection of a certain way of understanding the world.

What exactly are the ideals of rationality which Feyerabend is here attacking? Is he in fact appealing to other, hidden ideals of rationality in his attack on Cartesianism? Is he proposing that there is an ideal of rationality which lies wholly or in part outside the forces which have shaped the history of science, outside the forces which produce maximum 'progress' in science? What account can be given of *this* notion of rationality?

With this concluding essay, the consequences of the Duhem–Quine thesis have been extended not only to challenge Cartesian epistemology, but also to open the possibility of a reconsideration of the link between epistemology, on the one hand, and ethics and political theory, on the other hand. The Duhem–Quine thesis may well take its place in the history of ideas as signaling a radical change in our understanding of the nature of both human knowledge and human knowers.

SANDRA G. HARDING

NOTES

[1] *Modus tollens* is represented by the schema $[(H \rightarrow O) \cdot \sim O] \rightarrow \sim H$. If, from a given hypothesis, H, we predict a certain observation, O, then, if the prediction turns out to be false, $\sim O$, that would serve to refute the hypothesis from which the prediction was deduced, $\sim H$.

[2] Carl G. Hempel, 'Empiricist Criteria of Cognitive Significance: Problems and Changes', *below*, p. 75.

[3] Karl R. Popper, 'Truth, Rationality, and the Growth of Scientific Knowledge', *Conjectures and Refutations*, Routledge and Kegan Paul, London, 1963, p. 238–239, *below* p. 114.

[4] C. A. Hooker has also set Feyerabend's philosophy of science in this context. Cf. his interesting review of 'Against Method': 'Critical Notice of Minnesota Studies in the Philosophy of Science', Part II, *Canadian Journal of Philosophy* 1 (1972).

PIERRE DUHEM

PHYSICAL THEORY AND EXPERIMENT*

1. THE EXPERIMENTAL TESTING OF A THEORY DOES NOT HAVE THE SAME LOGICAL SIMPLICITY IN PHYSICS AS IN PHYSIOLOGY

The sole purpose of physical theory is to provide a representation and classification of experimental laws; the only test permitting us to judge a physical theory and pronounce it good or bad is the comparison between the consequences of this theory and the experimental laws it has to represent and classify. Now that we have minutely analyzed the characteristics of a physical experiment and of a physical law, we can establish the principles that should govern the comparison between experiment and theory; we can tell how we shall recognize whether a theory is confirmed or weakened by facts.

When many philosophers talk about experimental sciences, they think only of sciences still close to their origins, e.g., physiology or certain branches of chemistry where the experimenter reasons directly on the facts by a method which is only common sense brought to greater attentiveness but where mathematical theory has not yet introduced its symbolic representations. In such sciences the comparison between the deductions of a theory and the facts of experiment is subject to very simple rules. These rules were formulated in a particularly forceful manner by Claude Bernard, who would condense them into a single principle, as follows:

"The experimenter should suspect and stay away from fixed ideas, and always preserve his freedom of mind".

"The first condition that has to be fulfilled by a scientist who is devoted to the investigation of natural phenomena is to preserve a complete freedom of mind based on philosophical doubt."[1]

If a theory suggests experiments to be done, so much the better: "... we can follow our judgment and our thought, give free rein to our imagination provided that all our ideas are only pretexts for instituting new experiments that may furnish us probative facts or unexpected and fruitful ones".[2] Once the experiment is done and the results clearly established,

if a theory takes them over in order to generalize them, coordinate them, and draw from them new subjects for experiment, still so much the better: "... if one is imbued with the principles of experimental method, there is nothing to fear; for so long as the idea is a right one, it will go on being developed; when it is an erroneous idea, experiment is there to correct it".[3] But so long as the experiment lasts, the theory should remain waiting, under strict orders to stay outside the door of the laboratory; it should keep silent and leave the scientist without disturbing him while he faces the facts directly; the facts must be observed without a preconceived idea and gathered with the same scrupulous impartiality, whether they confirm or contradict the predictions of the theory. The report that the observer will give us of his experiment should be a faithful and scrupulously exact reproduction of the phenomena, and should not let us even guess what system the scientist places his confidence in or distrusts.

> Men who have an excessive faith in their theories or in their ideas are not only poorly disposed to make discoveries but they also make very poor observations. They necessarily observe with a preconceived idea and, when they have begun an experiment, they want to see in its results only a confirmation of their theory. Thus they distort observation and often neglect very important facts because they go counter to their goal. That is what made us say elsewhere that we must never do experiments in order to confirm our ideas but merely to check them.... But it quite naturally happens that those who believe too much in their own theories do not sufficiently believe in the theories of others. Then the dominant idea of these condemners of others is to find fault with the theories of the latter and to seek to contradict them. The setback for science remains the same. They are doing experiments only in order to destroy a theory instead of doing them in order to look for the truth. They also make poor observations because they take into the results of their experiments only what fits their purpose, by neglecting what is unrelated to it, and by very carefully avoiding whatever might go in the direction of the idea they wish to combat. Thus one is led by two parallel paths to the same result, that is to say, to falsifying science and the facts.
>
> The conclusion of all this is that it is necessary to obliterate one's opinion as well as that of others when faced with the decisions of the experiment; ... we must accept the results of experiment just as they present themselves with all that is unforeseen and accidental in them.[4]

Here, for example, is a physiologist who admits that the anterior roots of the spinal nerve contain the motor nerve-fibers and the posterior roots the sensory fibers. The theory he accepts leads him to imagine an experiment: if he cuts a certain anterior root, he ought to be suppressing the mobility of a certain part of the body without destroying its sensibility; after making the section of this root, when he observes the consequences

of his operation and when he makes a report of it, he must put aside all his ideas concerning the physiology of the spinal nerve; his report must be a raw description of the facts; he is not permitted to overlook or fail to mention any movement or quiver contrary to his predictions or to attribute it to some secondary cause unless some special experiment has given evidence of this cause; he must, if he does not wish to be accused of scientific bad faith, establish an absolute separation or watertight compartment between the consequences of his theoretical deductions and the establishing of the facts shown by his experiments.

Such a rule is not by any means easily followed; it requires of the scientist an absolute detachment from his own thought and a complete absence of animosity when confronted with the opinion of another person; neither vanity nor envy ought to be countenanced by him. As Bacon put it, he should never show eyes lustrous with human passions. Freedom of mind, which constitutes the sole principle of experimental method, according to Claude Bernard, does not depend merely on intellectual conditions, but also on moral conditions, making its practice rarer and more meritorious.

But if experimental method as just described is difficult to practice, the logical analysis of it is very simple. This is no longer the case when the theory to be subjected to test by the facts is not a theory of physiology but a theory of physics. In the latter case, in fact, it is impossible to leave outside the laboratory door the theory that we wish to test, for without theory it is impossible to regulate a single instrument or to interpret a single reading. We have seen that in the mind of the physicist there are contantly present two sorts of apparatus: one is the concrete apparatus in glass and metal, manipulated by him, the other is the schematic and abstract apparatus which theory substitutes for the concrete apparatus and on which the physicist does his reasoning. For these two ideas are indissolubly connected in his intelligence, and each necessarily calls on the other; the physicist can no sooner conceive the concrete apparatus without associating with it the idea of the schematic apparatus than a Frenchman can conceive an idea without associating it with the French word expressing it. This radical impossibility, preventing one from dissociating physical theories from the experimental procedures appropriate for testing these theories, complicates this test in a singular way, and obliges us to examine the logical meaning of it carefully.

Of course, the physicist is not the only one who appeals to theories at the very time he is experimenting or reporting the results of his experiments. The chemist and the physiologist when they make use of physical instruments, e.g., the thermometer, the manometer, the calorimeter, the galvanometer, and the saccharimeter, implicitly admit the accuracy of the theories justifying the use of these pieces of apparatus as well as of the theories giving meaning to the abstract ideas of temperature, pressure, quantity of heat, intensity of current, and polarized light, by means of which the concrete indications of these instruments are translated. But the theories used, as well as the instruments employed, belong to the domain of physics; by accepting with these instruments the theories without which their readings would be devoid of meaning, the chemist and the physiologist show their confidence in the physicist, whom they suppose to be infallible. The physicist, on the other hand, is obliged to trust his own theoretical ideas or those of his fellow-physicists. From the standpoint of logic, the difference is of little importance; for the physiologist and chemist as well as for the physicist, the statement of the result of an experiment implies, in general, an act of faith in a whole group of theories.

2. An experiment in physics can never condemn an isolated hypothesis but only a whole theoretical group

The physicist who carries out an experiment, or gives a report of one, implicitly recognizes the accuracy of a whole group of theories. Let us accept this principle and see what consequences we may deduce from it when we seek to estimate the role and logical import of a physical experiment.

In order to avoid any confusion we shall distinguish two sorts of experiments: experiments of *application*, which we shall first just mention, and experiments of *testing*, which will be our chief concern.

You are confronted with a problem in physics to be solved practically; in order to produce a certain effect you wish to make use of knowledge acquired by physicists; you wish to light an incandescent bulb; accepted theories indicate to you the means for solving the problem; but to make use of these means you have to secure certain information; you ought, I suppose, to determine the electromotive force of the battery of generators at your disposal; you measure this electromotive force: that is what I call an experiment of application. This experiment does not aim at discover-

ing whether accepted theories are accurate or not; it merely intends to draw on these theories. In order to carry it out, you make use of instruments that these same theories legitimize; there is nothing to shock logic in this procedure.

But experiments of application are not the only ones the physicist has to perform; only with their aid can science aid practice, but it is not through them that science creates and develops itself; besides experiments of application, we have experiments of testing.

A physicist disputes a certain law; he calls into doubt a certain theoretical point. How will he justify these doubts? How will he demonstrate the inaccuracy of the law? From the proposition under indictment he will derive the prediction of an experimental fact; he will bring into existence the conditions under which this fact should be produced; if the predicted fact is not produced, the proposition which served as the basis of the prediction will be irremediably condemned.

F. E. Neumann assumed that in a ray of polarized light the vibration is parallel to the plane of polarization, and many physicists have doubted this proposition. How did O. Wiener undertake to transform this doubt into a certainty in order to condemn Neumann's proposition? He deduced from this proposition the following consequence: If we cause a light beam reflected at 45° from a plate of glass to interfere with the incident beam polarized perpendicularly to the plane of incidence, there ought to appear alternately dark and light interference bands parallel to the reflecting surface; he brought about the conditions under which these bands should have been produced and showed that the predicted phenomenon did not appear, from which he concluded that Neumann's proposition is false, viz., that in a polarized ray of light the vibration is not parallel to the plane of polarization.

Such a mode of demonstration seems as convincing and as irrefutable as the proof by reduction to absurdity customary among mathematicians; moreover, this demonstration is copied from the reduction to absurdity, experimental contradiction playing the same role in one as logical contradiction plays in the other.

Indeed, the demonstrative value of experimental method is far from being so rigorous or absolute: the conditions under which it functions are much more complicated than is supposed in what we have just said; the evaluation of results is much more delicate and subject to caution.

A physicist decides to demonstrate the inaccuracy of a proposition; in order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced, in order to interpret the results of this experiment and establish that the predicted phenomenon is not produced, he does not confine himself to making use of the proposition in question; he makes use also of a whole group of theories accepted by him as beyond dispute. The prediction of the phenomenon, whose nonproduction is to cut off debate, does not derive from the proposition challenged if taken by itself, but from the proposition at issue joined to that whole group of theories; if the predicted phenomenon is not produced, not only is the proposition questioned at fault, but so is the whole theoretical scaffolding used by the physicist. The only thing the experiment teaches us is that among the propositions used to predict the phenomenon and to establish whether it would be produced, there is at least one error; but where this error lies is just what it does not tell us. The physicist may declare that this error is contained in exactly the proposition he wishes to refute, but is he sure it is not in another proposition? If he is, he accepts implicitly the accuracy of all the other propositions he has used, and the validity of his conclusion is as great as the validity of his confidence.

Let us take as an example the experiment imagined by Zenker and carried out by O. Wiener. In order to predict the formation of bands in certain circumstances and to show that these did not appear, Wiener did not make use merely of the famous proposition of F. E. Neumann, the proposition which he wished to refute; he did not merely admit that in a polarized ray vibrations are parallel to the plane of polarization; but he used, besides this, propositions, laws, and hypotheses constituting the optics commonly accepted: he admitted that light consists in simple periodic vibrations, that these vibrations are normal to the light ray, that at each point the mean kinetic energy of the vibratory motion is a measure of the intensity of light, that the more or less complete attack of the gelatine coating on a photographic plate indicates the various degrees of this intensity. By joining these propositions, and many others that would take too long to enumerate, to Neumann's proposition, Wiener was able to formulate a forecast and establish that the experiment belied it. If he attributed this solely to Neumann's proposition, if it alone bears the responsibility for the error this negative result has put in evidence, then

Wiener was taking all the other propositions he invoked as beyond doubt. But this assurance is not imposed as a matter of logical necessity; nothing stops us from taking Neumann's proposition as accurate and shifting the weight of the experimental contradiction to some other proposition of the commonly accepted optics; as H. Poincaré has shown, we can very easily rescue Neumann's hypothesis from the grip of Wiener's experiment on the condition that we abandon in exchange the hypothesis which takes the mean kinetic energy as the measure of the light intensity; we may, without being contradicted by the experiment, let the vibration be parallel to the plane of polarization, provided that we measure the light intensity by the mean potential energy of the medium deforming the vibratory motion.

These principles are so important that it will be useful to apply them to another example; again we choose an experiment regarded as one of the most decisive ones in optics.

We know that Newton conceived the emission theory for optical phenomena. The emission theory supposes light to be formed of extremely thin projectiles, thrown out with very great speed by the sun and other sources of light; these projectiles penetrate all transparent bodies; on account of the various parts of the media through which they move, they undergo attractions and repulsions; when the distance separating the acting particles is very small these actions are very powerful, and they vanish when the masses between which they act are appreciably far from each other. These essential hypotheses joined to several others, which we pass over without mention, lead to the formulation of a complete theory of reflection and refraction of light; in particular, they imply the following proposition: The index of refraction of light passing from one medium into another is equal to the velocity of the light projectile within the medium it penetrates, divided by the velocity of the same projectile in the medium it leaves behind.

This is the proposition that Arago chose in order to show that the theory of emission is in contradiction with the facts. From this proposition a second follows: Light travels faster in water than in air. Now Arago had indicated an appropriate procedure for comparing the velocity of light in air with the velocity of light in water; the procedure, it is true, was inapplicable, but Foucault modified the experiment in such a way that it could be carried out; he found that the light was propagated less rapidly

in water than in air. We may conclude from this, with Foucault, that the system of emission is incompatible with the facts.

I say the *system* of emission and not the *hypothesis* of emission; in fact, what the experiment declares stained with error is the whole group of propositions accepted by Newton, and after him by Laplace and Biot, that is, the whole theory from which we deduce the relation between the index of refraction and the velocity of light in various media. But in condemning this system as a whole by declaring it stained with error, the experiment does not tell us where the error lies. Is it in the fundamental hypothesis that light consists in projectiles thrown out with great speed by luminous bodies? Is it in some other assumption concerning the actions experienced by light corpuscles due to the media through which they move? We know nothing about that. It would be rash to believe, as Arago seems to have thought, that Foucault's experiment condemns once and for all the very hypothesis of emission, i.e., the assimilation of a ray of light to a swarm of projectiles. If physicists had attached some value to this task, they would undoubtedly have succeeded in founding on this assumption a system of optics that would agree with Foucault's experiment.

In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed.

We have gone a long way from the conception of the experimental method arbitrarily held by persons unfamiliar with its actual functioning. People generally think that each one of the hypotheses employed in physics can be taken in isolation, checked by experiment, and then, when many varied tests have established its validity, given a definitive place in the system of physics. In reality, this is not the case. Physics is not a machine which lets itself be taken apart; we cannot try each piece in isolation and, in order to adjust it, wait until its solidity has been carefully checked. Physical science is a system that must be taken as a whole; it is an organism in which one part cannot be made to function except when the parts that are most remote from it are called into play, some more so than others, but all to some degree. If something goes wrong, if some discomfort is felt in the functioning of the organism, the physicist

will have to ferret out through its effect on the entire system which organ needs to be remedied or modified without the possibility of isolating this organ and examining it apart. The watchmaker to whom you give a watch that has stopped separates all the wheelworks and examines them one by one until he finds the part that is defective or broken. The doctor to whom a patient appears cannot dissect him in order to establish his diagnosis; he has to guess the seat and cause of the ailment solely by inspecting disorders affecting the whole body. Now, the physicist concerned with remedying a limping theory resembles the doctor and not the watchmaker.

3. A "CRUCIAL EXPERIMENT" IS IMPOSSIBLE IN PHYSICS

Let us press this point further, for we are touching on one of the essential features of experimental method, as it is employed in physics.

Reduction to absurdity seems to be merely a means of refutation, but it may become a method of demonstration: in order to demonstrate the truth of a proposition it suffices to corner anyone who would admit the contradictory of the given proposition into admitting an absurd consequence. We know to what extent the Greek geometers drew heavily on this mode of demonstration.

Those who assimilate experimental contradiction to reduction to absurdity imagine that in physics we may use a line of argument similar to the one Euclid employed so frequently in geometry. Do you wish to obtain from a group of phenomena a theoretically certain and indisputable explanation? Enumerate all the hypotheses that can be made to account for this group of phenomena; then, by experimental contradiction eliminate all except one; the latter will no longer be a hypothesis, but will become a certainty.

Suppose, for instance, we are confronted with only two hypotheses. Seek experimental conditions such that one of the hypotheses forecasts the production of one phenomenon and the other the production of quite a different effect; bring these conditions into existence and observe what happens; depending on whether you observe the first or the second of the predicted phenomena, you will condemn the second or the first hypothesis; the hypothesis not condemned will be henceforth indisputable; debate will be cut off, and a new truth will be acquired by science. Such is

the experimental test that the author of the *Novum Organum* called the "*fact of the cross*, borrowing this expression from the crosses which at an intersection indicate the various roads".

We are confronted with two hypotheses concerning the nature of light; for Newton, Laplace, or Biot light consisted of projectiles hurled with extreme speed, but for Huygens, Young, or Fresnel light consisted of vibrations whose waves are propagated within an ether. These are the only two possible hypotheses as far as one can see: either the motion is carried away by the body it excites and remains attached to it, or else it passes from one body to another. Let us pursue the first hypothesis; it declares that light travels more quickly in water than in air; but if we follow the second, it declares that light travels more quickly in air than in water. Let us set up Foucault's apparatus; we set into motion the turning mirror; we see two luminous spots formed before us, one colorless, the other greenish. If the greenish band is to the left of the colorless one, it means that light travels faster in water than in air, and that the hypothesis of vibrating waves is false. If, on the contrary, the greenish band is to the right of the colorless one, that means that light travels faster in air than in water, and that the hypothesis of emissions is condemned. We look through the magnifying glass used to examine the two luminous spots, and we notice that the greenish spot is to the right of the colorless one; the debate is over; light is not a body, but a vibratory wave motion propagated by the ether; the emission hypothesis has had its day; the wave hypothesis has been put beyond doubt, and the crucial experiment has made it a new article of the scientific credo.

What we have said in the foregoing paragraph shows how mistaken we should be to attribute to Foucault's experiment so simple a meaning and so decisive an importance; for it is not between two hypotheses, the emission and wave hypotheses, that Foucault's experiment judges trenchantly; it decides rather between two sets of theories each of which has to be taken as a whole, i.e., between two entire systems, Newton's optics and Huygens' optics.

But let us admit for a moment that in each of these systems everything is compelled to be necessary by strict logic, except a single hypothesis; consequently, let us admit that the facts, in condemning one of the two systems, condemn once and for all the single doubtful assumption it contains. Does it follow that we can find in the 'crucial experiment' an irre-

futable procedure for transforming one of the two hypotheses before us into a demonstrated truth? Between two contradictory theorems of geometry there is no room for a third judgment; if one is false, the other is necessarily true. Do two hypotheses in physics ever constitute such a strict dilemma? Shall we ever dare to assert that no other hypothesis is imaginable? Light may be a swarm of projectiles, or it may be a vibratory motion whose waves are propagated in a medium; is it forbidden to be anything else at all? Arago undoubtedly thought so when he formulated this incisive alternative: Does light move more quickly in water than in air? "Light is a body. If the contrary is the case, then light is a wave". But it would be difficult for us to take such a decisive stand; Maxwell, in fact, showed that we might just as well attribute light to a periodical electrical disturbance that is propagated within a dielectric medium.

Unlike the reduction to absurdity employed by geometers, experimental contradiction does not have the power to transform a physical hypothesis into an indisputable truth; in order to confer this power on it, it would be necessary to enumerate completely the various hypotheses which may cover a determinate group of phenomena; but the physicist is never sure he has exhausted all the imaginable assumptions. The truth of a physical theory is not decided by heads or tails.

4. Criticism of the Newtonian Method. First Example: Celestial Mechanics

It is illusory to seek to construct by means of experimental contradiction a line of argument in imitation of the reduction to absurdity; but the geometer is acquainted with other methods for attaining certainty than the method of reducing to an absurdity; the direct demonstration in which the truth of a proposition is established by itself and not by the refutation of the contradictory proposition seems to him the most perfect of arguments. Perhaps physical theory would be more fortunate in its attempts if it sought to imitate direct demonstration. The hypotheses from which it starts and develops its conclusions would then be tested one by one; none would have to be accepted until it presented all the certainty that experimental method can confer on an abstract and general proposition; that is to say, each would necessarily be either a law drawn from observation by the sole use of those two intellectual operations called

induction and generalization, or else a corollary mathematically deduced from such laws. A theory based on such hypotheses would then not present anything arbitrary or doubtful; it would deserve all the confidence merited by the faculties which serve us in formulating natural laws.

It was this sort of physical theory that Newton had in mind when, in the 'General Scholium' which crowns his *Principia*, he rejected so vigorously as outside of natural philosophy any hypothesis that induction did not extract from experiment; when he asserted that in a sound physics every proposition should be drawn from phenomena and generalized by induction.

The ideal method we have just described therefore deserves to be named the Newtonian method. Besides, did not Newton follow this method when he established the system of universal attraction, thus adding to his precepts the most magnificent of examples? Is not his theory of gravitation derived entirely from the laws which were revealed to Kepler by observation, laws which problematic reasoning transforms and whose consequences induction generalizes?

This first law of Kepler's, "The radial vector from the sun to a planet sweeps out an area proportional to the time during which the planet's motion is observed", did, in fact, teach Newton that each planet is constantly subjected to a force directed toward the sun.

The second law of Kepler's, "The orbit of each planet is an ellipse having the sun at one focus", taught him that the force attracting a given planet varies with the distance of this planet from the sun, and that it is in an inverse ratio to the square of this distance.

The third law of Kepler's, "The squares of the periods of revolution of the various planets are proportional to the cubes of the major axes of their orbits", showed him that different planets would, if they were brought to the same distance from the sun, undergo in relation to it attractions proportional to their respective masses.

The experimental laws established by Kepler and transformed by geometric reasoning yield all the characteristics present in the action exerted by the sun on a planet; by induction Newton generalized the result obtained; he allowed this result to express the law according to which any portion of matter acts on any other portion whatsoever, and he formulated this great principle: "Any two bodies whatsoever attract each other with

a force which is proportional to the product of their masses and in inverse ratio to the square of the distance between them". The principle of universal gravitation was found, and it was obtained, without any use having been made of any fictive hypothesis, by the inductive method the plan of which Newton outlined.

Let us again examine this application of the Newtonian method, this time more closely; let us see if a somewhat strict logical analysis will leave intact the appearance of rigor and simplicity that this very summary exposition attributes to it.

In order to assure this discussion of all the clarity it needs, let us begin by recalling the following principle, familiar to all those who deal with mechanics: We cannot speak of the force which attracts a body in given circumstances before we have designated the supposedly fixed term of reference to which we relate the motion of all bodies; when we change this point of reference or term of comparison, the force representing the effect produced on the observed body by the other bodies surrounding it changes in direction and magnitude according to the rules stated by mechanics with precision.

That posited, let us follow Newton's reasoning.

Newton first took the sun as the fixed point of reference; he considered the motions affecting the different planets by reference to the sun; he admitted Kepler's laws as governing these motions, and derived the following proposition: If the sun is the point of reference in relation to which all forces are compared, each planet is subjected to a force directed toward the sun, a force proportional to the mass of the planet and to the inverse square of its distance from the sun. Since the latter is taken as the reference point, it is not subject to any force.

In an analogous manner Newton studied the motion of the satellites and for each of these he chose as a fixed reference point the planet which the satellite accompanies, the earth in the case of the moon, Jupiter in the case of the masses moving around Jupiter. Laws just like Kepler's were taken as governing these motions, from which it follows that we can formulate the following proposition: If we take as a fixed reference point the planet accompanied by a satellite, this satellite is subject to a force directed toward the planet varying inversely with the square of the distance. If, as happens with Jupiter, the same planet possesses several satellites, these satellites, were they at the same distance from the planet, would be acted

on by the latter with forces proportional to their respective masses. The planet is itself not acted on by the satellite.

Such, in very precise form, are the propositions which Kepler's laws of planetary motion and the extension of these laws to the motions of satellites authorize us to formulate. For these propositions Newton substituted another which may be stated as follows: Any two celestial bodies whatsoever exert on each other a force of attraction in the direction of the straight line joining them, a force proportional to the product of their masses and to the inverse square of the distance between them. This statement presupposes all motions and forces to be related to the same reference point; the latter is an ideal standard of reference which may well be conceived by the geometer but which does not characterize in an exact and concrete manner the position in the sky of any body.

Is this principle of universal gravitation merely a generalization of the two statements provided by Kepler's laws and their extension to the motion of satellites? Can induction derive it from these two statements? Not at all. In fact, not only is it more general than these two statements and unlike them, but it contradicts them. The student of mechanics who accepts the principle of universal attraction can calculate the magnitude and direction of the forces between the various planets and the sun when the latter is taken as the reference point, and if he does he finds that these forces are not what our first statement would require. He can determine the magnitude and direction of each of the forces between Jupiter and its satellites when we refer all the motions to the planet, assumed to be fixed, and if he does he notices that these forces are not what our second statement would require.

The principle of universal gravity, very far from being derivable by generalization and induction from the observational laws of Kepler, formally contradicts these laws. If Newton's theory is correct, Kepler's laws are necessarily false.

Kepler's laws based on the observation of celestial motions do not transfer their immediate experimental certainty to the principle of universal weight, since if, on the contrary, we admit the absolute exactness of Kepler's laws, we are compelled to reject the proposition on which Newton based his celestial mechanics. Far from adhering to Kepler's laws, the physicist who claims to justify the theory of universal gravitation finds that he has, first of all, to resolve a difficulty in these laws: he has to

prove that his theory, incompatible with the exactness of Kepler's laws, subjects the motions of the planets and satellites to other laws scarcely different enough from the first laws for Tycho Brahé, Kepler, and their contemporaries to have been able to discern the deviations between the Keplerian and Newtonian orbits. This proof derives from the circumstances that the sun's mass is very large in relation to the masses of the various planets and the mass of a planet is very large in relation to the masses of its satellites.

Therefore, if the certainty of Newton's theory does not emanate from the certainty of Kepler's laws, how will this theory prove its validity? It will calculate, with all the high degree of approximation that the constantly perfected methods of algebra involve, the perturbations which at each instant remove every heavenly body from the orbit assigned to it by Kepler's laws; then it will compare the calculated perturbations with the perturbations observed by means of the most precise instruments and the most scrupulous methods. Such a comparison will not only bear on this or that part of the Newtonian principle, but will involve all its parts at the same time; with those it will also involve all the principles of dynamics; besides, it will call in the aid of all the propositions of optics, the statics of gases, and the theory of heat, which are necessary to justify the properties of telescopes in their construction, regulation, and correction, and in the elimination of the errors caused by diurnal or annual aberration and by atmospheric refraction. It is no longer a matter of taking, one by one, laws justified by observation, and raising each of them by induction and generalization to the rank of a principle; it is a matter of comparing the corollaries of a whole group of hypotheses to a whole group of facts.

Now, if we seek out the causes which have made the Newtonian method fail in this case for which it was imagined and which seemed to be the most perfect application for it, we shall find them in that double character of any law made use of by theoretical physics: This law is symbolic and approximate.

Undoubtedly, Kepler's laws bear quite directly on the very objects of astronomical observation; they are as little symbolic as possible. But in this purely experimental form they remain inappropriate for suggesting the principle of universal gravitation; in order to acquire this fecundity they must be transformed and must yield the characters of the forces by which the sun attracts the various planets.

Now this new form of Kepler's laws is a symbolic form; only dynamics gives meanings to the words 'force' and 'mass', which serve to state it, and only dynamics permits us to substitute the new symbolic formulas for the old realistic formulas, to substitute statements relative to 'forces' and 'masses' for laws relative to orbits. The legitimacy of such a substitution implies full confidence in the laws of dynamics.

And in order to justify this confidence let us not proceed to claim that the laws of dynamics were beyond doubt at the time Newton made use of them in symbolically translating Kepler's laws; that they had received enough empirical confirmation to warrant the support of reason. In fact, the laws of dynamics had been subjected up to that time to only very limited and very crude tests. Even their enunciations had remained very vague and involved; only in Newton's *Principia* had they been for the first time formulated in a precise manner. It was in the agreement of the facts with the celestial mechanics which Newton's labors gave birth to that they received their first convincing verification.

Thus the translation of Kepler's laws into symbolic laws, the only kind useful for a theory, presupposed the prior adherence of the physicist to a whole group of hypotheses. But, in addition, Kepler's laws being only approximate laws, dynamics permitted giving them an infinity of different symbolic translations. Among these various forms, infinite in number, there is one and only one which agrees with Newton's principle. The observations of Tycho Brahé, so felicitously reduced to laws by Kepler, permit the theorist to choose this form, but they do not constrain him to do so, for there is an infinity of others they permit him to choose.

The theorist cannot, therefore, be content to invoke Kepler's laws in order to justify his choice. If he wishes to prove that the principle he has adopted is truly a principle of natural classification for celestial motions, he must show that the observed perturbations are in agreement with those which had been calculated in advance; he has to show how from the course of Uranus he can deduce the existence and position of a new planet, and find Neptune in an assigned direction at the end of his telescope.

5. Criticism of the Newtonian Method (Continued). Second Example: Electrodynamics

Nobody after Newton except Ampère has more clearly declared that all

physical theory should be derived from experience by induction only; no work has been more closely modelled after Newton's *Philosophiae naturalis Principia mathematica* than Ampère's *Théorie mathématique des phénomènes électrodynamiques uniquement déduite de l'expérience.*

The epoch marked by the works of Newton in the history of the sciences is not only one of the most important discoveries that man has made concerning the causes of the great phenomena of nature, but it is also the epoch in which the human mind opened a new route in the sciences whose object is the study of these phenomena.

These are the lines with which Ampère began the exposition of his *Théorie mathématique*; he continued in the following terms:

"Newton was far from thinking" that the law of universal weight

could be discovered by starting from more or less plausible abstract considerations. He established the fact that it had to be deduced from observed facts, or rather from those empirical laws which, like those of Kepler, are but results generalized from a great number of facts.

To observe the facts first, to vary their circumstances as far as possible, to make precise measurements along with this first task in order to deduce from them general laws based only on experience, and to deduce from these laws, independently of any hypothesis about the nature of the forces producing the phenomena, the mathematical value of these forces, i.e., the formula representing them – that is the course Newton followed. It has been generally adopted in France by the scientists to whom physics owes the enormous progress it has made in recent times, and it has served me as a guide in all my research on electrodynamic phenomena. I have consulted only experience in order to establish the laws of these phenomena, and I have deduced from them the formula which can only represent the forces to which they are due; I have made no investigation about the cause itself assignable to these forces, well convinced that any investigation of this kind should be preceded simply by experimental knowledge of the laws and of the determination, deduced solely from these laws, of the value of the elementary force.

Neither very close scrutiny nor great perspicacity is needed in order to recognize that the *Théorie mathématique des phénomènes électrodynamiques* does not in any way proceed according to the method prescribed by Ampère and to see that it is not "deduced only from experience" (*uniquement déduite de l'expérience*). The facts of experience taken in their primitive rawness cannot serve mathematical reasoning; in order to feed this reasoning they have to be transformed and put into a symbolic form. This transformation Ampère did make them undergo. He was not content merely with reducing the metal apparatus in which currents flow to simple geometric figures; such an assimilation imposes itself too naturally

to give way to any serious doubt. Neither was he content merely to use the notion of force, borrowed from mechanics, and various theorems constituting this science; at the time he wrote, these theorems might be considered as beyond dispute. Besides all this, he appealed to a whole set of entirely new hypotheses which are entirely gratuitous and sometimes even rather surprising. Foremost among these hypotheses it is appropriate to mention the intellectual operation by which he decomposed into infinitely small elements the electric current, which, in reality, cannot be broken without ceasing to exist; then the supposition that all real electrodynamic actions are resolved into fictive actions involving the pairs that the elements of current form, one pair at a time; then the postulate that the mutual actions of two elements are reduced to two forces applied to the elements in the direction of the straight line joining them, forces equal and opposite in direction; then the postulate that the distance between two elements enters simply into the formula of their mutual action by the inverse of a certain power.

These diverse assumptions are so little self-evident and so little necessary that several of them have been criticized or rejected by Ampère's successors; other hypotheses equally capable of translating symbolically the fundamental experiments of electrodynamics have been proposed by other physicists, but none of them has succeeded in giving this translation without formulating some new postulate, and it would be absurd to claim to do so.

The necessity which leads the physicist to translate experimental facts symbolically before introducing them into his reasoning, renders the purely inductive path Ampère drew impracticable; this path is also forbidden to him because each of the observed laws is not exact but merely approximate.

Ampère's experiments have the grossest degree of approximation. He gave a symbolic translation of the facts observed in a form appropriate for the success of his theory, but how easily he might have taken advantage of the uncertainty of the observations in order to give quite a different translation! Let us listen to Wilhelm Weber:

Ampère made a point of expressly indicating in the title of his memoir that his mathematical theory of electrodynamic phenomena is *deduced only from experiment*, and indeed in his book we find expounded in detail the simple as well as ingenious method which led him to his goal. There we find, presented with all the precision and scope

desirable, the exposition of his experiments, the deductions that he draws from them for theory, and the description of the instruments he employs. But in fundamental experiments, such as we have here, it is not enough to indicate the general meaning of an experiment, to describe the instruments used in performing it, and to tell in a general way that it has yielded the result expected; it is indispensable to go into the details of the experiment itself, to say how often it has been repeated, how the conditions were modified, and what the effect of these modifications has been; in a word, to compose a sort of brief of all the circumstances permitting the reader to sit in judgment on the degree of reliability and certainty of the result. Ampère does *not* give these precise details concerning his experiments, and the demonstration of the fundamental law of electrodynamics still awaits this indispensable supplementation. The fact of the mutual attraction of two conducting wires has been verified over an over again and is beyond all dispute; but these verifications have always been made under conditions and by such means that no *quantitative* measurement was possible and these measurements are far from having reached the degree of precision required for considering the law of these phenomena demonstrated.

More than once, Ampère has drawn from the *absence* of any electrodynamic action the same consequences as from a measurement that would have given him a result equal to *zero*, and by this artifice, with great sagacity and with even greater skill, he has succeeded in bringing together the data necessary for the establishment and demonstration of his theory; but these *negative* experiments with which we must be content in the absence of direct *positive* measurements,

those experiments in which all passive resistances, all friction, all causes of error tend precisely to produce the effect we wish to observe,

cannot have all the value or demonstrative force of those positive measurements, especially when they are not obtained with the procedures and under the conditions of true measurement, which are moreover impossible to obtain with the instruments Ampère has employed.[5]

Experiments with so little precision leave the physicist with the problem of choosing between an infinity of equally possible symbolic translations, and confer no certainty on a choice they do not impose; only intuition, guessing the form of theory to be established, directs this choice. This role of intuition is particularly important in the work of Ampère; it suffices to run through the writings of this great geometer in order to recognize that his fundamental formula of electrodynamics was found quite completely by a sort of divination, that his experiments were thought up by him as afterthoughts and quite purposefully combined so that he might be able to expound according to the Newtonian method a theory that he had constructed by a series of postulates.

Besides, Ampère had too much candor to dissimulate very learnedly that what was artificial in his exposition was *entirely deduced from experi-*

ment; at the end of his *Théorie mathématique des phénomènes électrodynamiques* he wrote the following lines: "I think I ought to remark in finishing this memoir that I have not yet had the time to construct the instruments represented in Diagram 4 of the first plate and in Diagram 20 of the second plate. The experiments for which they were intended have not yet been done." Now the first of the two sets of apparatus in question aimed to bring into existence the last of the four fundamental cases of equilibrium which are like columns in the edifice constructed by Ampère: it is with the aid of the experiment for which this apparatus was intended that we were to determine the power of the distance according to which electrodynamic actions proceed. Very far from its being the case that Ampère's electrodynamic theory was *entirely deduced from experiment*, experiment played a very feeble role in its formation: it was merely the occasion which awakened the intuition of this physicist of genius, and his intuition did the rest.

It was through the research of Wilhelm Weber that the very intuitive theory of Ampère was first subjected to a detailed comparison with the facts; but this comparison was not guided by the Newtonian method. Weber deduced from Ampère's theory, taken as a whole, certain effects capable of being calculated; the theorems of statics and of dynamics, and also even certain propositions of optics, permitted him to conceive an apparatus, the electrodynamometer, by means of which these same effects may be subjected to precise measurements; the agreement of the calculated predictions with the results of the measurements no longer, then, confirms this or that isolated proposition of Ampère's theory, but the whole set of electrodynamical, mechanical, and optical hypotheses that must be invoked in order to interpret each of Weber's experiments.

Hence, where Newton had failed, Ampère in his turn just stumbled. That is because two inevitable rocky reefs make the purely inductive course impracticable for the physicist. In the first place, no experimental law can serve the theorist before it has undergone an interpretation transforming it into a symbolic law; and this interpretation implies adherence to a whole set of theories. In the second place, no experimental law is exact but only approximate, and is therefore susceptible to an infinity of distinct symbolic translations; and among all these translations the physicist has to choose one which will provide him with a fruitful hypothesis without his choice being guided by experiment at all.

This criticism of the Newtonian method brings us back to the conclusions to which we have already been led by the criticism of experimental contradiction and of the crucial experiment. These conclusions merit our formulating them with the utmost clarity. Here they are:

To seek to separate each of the hypotheses of theoretical physics from the other assumptions on which this science rests in order to subject it in isolation to observational test is to pursue a chimera; for the realization and interpretation of no matter what experiment in physics imply adherence to a whole set of theoretical propositions.

The only experimental check on a physical theory which is not illogical consists in comparing the *entire system of the physical theory with the whole group of experimental laws*, and in judging whether the latter is represented by the former in a satisfactory manner.

6. Consequences relative to the teaching of physics

Contrary to what we have made every effort to establish, it is generally accepted that each hypothesis of physics may be separated from the group and subjected in isolation to experimental test. Of course, from this erroneous principle false consequences are deduced concerning the method by which physics should be taught. People would like the professor to arrange all the hypotheses of physics in a certain order, to take the first one, enounce it, expound its experimental verifications, and then when the latter have been recognized as sufficient, declare the hypothesis accepted. Better still, people would like him to formulate this first hypothesis by inductive generalization of a purely experimental law; he would begin this operation again on the second hypothesis, on the third, and so on until all of physics was constituted. Physics would be taught as geometry is: hypotheses would follow one another as theorems follow one another; the experimental test of each assumption would replace the demonstration of each proposition; nothing which is not drawn from facts or immediately justified by facts would be promulgated.

Such is the ideal which has been proposed by many teachers, and which several perhaps think they have attained. There is no lack of authoritative voices inviting them to the pursuit of this ideal. M. Poincaré says:

It is important not to multiply hypotheses excessively, but to make them only one after the other. If we construct a theory based on multiple hypotheses, and experiment

condemns the theory, which one among our premises is it necessary to change? It will be impossible to know. And if, on the other hand, the experiment succeeds, shall we think we have verified all these hypotheses at the same time? Shall we think we have determined several unknowns with a single equation?[6]

In particular, the purely inductive method whose laws Newton formulated is given by many physicists as the only method permitting one to expound rationally the science of nature. Gustave Robin says:

The science we shall make will be only a combination of simple inductions suggested by experience. As to these inductions, we shall formulate them always in propositions easy to retain and *susceptible of direct verification*, never losing sight of the fact that *a hypothesis cannot be verified by its consequences.*[7]

This is the Newtonian method recommended if not prescribed for those who plan to teach physics in the secondary schools. They are told:

The procedures of mathematical physics are not adequate for secondary-school instruction, for they consist in starting from hypotheses or from definitions posited a priori in order to deduce from them conclusions which will be subjected to experimental check. This method may be suitable for specialized classes in mathematics, but it is wrong to apply it at present in our elementary courses in mechanics, hydrostatics, and optics. Let us replace it by the inductive method.[8]

The arguments we have developed have established more than sufficiently the following truth: It is as impracticable for the physicist to follow the inductive method whose practice is recommended to him as it is for the mathematician to follow that perfect deductive method which would consist in defining and demonstrating everything, a method of inquiry to which certain geometers seem passionately attached, although Pascal properly and rigorously disposed of it a long time ago. Therefore, it is clear that those who claim to unfold the series of physical principles by means of this method are naturally giving an exposition of it that is faulty at some point.

Among the vulnerable points noticeable in such an exposition, the most frequent and, at the same time, the most serious, because of the false ideas it deposits in the minds of students, is the 'fictitious experiment'. Obliged to invoke a principle which has not really been drawn from facts or obtained by induction, and averse, moreover, to offering this principle for what it is, namely, a postulate, the physicist invents an imaginary experiment which, were it carried out with success, would possibly lead to the principle whose justification is desired.

To invoke such a fictitious experiment is to offer an experiment to be done for an experiment done; this is justifying a principle not by means of facts observed but by means of facts whose existence is predicted, and this prediction has no other foundation than the belief in the principle supported by the alleged experiment. Such a method of demonstration implicates him who trusts it in a vicious circle; and he who teaches it without making it exactly clear that the experiment cited has not been done commits an act of bad faith.

At times the fictitious experiment described by the physicist could not, if we attempted to bring it about, yield a result of any precision; the very indecisive and rough results it would produce could undoubtedly be put into agreement with the proposition claimed to be warranted; but they would agree just as well with certain very different propositions; the demonstrative value of such an experiment would therefore be very weak and subject to caution. The experiment that Ampère imagined in order to prove that electrodynamic actions proceed according to the inverse square of the distance, but which he did not perform, gives us a striking example of such a fictitious experiment.

But there are worse things. Very often the fictitious experiment invoked is not only not realized but incapable of being realized; it presupposes the existence of bodies not encountered in nature and of physical properties which have never been observed. Thus Gustave Robin, in order to give the principles of chemical mechanics the purely inductive exposition that he wishes, creates at will what he calls witnessing bodies (*corps témoins*), bodies which by their presence alone are capable of agitating or stopping a chemical reaction.[9] Observation has never revealed such bodies to chemists.

The unperformed experiment, the experiment which would not be performed with precision, and the absolutely unperformable experiment do not exhaust the diverse forms assumed by the fictitious experiment in the writings of physicists who claim to be following the experimental method; there remains to be pointed out a form more illogical than all the others, namely, the absurd experiment. The latter claims to prove a proposition which is contradictory if regarded as the statement of an experimental fact.

The most subtle physicists have not always known how to guard against the intervention of the absurd experiment in their expositions. Let us quote, for instance, some lines taken from J. Bertrand:

If we accept it as an experimental fact that electricity is carried to the surface of bodies, and as a necessary principle that the action of free electricity on the points of conductors should be null, we can deduce from these two conditions, supposing they are strictly satisfied, that electrical attractions and repulsions are inversely proportional to the square of the distance.[10]

Let us take the proposition "There is no electricity in the interior of a conducting body when electrical equilibrium is established in it", and let us inquire whether it is possible to regard it as the statement of an experimental fact. Let us weigh the exact sense of the words figuring in the statement, and particularly, of the word interior. In the sense we must give this word in this proposition, a point interior to a piece of electrified copper is a point taken within the mass of copper. Consequently, how can we go about establishing whether there is or is not any electricity at this point? It would be necessary to place a testing body there, and to do that it would be necessary to take away beforehand the copper that is there, but then this point would no longer be within the mass of copper; it would be outside that mass. We cannot without falling into a logical contradiction take our proposition as a result of observation.

What, therefore, is the meaning of the experiments by which we claim to prove this proposition? Certainly, something quite different from what we make them say. We hollow out a cavity in a conducting mass and note that the walls of this cavity are not charged. This observation proves nothing concerning the presence or absence of electricity at points deep within the conducting mass. In order to pass from the experimental law noted to the law stated we play on the word interior. Afraid to base electrostatics on a postulate, we base it on a pun.

If we simply turn the pages of the treatises and manuals of physics we can collect any number of fictitious experiments; we should find there abundant illustrations of the various forms that such an experiment can assume, from the merely unperformed experiment to the absurd experiment. Let us not waste time on such a fastidious task. What we have said suffices to warrant the following conclusion: The teaching of physics by the purely inductive method such as Newton defined it is a chimera. Whoever claims to grasp this mirage is deluding himself and deluding his pupils. He is giving them, as facts seen, facts merely foreseen; as precise observations, rough reports; as performable procedures, merely ideal experiments; as experimental laws, propositions whose terms cannot be

taken as real without contradiction. The physics he expounds is false and falsified.

Let the teacher of physics give up this ideal inductive method which proceeds from a false idea, and reject this way of conceiving the teaching of experimental science, a way which dissimulates and twists its essential character. If the interpretation of the slightest experiment in physics presupposes the use of a whole set of theories, and if the very description of this experiment requires a great many abstract symbolic expressions whose meaning and correspondence with the facts are indicated only by theories, it will indeed be necessary for the physicist to decide to develop a long chain of hypotheses and deductions before trying the slightest comparison between the theoretical structure and the concrete reality; also, in describing experiments verifying theories already developed, he will very often have to anticipate theories to come. For example, he will not be able to attempt the slightest experimental verification of the principles of dynamics before he has not only developed the chain of propositions of general mechanics but also laid the foundations of celestial mechanics; and he will also have to suppose as known, in reporting the observations verifying this set of theories, the laws of optics which alone warrant the use of astronomical instruments.

Let the teacher therefore develop, in the first place, the essential theories of the science; without doubt, by presenting the hypotheses on which these theories rest, it is necessary for him to prepare their acceptance; it is good for him to point out the data of common sense, the facts gathered by ordinary observation or simple experiments or those scarcely analyzed which have led to formulating these hypotheses. To this point, moreover, we shall insist on returning in the next chapter; but we must proclaim loudly that these facts sufficient for suggesting hypotheses are not sufficient to verify them; it is only after he has constituted an extensive body of doctrine and constructed a complete theory that he will be able to compare the consequences of this theory with experiment.

Instruction ought to get the student to grasp this primary truth: Experimental verifications are not the base of theory but its crown. Physics does not make progress in the way geometry does: the latter grows by the continual contribution of a new theorem demonstrated once and for all and added to theorems already demonstrated; the former is a symbolic painting in which continual retouching gives greater comprehensiveness and

unity, and the *whole* of which gives a picture resembling more and more the *whole* of the experimental facts, whereas each detail of the picture cut off and isolated from the whole loses all meaning and no longer represents anything.

To the student who will not have perceived this truth, physics will appear as a monstrous confusion of fallacies of reasoning in circles and begging the question; if he is endowed with a mind of high accuracy, he will repel with disgust these perpetual defiances of logic; if he has a less accurate mind, he will learn by heart here words with inexact meaning, these descriptions of unperformed and unperformable experiments, and lines of reasoning which are sleight-of-hand passes, thus losing in such unreasoned memory work the little correct sense and critical mind he used to possess.

The student who, on the other hand, will have seen clearly the ideas we have just formulated will have done more than learned a certain number of propositions of physics; he will have understood the nature and true method of experimental science.[11]

7. Consequences relative to the mathematical development of physical theory

Through the preceding discussions the exact nature of physical theory and of its relations with experiment emerge more and more clearly and precisely.

The materials with which this theory is constructed are, on the one hand, the mathematical symbols serving to represent the various quantities and qualities of the physical world, and, on the other hand, the general postulates serving as principles. With these materials theory builds a logical structure; in drawing the plan of this structure it is hence bound to respect scrupulously the laws that logic imposes on all deductive reasoning and the rules that algebra prescribes for any mathematical operation.

The mathematical symbols used in theory have meaning only under very definite conditions; to define these symbols is to enumerate these conditions. Theory is forbidden to make use of these signs outside these conditions. Thus, an absolute temperature by definition can be positive only, and by definition the mass of a body is invariable; never will theory in its formulas give a zero or negative value to absolute temperature, and

never in its calculations will it make the mass of a given body vary.

Theory is in principle grounded on postulates, that is to say, on propositions that it is at leisure to state as it pleases, provided that no contradiction exists among the terms of the same postulate or between two distinct postulates. But once these postulates are set down it is bound to guard them with jealous rigor. For instance, if it has placed at the base of its system the principle of the conservation of energy, it must forbid any assertion in disagreement with this principle.

These rules bring all their weight to bear on a physical theory that is being constructed; a single default would make the system illogical and would oblige us to upset it in order to reconstruct another; but they are the only limitations imposed. IN THE COURSE OF ITS DEVELOPMENT, *a physical theory is free to choose any path it pleases provided that it avoids any logical contradiction; in particular, it is free not to take account of experimental facts.*

This is no longer the case WHEN THE THEORY HAS REACHED ITS COMPLETE DEVELOPMENT. When the logical structure has reached its highest point it becomes necessary to compare the set of mathematical propositions obtained as conclusions from these long deductions with the set of experimental facts; by employing the adopted procedures of measurement we must be sure that the second set finds in the first a sufficiently similar image, a sufficiently precise and complete symbol. If this agreement between the conclusions of theory and the facts of experiment were not to manifest a satisfactory approximation, the theory might well be logically constructed, but it should nonetheless be rejected because it would be contradicted by observation, because it would be *physically* false.

This comparison between the conclusions of theory and the truths of experiment is therefore indispensable, since only the test of facts can give physical validity to a theory. But this test by facts should bear exclusively on the conclusions of a theory, for only the latter are offered as an image of reality; the postulates serving as points of departure for the theory and the intermediary steps by which we go from the postulates to the conclusions do not have to be subject to this test.

We have in the foregoing pages very thoroughly analyzed the error of those who claim to subject one of the fundamental postulates of physics directly to the test of facts through a procedure such as a crucial experiment; and especially the error of those who accept as principles only "in-

ductions consisting exclusively in erecting into general laws not the interpretation but *the very result of a very large number of experiments*".[12]

There is another error lying very close to this one; it consists in requiring that all the operations performed by the mathematician connecting postulates with conclusions should have *a physical meaning*, in wishing "to reason only about *performable operations*", and in "introducing only magnitudes accessible to experiment".[13]

According to this requirement any magnitude introduced by the physicist in his formulas should be connected through a process of measurement to a property of a body; any algebraic operation performed on these magnitudes should be translated into concrete language by the employment of these processes of measurement; thus translated, it should express a real or possible fact.

Such a requirement, legitimate when it comes to the final formulas at the end of a theory, has no justification if applied to the intermediary formulas and operations establishing the transition from postulates to conclusions.

Let us take an example.

J. Willard Gibbs studied the theory of the dissociation of a perfect composite gas into its elements, also regarded as perfect gases. A formula was obtained expressing the law of chemical equilibrium internal to such a system. I propose to discuss this formula. For this purpose, keeping constant the pressure supporting the gaseous mixture, I consider the absolute temperature appearing in the formula and I make it vary from 0 to $+\infty$.

If we wish to attribute a physical meaning to this mathematical operation, we shall be confronted with a host of objections and difficulties. No thermometer can reveal temperatures below a certain limit, and none can determine temperatures high enough; this symbol which we call 'absolute temperature' cannot be translated through the means of measurement at our disposal into something having a concrete meaning unless its numerical value remains between a certain minimum and a certain maximum. Moreover, at temperatures sufficiently low this other symbol which thermodynamics calls 'a perfect gas' is no longer even an approximate image of any real gas.

These difficulties and many others, which it would take too long to enumerate, disappear if we heed the remarks we have formulated. In the

construction of the theory, the discussion we have just given is only an intermediary step, and there is no justification for seeking a physical meaning in it. Only when this discussion shall have led us to a series of propositions, shall we have to submit these propositions to the test of facts; then we shall inquire whether, within the limits in which the absolute temperature may be translated into concrete readings of a thermometer and the idea of a perfect gas is approximately embodied in the fluids we observe, the conclusions of our discussion agree with the results of experiment.

By requiring that mathematical operations by which postulates produce their consequences shall always have a physical meaning, we set unjustifiable obstacles before the mathematician and cripple his progress. G. Robin goes so far as to question the use of the differential calculus; if Professor Robin is intent on constantly and scrupulously satisfying this requirement, he would practically be unable to develop any calculation; theoretical deduction would be stopped in its tracks from the start. A more accurate idea of the method of physics and a more exact line of demarcation between the propositions which have to submit to factual test and those which are free to dispense with it would give back to the mathematician all his freedom and permit him to use all the resources of algebra for the greatest development of physical theories.

8. Are certain postulates of physical theory incapable of being refuted by experiment?

We recognize a correct principle by the facility with which it straightens out the complicated difficulties into which the use of erroneous principles brought us.

If, therefore, the idea we have put forth is correct, namely, that comparison is established necessarily between the *whole* of theory and the *whole* of experimental facts, we ought in the light of this principle to see the disappearance of the obscurities in which we should be lost by thinking that we are subjecting each isolated theoretical hypothesis to the test of facts.

Foremost among the assertions in which we shall aim at eliminating the appearance of paradox, we shall place one that has recently been often formulated and discussed. Stated first by G. Milhaud in connection with

the '*pure bodies*' of chemistry,[14] it has been developed at length and forcefully by H. Poincaré with regard to principles of mechanics;[15] Edouard Le Roy has also formulated it with great clarity.[16]

That assertion is as follows: Certain fundamental hypotheses of physical theory cannot be contradicted by any experiment, because they constitute in reality *definitions*, and because certain expressions in the physicist's usage take their meaning only through them.

Let us take one of the examples cited by Le Roy:

When a heavy body falls freely, the acceleration of its fall is constant. Can such a law be contradicted by experiment? No, for it constitutes the very definition of what is meant by "falling freely." If while studying the fall of a heavy body we found that this body does not fall with uniform acceleration, we should conclude not that the stated law is false, but that the body does not fall freely, that some cause obstructs its motion, and that the deviations of the observed facts from the law as stated would serve to discover this cause and to analyze its effects.

Thus, M. Le Roy concludes,

> laws are verifiable, taking things strictly ..., because they constitute the very criterion by which we judge appearances as well as the methods that it would be necessary to utilize in order to submit them to an inquiry whose precision is capable of exceeding any assignable limit.

Let us study again in greater detail, in the light of the principles previously set down, what this comparison is between the law of falling bodies and experiment.

Our daily observations have made us acquainted with a whole category of motions which we have brought together under the name of motions of heavy bodies; among these motions is the falling of a heavy body when it is not hindered by any obstacle. The result of this is that the words "free fall of a heavy body" have a meaning for the man who appeals only to the knowledge of common sense and who has no notion of physical theories.

On the other hand, in order to classify the laws of motion in question the physicist has created a theory, the theory of weight, an important application of rational mechanics. In that theory, intended to furnish a symbolic representation of reality, there is also the question of "free fall of a heavy body", and as a consequence of the hypotheses supporting this

whole scheme free fall must necessarily be a uniformly accelerated motion.

The words "free fall of a heavy body" now have two distinct meanings. For the man ignorant of physical theories, they have their *real* meaning, and they mean what common sense means in pronouncing them; for the physicist they have a *symbolic* meaning, and mean "uniformly accelerated motion". Theory would not have realized its aim if the second meaning were not the sign of the first, if a fall regarded as free by common sense were not also regarded as uniformly accelerated, or *nearly* uniformly accelerated, since common-sense observations are essentially devoid of precision, according to what we have already said.

This agreement, without which the theory would have been rejected without further examination, is finally arrived at: a fall declared by common sense to be nearly free is also a fall whose acceleration is nearly constant. But noticing this crudely approximate agreement does not satisfy us; we wish to push on and surpass the degree of precision which common sense can claim. With the aid of the theory that we have imagined, we put together apparatus enabling us to recognize with sensitive accuracy whether the fall of a body is or is not uniformly accelerated; this apparatus shows us that a certain fall regarded by common sense as a free fall has a slightly variable acceleration. The proposition which in our theory gives its symbolic meaning to the words "free fall" does not represent with sufficient accuracy the properties of the real and concrete fall that we have observed.

Two alternatives are then open to us.

In the first place, we can declare that we were right in regarding the fall studied as a free fall and in requiring that the theoretical definition of these words agree with our observations. In this case, since our theoretical definition does not satisfy this requirement, it must be rejected; we must construct another mechanics on new hypotheses, a mechanics in which the words "free fall" no longer signify "uniformly accelerated motion", but "fall whose acceleration varies according to a certain law".

In the second alternative, we may declare that we were wrong in establishing a connection between the concrete fall we have observed and the symbolic free fall defined by our theory, that the latter was too simplified a scheme of the former, that in order to represent suitably the fall as our experiments have reported it the theorist should give up imagining a weight falling freely and think in terms of a weight hindered by certain

obstacles like the resistance of the air, that in picturing the action of these obstacles by means of appropriate hypotheses he will compose a more complicated scheme than a free weight but one more apt to reproduce the details of the experiment; in short, in accord with the language we have previously established (Ch. IV, Sec. 3), we may seek to eliminate by means of suitable "corrections" the "causes of error", such as air resistance, which influenced our experiment.

M. Le Roy asserts that we shall prefer the second to the first alternative, and he is surely right in this. The reasons dictating this choice are easy to perceive. By taking the first alternative we should be obliged to destroy from top to bottom a very vast theoretical system which represents in a most satisfactory manner a very extensive and complex set of experimental laws. The second alternative, on the other hand, does not make us lose anything of the terrain already conquered by physical theory; in addition, it has succeeded in so large a number of cases that we can bank with interest on a new success. But in this confidence accorded the law of fall of weights, we see nothing analogous to the certainty that a mathematical definition draws from its very essence, that is, to the kind of certainty we have when it would be foolish to doubt that the various points on a circumference are all equidistant from the center.

We have here nothing more than a particular application of the principle set down in Section 2 of this chapter. A disagreement between the concrete facts constituting an experiment and the symbolic representation which theory substitutes for this experiment proves that some part of this symbol is to be rejected. But which part? This the experiment does not tell us; it leaves to our sagacity the burden of guessing. Now among the theoretical elements entering into the composition of this symbol there is always a certain number which the physicists of a certain epoch agree in accepting without test and which they regard as beyond dispute. Hence, the physicist who wishes to modify this symbol will surely bring his modification to bear on elements other than those just mentioned.

But what impels the physicist to act thus is *not* logical necessity. It would be awkward and ill inspired for him to do otherwise, but it would not be doing something logically absurd; he would not for all that be walking in the footsteps of the mathematician mad enough to contradict his own definitions. More than this, perhaps some day by acting differently, by refusing to invoke causes of error and take recourse to corrections

in order to reestablish agreement between the theoretical scheme and the fact, and by resolutely carrying out a reform among the propositions declared untouchable by common consent, he will accomplish the work of a genius who opens a new career for a theory.

Indeed, we must really guard ourselves against believing forever warranted those hypotheses which have become universally adopted conventions, and whose certainty seems to break through experimental contradiction by throwing the latter back on more doubtful assumptions. The history of physics shows us that very often the human mind has been led to overthrow such principles completely, though they have been regarded by common consent for centuries as inviolable axioms, and to rebuild its physical theories on new hypotheses.

Was there, for instance, a clearer or more certain principle for thousands of years than this one: In a homogeneous medium, light is propagated in a straight line? Not only did this hypothesis carry all former optics, catoptrics, and dioptrics, whose elegant geometric deductions represented at will an enormous number of facts, but it had become, so to speak, the physical definition of a straight line. It is to this hypothesis that any man wishing to make a straight line appeals, the carpenter who verifies the straightness of a piece of wood, the surveyor who lines up his sights, the geodetic surveyor who obtains a direction with the help of the pinholes of his alidade, the astronomer who defines the position of stars by the optical axis of his telescope. However, the day came when physicists tired of attributing to some cause of error the diffraction effects observed by Grimaldi, when they resolved to reject the law of the rectilinear propagation of light and to give optics entirely new foundations; and this bold resolution was the signal of remarkable progress for physical theory.

9. On hypotheses whose statement has no experimental meaning

This example, as well as others we could add from the history of science, should show that it would be very imprudent for us to say concerning a hypothesis commonly accepted today: "We are certain that we shall never be led to abandon it because of a new experiment, no matter how precise it is". Yet M. Poincaré does not hesitate to enunciate it concerning the principles of mechanics.[17]

To the reasons already given to prove that these principles cannot be reached by experimental refutation, M. Poincaré adds one which seems even more convincing: Not only can these principles not be refuted by experiment because they are the universally accepted rules serving to discover in our theories the weak spots indicated by these refutations, but also, they cannot be refuted by experiment because *the operation which would claim to compare them with the facts would have no meaning.*

Let us explain that by an illustration.

The principle of inertia teaches us that a material point removed from the action of any other body moves in a straight line with uniform motion. Now, we can observe only relative motions; we cannot, therefore, give an experimental meaning to this principle unless we assume a certain point chosen or a certain geometric solid taken as a fixed reference point to which the motion of the material point is related. The fixation of this reference frame constitutes an integral part of the statement of the law, for if we omitted it, this statement would be devoid of meaning. There are as many different laws as there are distinct frames of reference. We shall be stating one law of inertia when we say that the motion of an isolated point assumed to be seen from the earth is rectilinear and uniform, and another when we repeat the same sentence in referring the motion to the sun, and still another if the frame of reference chosen is the totality of fixed stars. But then, one thing is indeed certain, namely, that whatever the motion of a material point is, when seen from a first frame of reference, we can always and in infinite ways choose a second frame of reference such that seen from the latter our material point appears to move in a straight line with uniform motion. We cannot, therefore, attempt an experimental verification of the principle of inertia; false when we refer the motions to one frame of reference, it will become true when selection is made of another term of comparison, and we shall always be free to choose the latter. If the law of inertia stated by taking the earth as a frame of reference is contradicted by an observation, we shall substitute for it the law of inertia whose statement refers the motion to the sun; if the latter in its turn is contraverted, we shall replace the sun in the statement of the law by the system of fixed stars, and so forth. It is impossible to stop this loophole.

The principle of the equality of action and reaction, analyzed at length by M. Poincaré,[18] provides room for analogous remarks. This principle

may be stated thus: "The center of gravity of an isolated system can have only a uniform rectilinear motion".

This is the principle that we propose to verify by experiment.

Can we make this verification? For that it would be necessary for isolated systems to exist. Now, these systems do not exist; the only isolated system is the whole universe.

But we can observe only relative motions; the absolute motion of the center of the universe will therefore be forever unknown. We shall never be able to know if it is rectilinear and uniform or, better still, the question has no meaning. Whatever facts we may observe, we shall hence always be free to assume our principle is true.

Thus many a principle of mechanics has a form such that it is absurd to ask one's self: "Is this principle in agreement with experiment or not?" This strange character is not peculiar to the principles of mechanics; it also marks certain fundamental hypotheses of our physical or chemical theories.[19]

For example, chemical theory rests entirely on the "law of multiple proportions"; here is the exact statement of this law:

Simple bodies A, B, and C may by uniting in various proportions form various compounds M, M',.... The masses of the bodies A, B, and C combining to form the compound M are to one another as the three numbers a, b, and c. Then the masses of the elements A, B, and C combining to form the compound M' will be to one another as the numbers xa, yb, and zc (x, y, and z being three whole numbers).

Is this law perhaps subject to experimental test? Chemical analysis will make us acquainted with the chemical composition of the body M' not exactly but with a certain approximation. The uncertainty of the results obtained can be extremely small; it will never be strictly zero. Now, in whatever relations the elements A, B, and C are combined within the compound M', we can always represent these relations, with as close an approximation as you please, by the mutual relations of three products xa, yb, and zc, where x, y, and z are whole numbers; in other words, whatever the results given by the chemical analysis of the compound M', we are always sure to find three integers x, y, and z thanks to which the law of multiple proportions will be verified with a precision greater than that of the experiment. Therefore, no chemical analysis, no matter how refined, will ever be able to show the law of multiple proportions to be wrong.

In like manner, all crystallography rests entirely on the "law of rational indices" which is formulated in the following way:

A trihedral being formed by three faces of a crystal, a fourth face cuts the three edges of this trihedral at distances from the summit which are proportional to one another as three given numbers, the parameters of the crystal. Any other face whatsoever should cut these same edges at distances from the summit which are to one another as xa, yb, and zc, where x, y, and z are three integers, the indices of the new face of the crystal.

The most perfect protractor determines the direction of a crystal's face only with a certain degree of approximation; the relations among the three segments that such a face makes on the edges of the fundamental trihedral are always able to get by with a certain error; now, however small this error is, we can always choose three numbers x, y, and z such that the mutual relations of these segments are represented with the least amount of error by the mutual relations of the three numbers xa, yb, and zc; the crystallographer who would claim that the law of rational indices is made justifiable by his protractor would surely not have understood the very meaning of the words he is employing.

The law of multiple proportions and the law of rational indices are mathematical statements deprived of all physical meaning. A mathematical statement has physical meaning only if it retains a meaning when we introduce the word 'nearly' or 'approximately'. This is not the case with the statements we have just alluded to. Their object really is to assert that certain relations are *commensurable* numbers. They would degenerate into mere truisms if they were made to declare that these relations are approximately commensurable, for any incommensurable relation whatever is always approximately commensurable; it is even as near as you please to being commensurable.

Therefore, it would be absurd to wish to subject certain principles of mechanics to *direct* experimental test; it would be absurd to subject the law of multiple proportions or the law of rational indices to this *direct* test.

Does it follow that these hypotheses placed beyond the reach of direct experimental refutation have nothing more to fear from experiment? That they are guaranteed to remain immutable no matter what discoveries observation has in store for us? To pretend so would be a serious error.

Taken in isolation these different hypotheses have no experimental meaning; there can be no question of either confirming or contradicting

them by experiment. But these hypotheses enter as essential foundations into the construction of certain theories of rational mechanics, of chemical theory, of crystallography. The object of these theories is to represent experimental laws; they are schematisms intended essentially to be compared with facts.

Now this comparison might some day very well show us that one of our representations is ill adjusted to the realities it should picture, that the corrections which come and complicate our schematism do not produce sufficient concordance between this schematism and the facts, that the theory accepted for a long time without dispute should be rejected, and that an entirely different theory should be constructed on entirely different or new hypotheses. On that day some one of our hypotheses, which taken in isolation defied direct experimental refutation, will crumble with the system it supported under the weight of the contradictions inflicted by reality on the consequences of this system taken as a whole.[20]

In truth, hypotheses which by themselves have no physical meaning undergo experimental testing in exactly the same manner as other hypotheses. Whatever the nature of the hypothesis is, we have seen at the beginning of this chapter that it is never in isolation contradicted by experiment; experimental contradiction always bears as a whole on the entire group constituting a theory without any possibility of designating which proposition in this group should be rejected.

There thus disappears what might have seemed paradoxical in the following assertion: Certain physical theories rest on hypotheses which do not by themselves have any physical meaning.

10. Good sense is the judge of hypotheses which ought to be abandoned

When certain consequences of a theory are struck by experimental contradiction, we learn that this theory should be modified but we are not told by the experiment what must be changed. It leaves to the physicist the task of finding out the weak spot that impairs the whole system. No absolute principle directs this inquiry, which different physicists may conduct in very different ways without having the right to accuse one another of illogicality. For instance, one may be obliged to safeguard certain fundamental hypotheses while he tries to reestablish harmony be-

tween the consequences of the theory and the facts by complicating the schematism in which these hypotheses are applied, by invoking various causes of error, and by multiplying corrections. The next physicist, disdainful of these complicated artificial procedures, may decide to change some one of the essential assumptions supporting the entire system. The first physicist does not have the right to condemn in advance the boldness of the second one, nor does the latter have the right to treat the timidity of the first physicist as absurd. The methods they follow are justifiable only by experiment, and if they both succeed in satisfying the requirements of experiment each is logically permitted to declare himself content with the work that he has accomplished.

That does not mean that we cannot very properly prefer the work of one of the two to that of the other. Pure logic is not the only rule for our judgments; certain opinions which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable. These motives which do not proceed from logic and yet direct our choices, these "reasons which reason does not know" and which speak to the ample "mind of finesse" but not to the "geometric mind", constitute what is appropriately called good sense.

Now, it may be good sense that permits us to decide between two physicists. It may be that we do not approve of the haste with which the second one upsets the principles of a vast and harmoniously constructed theory whereas a modification of detail, a slight correction, would have sufficed to put these theories in accord with the facts. On the other hand, it may be that we may find it childish and unreasonable for the first physicist to maintain obstinately at any cost, at the price of continual repairs and many tangled-up stays, the worm-eaten columns of a building tottering in every part, when by razing these columns it would be possible to construct a simple, elegant, and solid system.

But these reasons of good sense do not impose themselves with the same implacable rigor that the prescriptions of logic do. There is something vague and uncertain about them; they do not reveal themselves at the same time with the same degree of clarity to all minds. Hence, the possibility of lengthy quarrels between the adherents of an old system and the partisans of a new doctrine, each camp claiming to have good sense on its side, each party finding the reasons of the adversary inadequate. The history of physics would furnish us with innumerable illustrations of

these quarrels at all times and in all domains. Let us confine ourselves to the tenacity and ingenuity with which Biot by a continual bestowal of corrections and accessory hypotheses maintained the emissionist doctrine in optics, while Fresnel opposed this doctrine constantly with new experiments favoring the wave theory.

In any event this state of indecision does not last forever. The day arrives when good sense comes out so clearly in favor of one of the two sides that the other side gives up the struggle even though pure logic would not forbid its continuation. After Foucault's experiment had shown that light traveled faster in air than in water, Biot gave up supporting the emission hypothesis; strictly, pure logic would not have compelled him to give it up, for Foucault's experiment was *not* the crucial experiment that Arago thought he saw in it, but by resisting wave optics for a longer time Biot would have been lacking in good sense.

Since logic does not determine with strict precision the time when an inadequate hypothesis should give way to a more fruitful assumption, and since recognizing this moment belongs to good sense, physicists may hasten this judgment and increase the rapidity of scientific progress by trying consciously to make good sense within themselves more lucid and more vigilant. Now nothing contributes more to entangle good sense and to disturb its insight than passions and interests. Therefore, nothing will delay the decision which should determine a fortunate reform in a physical theory more than the vanity which makes a physicist too indulgent towards his own system and too severe towards the system of another. We are thus led to the conclusion so clearly expressed by Claude Bernard: The sound experimental criticism of a hypothesis is subordinated to certain moral conditions; in order to estimate correctly the agreement of a physical theory with the facts, it is not enough to be a good mathematician and skillful experimenter; one must also be an impartial and faithful judge.

NOTES

* Chapter VI of *The Aim and Structure of Physical Theory*, translated by Philip Wiener (copyright © 1954 by Princeton University Press), pp. 180–218. Reprinted by permission of Princeton University Press. Originally published in French in 1906.

[1] Claude Bernard, *Introduction à la Médecine expérimentale*, Paris, 1865, p. 63. (Translator's note: Translated into English by H. C. Greene, *An Introduction to Experimental Medicine*, Henry Schuman, New York, 1949.)

[2] Claude Bernard, *Introduction à la Médecine expérimentale*, Paris, 1865, p. 64.
[3] *Ibid.*, p. 70.
[4] *Ibid.*, p. 67.
[5] Wilhelm Weber, *Electrodynamische Maassbestimmungen*, Leipzig, 1846. Translated into French in *Collection de Mémoires relatifs à la Physique* (Société française de Physique), Vol. III: *Mémoires sur l'Electrodynamique.*
[6] H. Poincaré, *Science et Hypothèse*, p. 179.
[7] G. Robin, *Oeuvres scientifiques, Thermodynamique générale*, Paris, 1901, Introduction, p. xii.
[8] Note on a lecture of M. Joubert, inspector-general of secondary-school instruction, *L'Enseignement secondaire*, April 15, 1903.
[9] G. Robin, *op. cit.*, p. ii.
[10] J. Bertrand, *Leçons sur la Théorie mathématique de l'Electricité*, Paris, 1890, p. 71.
[11] It will be objected undoubtedly that such teaching of physics would be hardly accessible to young minds; the answer is simple: Do not teach physics to minds not yet ready to assimilate it. Mme. de Sévigné used to say, speaking of young children: "Before you give them the food of a truckdriver, find out if they have the stomach of a truckdriver".
[12] G. Robin, *op. cit.*, p. xiv.
[13] *loc. cit.*
[14] G. Milhaud, 'La Science rationnelle', *Revue de Métaphysique et de Morale*, IV, 1896, p. 280. Reprinted in *Le Rationnel*, Paris, 1898, p. 45.
[15] H. Poincaré, 'Sur les Principes de la Mécanique', *Bibliothèque du Congrès International de Philosophie*, III: *Logique et Histoire des Sciences*, Paris, 1901, p. 457; 'Sur la Valeur objective des Théories physiques', *Revue de Métaphysique et de Morale*, X, 1902, p. 263; *La Science et l'Hypothèse*, p. 110.
[16] E. Le Roy, 'Un Positivisme Nouveau', *Revue de Métaphysique et de Morale*, IX, 1901, pp. 143–144.
[17] H. Poincaré, 'Sur les Principes de la Mécanique', *Bibliothèque du Congrès international de Philosophie*, Sec. III: 'Logique et Histoire des Sciences', Paris, 1901, pp. 475, 491.
[18] *Ibid.*, pp. 472ff.
[19] P. Duhem, *Le Mixte et la Combinaison chimique: Essai sur l'évolution d'une idée*, Paris, 1902, pp. 159–161.
[20] At the International Congress of Philosophy held in Paris in 1900, M. Poincaré developed this conclusion: "Thus is explained how experiment may have been able to edify (or suggest) the principles of mechanics, but will never be able to overthrow them". Against this conclusion, M. Hadamard offered various remarks, among them the following: "Moreover, in conformity with a remark of M. Duhem, it is not *an* isolated hypothesis but the whole group of the hypotheses of mechanics that we can try to verify experimentally". *Revue de Métaphysique et de Morale*, VIII, 1900, p. 559.

WILLARD VAN ORMAN QUINE

TWO DOGMAS OF EMPIRICISM*

0. Introduction

Modern empiricism has been conditioned in large part by two dogmas. One is a belief in some fundamental cleavage between truths which are *analytic*, or grounded in meanings independently of matters of fact, and truths which are *synthetic*, or grounded in fact. The other dogma is *reductionism*: the belief that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience. Both dogmas, I shall argue, are ill-founded. One effect of abandoning them is, as we shall see, a blurring of the supposed boundary between speculative metaphysics and natural science. Another effect is a shift toward pragmatism.

1. Background for analyticity

Kant's cleavage between analytic and synthetic truths was foreshadowed in Hume's distinction between relations of ideas and matters of fact, and in Leibniz's distinction between truths of reason and truths of fact. Leibniz spoke of the truths of reason as true in all possible worlds. Picturesqueness aside, this is to say that the truths of reason are those which could not possibly be false. In the same vein we hear analytic statements defined as statements whose denials are self-contradictory. But this definition has small explanatory value; for the notion of self-contradictoriness, in the quite broad sense needed for this definition of analyticity, stands in exactly the same need of clarification as does the notion of analyticity itself. The two notions are the two sides of a single dubious coin.

Kant conceived of an analytic statement as one that attributes to its subject no more than is already conceptually contained in the subject. This formulation has two shortcomings: it limits itself to statements of subject-predicate form, and it appeals to a notion of containment which is left at a metaphorical level. But Kant's intent, evident more from the use he makes of the notion of analyticity than from his definition of it,

can be restated thus: a statement is analytic when it is true by virtue of meanings and independently of fact. Pursuing this line, let us examine the concept of *meaning* which is presupposed.

Meaning, let us remember, is not to be identified with naming.[1] Frege's example of 'Evening Star' and 'Morning Star', and Russell's of 'Scott' and 'the author of *Waverley*', illustrate that terms can name the same thing but differ in meaning. The distinction between meaning and naming is no less important at the level of abstract terms. The terms '9' and 'the number of the planets' name one and the same abstract entity but presumably must be regarded as unlike in meaning; for astronomical observation was needed, and not mere reflection on meanings, to determine the sameness of the entity in question.

The above examples consist of singular terms, concrete and abstract. With general terms, or predicates, the situation is somewhat different but parallel. Whereas a singular term purports to name an entity, abstract or concrete, a general term does not; but a general term is *true* of an entity, or of each of many, or of none.[2] The class of all entities of which a general term is true is called the *extension* of the term. Now paralleling the contrast between the meaning of a singular term and the entity named, we must distinguish equally between the meaning of a general term and its extension. The general terms 'creature with a heart' and 'creature with kidneys', for example, are perhaps alike in extension but unlike in meaning.

Confusion of meaning with extension, in the case of general terms, is less common than confusion of meaning with naming in the case of singular terms. It is indeed a commonplace in philosophy to oppose intension (or meaning) to extension, or, in a variant vocabulary, connotation to denotation.

The Aristotelian notion of essence was the forerunner, no doubt, of the modern notion of intension or meaning. For Aristotle it was essential in men to be rational, accidental to be two-legged. But there is an important difference between this attitude and the doctrine of meaning. From the latter point of view it may indeed be conceded (if only for the sake of argument) that rationality is involved in the meaning of the word 'man' while two-leggedness is not; but two-leggedness may at the same time be viewed as involved in the meaning of 'biped' while rationality is not. Thus from the point of view of the doctrine of meaning it makes no sense to

say of the actual individual, who is at once a man and a biped, that his rationality is essential and his two-leggedness accidental or vice versa. Things had essences, for Aristotle, but only linguistic forms have meaning. Meaning is what essence becomes when it is divorced from the object of reference and wedded to the word.

For the theory of meaning a conspicuous question is the nature of its objects: what sort of things are meanings? A felt need for meant entities may derive from an earlier failure to appreciate that meaning and reference are distinct. Once the theory of meaning is sharply separated from the theory of reference, it is a short step to recognizing as the primary business of the theory of meaning simply the synonymy of linguistic forms and the analyticity of statements; meanings themselves, as obscure intermediary entities, may well be abandoned.[3]

The problem of analyticity then confronts us anew. Statements which are analytic by general philosophical acclaim are not, indeed, far to seek. They fall into two classes. Those of the first class, which may be called *logically true*, are typified by:

(1) No unmarried man is married.

The relevant feature of this example is that it not merely is true as it stands, but remains true under any and all reinterpretations of 'man' and 'married'. If we suppose a prior inventory of *logical* particles, comprising 'no', 'un-', 'not', 'if', 'then', 'and', etc., then in general a logical truth is a statement which is true and remains true under all reinterpretations of its components other than the logical particles.

But there is also a second class of analytic statements, typified by:

(2) No bachelor is married.

The characteristic of such a statement is that it can be turned into a logical truth by putting synonyms for synonyms; thus (2) can be turned into (1) by putting 'unmarried man' for its synonym 'bachelor'. We still lack a proper characterization of this second class of analytic statements, and therewith of analyticity generally, inasmuch as we have had in the above description to lean on a notion of "synonymy" which is no less in need of clarification than analyticity itself.

In recent years Carnap has tended to explain analyticity by appeal to what he calls state-descriptions.[4] A state-description is any exhaustive

assignment of truth values to the atomic, or noncompound, statements of the language. All other statements of the language are, Carnap assumes, built up of their component clauses by means of the familiar logical devices, in such a way that the truth value of any complex statement is fixed for each state-description by specifiable logical laws. A statement is then explained as analytic when it comes out true under every state-description. This account is an adaptation of Leibniz's "true in all possible worlds". But note that this version of analyticity serves its purpose only if the atomic statements of the language are, unlike 'John is a bachelor' and 'John is married', mutually independent. Otherwise there would be a state-description which assigned truth to 'John is a bachelor' and to 'John is married', and consequently 'No bachelors are married' would turn out synthetic rather than analytic under the proposed criterion. Thus the criterion of analyticity in terms of state-descriptions serves only for languages devoid of extralogical synonym-pairs, such as 'bachelor' and 'unmarried man' – synonym-pairs of the type which give rise to the "second class" of analytic statements. The criterion in terms of state-descriptions is a reconstruction at best of logical truth, not of analyticity.

I do not mean to suggest that Carnap is under any illusions on this point. His simplified model language with its state-descriptions is aimed primarily not at the general problem of analyticity but at another purpose, the clarification of probability and induction. Our problem, however, is analyticity; and here the major difficulty lies not in the first class of analytic statements, the logical truths, but rather in the second class, which depends on the notion of synonymy.

2. Definition

There are those who find it soothing to say that the analytic statements of the second class reduce to those of the first class, the logical truths, by *definition*; 'bachelor', for example, is *defined* as 'unmarried man'. But how do we find that 'bachelor' is defined as 'unmarried man'? Who defined it thus, and when? Are we to appeal to the nearest dictionary, and accept the lexicographer's formulation as law? Clearly this would be to put the cart before the horse. The lexicographer is an empirical scientist, whose business is the recording of antecedent facts; and if he glosses 'bachelor'

as 'unmarried man' it is because of his belief that there is a relation of synonymy between those forms, implicit in general or preferred usage prior to his own work. The notion of synonymy presupposed here has still to be clarified, presumably in terms relating to linguistic behavior. Certainly the "definition" which is the lexicographer's report of an observed synonymy cannot be taken as the ground of the synonymy.

Definition is not, indeed, an activity exclusively of philologists. Philosophers and scientists frequently have occasion to "define" a recondite term by paraphrasing it into terms of a more familiar vocabulary. But ordinarily such a definition, like the philologist's, is pure lexicography, affirming a relation of synonymy antecedent to the exposition in hand.

Just what it means to affirm synonymy, just what the interconnections may be which are necessary and sufficient in order that two linguistic forms be properly describable as synonymous, is far from clear; but, whatever these interconnections may be, ordinarily they are grounded in usage. Definitions reporting selected instances of synonymy come then as reports upon usage.

There is also, however, a variant type of definitional activity which does not limit itself to the reporting of preëxisting synonymies. I have in mind what Carnap calls *explication* – an activity to which philosophers are given, and scientists also in their more philosophical moments. In explication the purpose is not merely to paraphrase the definiendum into an outright synonym, but actually to improve upon the definiendum by refining or supplementing its meaning. But even explication, though not merely reporting a preëxisting synonymy between definiendum and definiens, does rest nevertheless on *other* preëxisting synonymies. The matter may be viewed as follows. Any word worth explicating has some contexts which, as wholes, are clear and precise enough to be useful; and the purpose of explication is to preserve the usage of these favored contexts while sharpening the usage of other contexts. In order that a given definition be suitable for purposes of explication, therefore, what is required is not that the definiendum in its antecedent usage be synonymous with the definiens, but just that each of these favored contexts of the definiendum, taken as a whole in its antecedent usage, be synonymous with the corresponding context of the definiens.

Two alternative definientia may be equally appropriate for the purposes of a given task of explication and yet not be synonymous with each

other; for they may serve interchangeably within the favored contexts but diverge elsewhere. By cleaving to one of these definientia rather than the other, a definition of explicative kind generates, by fiat, a relation of synonymy between definiendum and definiens which did not hold before. But such a definition still owes its explicative function, as seen, to preëxisting synonymies.

There does, however, remain still an extreme sort of definition which does not hark back to prior synonymies at all: namely, the explicitly conventional introduction of novel notations for purposes of sheer abbreviation. Here the definiendum becomes synonymous with the definiens simply because it has been created expressly for the purpose of being synonymous with the definiens. Here we have a really transparent case of synonymy created by definition; would that all species of synonymy were as intelligible. For the rest, definition rests on synonymy rather than explaining it.

The word 'definition' has come to have a dangerously reassuring sound, owing no doubt to its frequent occurrence in logical and mathematical writings. We shall do well to digress now into a brief appraisal of the role of definition in formal work.

In logical and mathematical systems either of two mutually antagonistic types of economy may be striven for, and each has its peculiar practical utility. On the one hand we may seek economy of practical expression – ease and brevity in the statement of multifarious relations. This sort of economy calls usually for distinctive concise notations for a wealth of concepts. Second, however, and oppositely, we may seek economy in grammar and vocabulary; we may try to find a minimum of basic concepts such that, once a distinctive notation has been appropriated to each of them, it becomes possible to express any desired further concept by mere combination and iteration of our basic notations. This second sort of economy is impractical in one way, since a poverty in basic idioms tends to a necessary lengthening of discourse. But it is practical in another way: it greatly simplifies theoretical discourse *about* the language, through minimizing the terms and the forms of construction wherein the language consists.

Both sorts of economy, though prima facie incompatible, are valuable in their separate ways. The custom has consequently arisen of combining both sorts of economy by forging in effect two languages, the one a part

of the other. The inclusive language, though redundant in grammar and vocabulary, is economical in message lengths, while the part, called primitive notation, is economical in grammar and vocabulary. Whole and part are correlated by rules of translation whereby each idiom not in primitive notation is equated to some complex built up of primitive notation. These rules of translation are the so-called *definitions* which appear in formalized systems. They are best viewed not as adjuncts to one language but as correlations between two languages, the one a part of the other.

But these correlations are not arbitrary. They are supposed to show how the primitive notations can accomplish all purposes, save brevity and convenience, of the redundant language. Hence the definiendum and its definiens may be expected, in each case, to be related in one or another of the three ways lately noted. The definiens may be a faithful paraphrase of the definiendum into the narrower notation, preserving a direct synonymy[5] as of antecedent usage; or the definiens may, in the spirit of explication, improve upon the antecedent usage of the definiendum; or finally, the definiendum may be a newly created notation, newly endowed with meaning here and now.

In formal and informal work alike, thus we find that definition – except in the extreme case of the explicitly conventional introduction of new notations – hinges on prior relations of synonymy. Recognizing then that the notion of definition does not hold the key to synonymy and analyticity, let us look further into synonymy and say no more of definition.

3. Interchangeability

A natural suggestion, deserving close examination, is that the synonymy of two linguistic forms consists simply in their interchangeability in all contexts without change of truth value – interchangeability, in Leibniz's phrase, *salva veritate*.[6] Note that synonyms so conceived need not even be free from vagueness, as long as the vaguenesses match.

But it is not quite true that the synonyms 'bachelor' and 'unmarried man' are everywhere interchangeable *salva veritate*. Truths which become false under substitution of 'ummarried man' for 'bachelor' are easily constructed with the help of 'bachelor of arts' or 'bachelor's buttons'; also with the help of quotation, thus:

'Bachelor' has less than ten letters.

Such counterinstances can, however, perhaps be set aside by treating the phrases 'bachelor of arts' and 'bachelor's buttons' and the quotation "bachelor" each as a single indivisible word and then stipulating that the interchangeability *salva veritate* which is to be the touchstone of synonymy is not supposed to apply to fragmentary occurrences inside of a word. This account of synonymy, supposing it acceptable on other counts, has indeed the drawback of appealing to a prior conception of "word" which can be counted on to present difficulties of formulation in its turn. Nevertheless some progress might be claimed in having reduced the problem of synonymy to a problem of wordhood. Let us pursue this line a bit, taking "word" for granted.

The question remains whether interchangeability *salva veritate* (apart from occurrences within words) is a strong enough condition for synonymy, or whether, on the contrary, some heteronymous expressions might be thus interchangeable. Now let us be clear that we are not concerned here with synonymy in the sense of complete identity in psychological associations or poetic quality; indeed no two expressions are synonymous in such a sense. We are concerned only with what may be called *cognitive* synonymy. Just what this is cannot be said without successfully finishing the present study; but we know something about it from the need which arose for it in connection with analyticity in Section 1. The sort of synonymy needed there was merely such that any analytic statement could be turned into a logical truth by putting synonyms for synonyms. Turning the tables and assuming analyticity, indeed, we could explain cognitive synonymy of terms as follows (keeping to the familiar example): to say that 'bachelor' and 'unmarried man' are cognitively synonymous is to say no more nor less than that the statement:

(3) All and only bachelors are unmarried men

is analytic.[7]

What we need is an account of cognitive synonymy not presupposing analyticity – if we are to explain analyticity conversely with help of cognitive synonymy as undertaken in Section 1. And indeed such an independent account of cognitive synonymy is at present up for consideration, namely, interchangeability *salva veritate* everywhere except within words. The question before us, to resume the thread at last, is whether such interchangeability is a sufficient condition for cognitive synonymy.

We can quickly assure ourselves that it is, by examples of the following sort. The statement:

(4) Necessarily all and only bachelors are bachelors

is evidently true, even supposing 'necessarily' so narrowly construed as to be truly applicable only to analytic statements. Then, if 'bachelor' and 'unmarried man' are interchangeable *salva veritate*, the result:

(5) Necessarily all and only bachelors are unmarried men

of putting 'unmarried man' for an occurrence of 'bachelor' in (4) must, like (4), be true. But to say that (5) is true is to say that (3) is analytic, and hence that 'bachelor' and 'unmarried man' are cognitively synonymous.

Let us see what there is about the above argument that gives it its air of hocus-pocus. The condition of interchangeability *salva veritate* varies in its force with variations in the richness of the language at hand. The above argument supposes we are working with a language rich enough to contain the adverb 'necessarily', this adverb being so construed as to yield truth when and only when applied to an analytic statement. But can we condone a language which contains such an adverb? Does the adverb really make sense? To suppose that it does is to suppose that we have already made satisfactory sense of 'analytic'. Then what are we so hard at work on right now?

Our argument is not flatly circular, but something like it. It has the form, figuratively speaking, of a closed curve in space.

Interchangeability *salva veritate* is meaningless until relativized to a language whose extent is specified in relevant respects. Suppose now we consider a language containing just the following materials. There is an indefinitely large stock of one-place predicates (for example, 'F' where 'Fx' means that x is a man) and many-place predicates (for example, 'G' where 'Gxy' means that x loves y), mostly having to do with extralogical subject matter. The rest of the language is logical. The atomic sentences consist each of a predicate followed by one or more variables 'x', 'y', etc.; and the complex sentences are built up of the atomic ones by truth functions ('not', 'and', 'or', etc.) and quantification.[8] In effect such a language enjoys the benefits also of descriptions and indeed singular terms generally, these being contextually definable in known ways.[9] Even abstract singular terms naming classes, classes of classes, etc., are contextually de-

finable in case the assumed stock of predicates includes the two-place predicate of class membership.[10] Such a language can be adequate to classical mathematics and indeed to scientific discourse generally, except insofar as the latter involves debatable devices such as contrary-to-fact conditionals or modal adverbs like 'necessarily'.[11] Now a language of this type is extensional, in this sense: any two predicates which agree extensionally (that is, are true of the same objects) are interchangeable *salva veritate*.[12]

In an extensional language, therefore, interchangeability *salva veritate* is no assurance of cognitive synonymy of the desired type. That 'bachelor' and 'unmarried man' are interchangeable *salva veritate* in an extensional language assures us of no more than that (3) is true. There is no assurance here that the extensional agreement of 'bachelor' and 'unmarried man' rests on meaning rather than merely on accidental matters of fact, as does the extensional agreement of 'creature with a heart' and 'creature with kidneys'.

For most purposes extensional agreement is the nearest approximation to synonymy we need care about. But the fact remains that extensional agreement falls far short of cognitive synonymy of the type required for explaining analyticity in the manner of Section 1. The type of cognitive synonymy required there is such as to equate the synonymy of 'bachelor' and 'unmarried man' with the analyticity of (3), not merely with the truth of (3).

So we must recognize that interchangeability *salva veritate*, if construed in relation to an extensional language, is not a sufficient condition of cognitive synonymy in the sense needed for deriving analyticity in the manner of Section 1. If a language contains an intensional adverb 'necessarily' in the sense lately noted, or other particles to the same effect, then interchangeability *salva veritate* in such a language does afford a sufficient condition of cognitive synonymy; but such a language is intelligible only insofar as the notion of analyticity is already understood in advance.

The effort to explain cognitive synonymy first, for the sake of deriving analyticity from it afterward as in Section 1, is perhaps the wrong approach. Instead we might try explaining analyticity somehow without appeal to cognitive synonymy. Afterward we could doubtless derive cognitive synonymy from analyticity satisfactorily enough if desired. We have seen that cognitive synonymy of 'bachelor' and 'unmarried man'

can be explained as analyticity of (3). The same explanation works for any pair of one-place predicates, of course, and it can be extended in obvious fashion to many-place predicates. Other syntactical categories can also be accommodated in fairly parallel fashion. Singular terms may be said to be cognitively synonymous when the statement of identity formed by putting '=' between them is analytic. Statements may be said simply to be cognitively synonymous when their biconditional (the result of joining them by 'if and only if') is analytic.[13] If we care to lump all categories into a single formulation, at the expense of assuming again the notion of "word" which was appealed to early in this section, we can describe any two linguistic forms as cognitively synonymous when the two forms are interchangeable (apart from occurrences within "words") *salva* (no longer *veritate* but) *analyticitate*. Certain technical questions arise, indeed, over cases of ambiguity or homonymy; let us not pause for them, however, for we are already digressing. Let us rather turn our backs on the problem of synonymy and address ourselves anew to that of analyticity.

4. Semantical rules

Analyticity at first seemed most naturally definable by appeal to a realm of meanings. On refinement, the appeal to meanings gave way to an appeal to synonymy or definition. But definition turned out to be a will-o'-the-wisp, and synonymy turned out to be best understood only by dint of a prior appeal to analyticity itself. So we are back at the problem of analyticity.

I do not know whether the statement 'Everything green is extended' is analytic. Now does my indecision over this example really betray an incomplete understanding, an incomplete grasp of the "meanings" of 'green' and 'extended'? I think not. The trouble is not with 'green' or 'extended', but with 'analytic'.

It is often hinted that the difficulty in separating analytic statements from synthetic ones in ordinary language is due to the vagueness of ordinary language and that the distinction is clear when we have a precise artificial language with explicit "semantical rules". This, however, as I shall now attempt to show, is a confusion.

The notion of analyticity about which we are worrying is a purported relation between statements and languages: a statement S is said to be

analytic for a language L, and the problem is to make sense of this relation generally, that is, for variable 'S' and 'L'. The gravity of this problem is not perceptibly less for artificial languages than for natural ones. The problem of making sense of the idiom 'S is analytic for L', with variable 'S' and 'L', retains its stubbornness even if we limit the range of the variable 'L' to artificial languages. Let me now try to make this point evident.

For artificial languages and semantical rules we look naturally to the writings of Carnap. His semantical rules take various forms, and to make my point I shall have to distinguish certain of the forms. Let us suppose, to begin with, an artificial language L_0 whose semantical rules have the form explicitly of a specification, by recursion or otherwise, of all the analytic statements of L_0. The rules tell us that such and such statements, and only those, are the analytic statements of L_0. Now here the difficulty is simply that the rules contain the word 'analytic', which we do not understand! We understand what expressions the rules attribute analyticity to, but we do not understand what the rules attribute to those expressions. In short, before we can understand a rule which begins 'A statement S is analytic for language L_0 if and only if...', we must understand the general relative term 'analytic for'; we must understand 'S is analytic for L' where 'S' and 'L' are variables.

Alternatively we may, indeed, view the so-called rule as a conventional definition of a new simple symbol 'analytic-for-L_0', which might better be written untendentiously as 'K' so as not to seem to throw light on the interesting word 'analytic'. Obviously any number of classes K, M, N, etc. of statements of L_0 can be specified for various purposes or for no purpose; what does it mean to say that K, as against M, N, etc., is the class of the "analytic" statements of L_0?

By saying what statements are analytic for L_0 we explain 'analytic-for-L_0' but not 'analytic', not 'analytic for'. We do not begin to explain the idiom 'S is analytic for L' with variable 'S' and 'L', even if we are content to limit the range of 'L' to the realm of artificial languages.

Actually we do know enough about the intended significance of 'analytic' to know that analytic statements are supposed to be true. Let us then turn to a second form of semantical rule, which says not that such and such statements are analytic but simply that such and such statements are included among the truths. Such a rule is not subject to the criticism of containing the un-understood word 'analytic'; and we may grant for the

sake of argument that there is no difficulty over the broader term 'true'. A semantical rule of this second type, a rule of truth, is not supposed to specify all the truths of the language; it merely stipulates, recursively or otherwise, a certain multitude of statements which, along with others unspecified, are to count as true. Such a rule may be conceded to be quite clear. Derivatively, afterward, analyticity can be demarcated thus: a statement is analytic if it is (not merely true but) true according to the semantical rule.

Still there is really no progress. Instead of appealing to an unexplained word 'analytic', we are now appealing to an unexplained phrase 'semantical rule'. Not every true statement which says that the statements of some class are true can count as a semantical rule – otherwise *all* truths would be "analytic" in the sense of being true according to semantical rules. Semantical rules are distinguishable, apparently, only by the fact of appearing on a page under the heading 'Semantical Rules'; and this heading is itself then meaningless.

We can say indeed that a statement is *analytic-for-*L_0 if and only if it is true according to such and such specifically appended "semantical rules", but then we find ourselves back at essentially the same case which was originally discussed: 'S is analytic-for-L_0 if and only if...'. Once we seek to explain 'S is analytic for L' generally for variable 'L' (even allowing limitation of 'L' to artificial languages), the explanation 'true according to the semantical rules of L' is unavailing; for the relative term 'semantical rule of' is as much in need of clarification, at least, as 'analytic for'.

It may be instructive to compare the notion of semantical rule with that of postulate. Relative to a given set of postulates, it is easy to say what a postulate is: it is a member of the set. Relative to a given set of semantical rules, it is equally easy to say what a semantical rule is. But given simply a notation, mathematical or otherwise, and indeed as thoroughly understood a notation as you please in point of the translations or truth conditions of its statements, who can say which of its true statements rank as postulates? Obviously the question is meaningless – as meaningless as asking which points in Ohio are starting points. Any finite (or effectively specifiable infinite) selection of statements (preferably true ones, perhaps) is as much a set of postulates as any other. The word 'postulate' is significant only relative to an act of inquiry; we apply the word to a set of statements just insofar as we happen, for the year or the moment, to be

thinking of those statements in relation to the statements which can be reached from them by some set of transformations to which we have seen fit to direct our attention. Now the notion of semantical rule is as sensible and meaningful as that of postulate, if conceived in a similarly relative spirit – relative, this time, to one or another particular enterprise of schooling unconversant persons in sufficient conditions for truth of statements of some natural or artificial language *L*. But from this point of view no one signalization of a subclass of the truths of *L* is intrinsically more a semantical rule than another; and, if 'analytic' means 'true by semantical rules', no one truth of *L* is analytic to the exclusion of another.[14]

It might conceivably be protested that an artificial language *L* (unlike a natural one) is a language in the ordinary sense *plus* a set of explicit semantical rules – the whole constituting, let us say, an ordered pair; and that the semantical rules of *L* then are specifiable simply as the second component of the pair *L*. But, by the same token and more simply, we might construe an artificial language *L* outright as an ordered pair whose second component is the class of its analytic statements; and then the analytic statements of *L* become specifiable simply as the statements in the second component of *L*. Or better still, we might just stop tugging at our bootstraps altogether.

Not all the explanations of analyticity known to Carnap and his readers have been covered explicitly in the above considerations, but the extension to other forms is not hard to see. Just one additional factor should be mentioned which sometimes enters: sometimes the semantical rules are in effect rules of translation into ordinary language, in which case the analytic statements of the artificial language are in effect recognized as such from the analyticity of their specified translations in ordinary language. Here certainly there can be no thought of an illumination of the problem of analyticity from the side of the artificial language.

From the point of view of the problem of analyticity the notion of an artificial language with semantical rules is a *feu follet par excellence*. Semantical rules determining the analytic statements of an artificial language are of interest only insofar as we already understand the notion of analyticity; they are of no help in gaining this understanding.

Appeal to hypothetical languages of an artificially simple kind could conceivably be useful in clarifying analyticity, if the mental or behavioral or cultural factors relevant to analyticity – whatever they may be – were

somehow sketched into the simplified model. But a model which takes analyticity merely as an irreducible character is unlikely to throw light on the problem of explicating analyticity.

It is obvious that truth in general depends on both language and extralinguistic fact. The statement 'Brutus killed Caesar' would be false if the world had been different in certain ways, but it would also be false if the word 'killed' happened rather to have the sense of 'begat'. Thus one is tempted to suppose in general that the truth of a statement is somehow analyzable into a linguistic component and a factual component. Given this supposition, it next seems reasonable that in some statements the factual component should be null; and these are the analytic statements. But, for all its a priori reasonableness, a boundary between analytic and synthetic statements simply has not been drawn. That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith.

5. The verification theory and reductionism

In the course of these somber reflections we have taken a dim view first of the notion of meaning, then of the notion of cognitive synonymy, and finally of the notion of analyticity. But what, it may be asked, of the verification theory of meaning? This phrase has established itself so firmly as a catchword of empiricism that we should be very unscientific indeed not to look beneath it for a possible key to the problem of meaning and the associated problems.

The verification theory of meaning, which has been conspicuous in the literature from Peirce onward, is that the meaning of a statement is the method of empirically confirming or infirming it. An analytic statement is that limiting case which is confirmed no matter what.

As urged in Section 1, we can as well pass over the question of meanings as entities and move straight to sameness of meaning, or synonymy. Then what the verification theory says is that statements are synonymous if and only if they are alike in point of method of empirical confirmation or infirmation.

This is an account of cognitive synonymy not of linguistic forms generally, but of statements.[15] However, from the concept of synonymy of statements we could derive the concept of synonymy for other linguistic

forms, by considerations somewhat similar to those at the end of Section 3. Assuming the notion of "word", indeed, we could explain any two forms as synonymous when the putting of the one form for an occurrence of the other in any statement (apart from occurrences within "words") yields a synonymous statement. Finally, given the concept of synonymy thus for linguistic forms generally, we could define analyticity in terms of synonymy and logical truth as in Section 1. For that matter, we could define analyticity more simply in terms of just synonymy of statements together with logical truth; it is not necessary to appeal to synonymy of linguistic forms other than statements. For a statement may be described as analytic simply when it is synonymous with a logically true statement.

So, if the verification theory can be accepted as an adequate account of statement synonymy, the notion of analyticity is saved after all. However, let us reflect. Statement synonymy is said to be likeness of method of empirical confirmation or infirmation. Just what are these methods which are to be compared for likeness? What, in other words, is the nature of the relation between a statement and the experiences which contribute to or detract from its confirmation?

The most naive view of the relation is that it is one of direct report. This is *radical reductionism*. Every meaningful statement is held to be translatable into a statement (true or false) about immediate experience. Radical reductionism, in one form or another, well antedates the verification theory of meaning explicitly so called. Thus Locke and Hume held that every idea must either originate directly in sense experience or else be compounded of ideas thus originating; and taking a hint from Tooke we might rephrase this doctrine in semantical jargon by saying that a term, to be significant at all, must be either a name of a sense datum or a compound of such names or an abbreviation of such a compound. So stated, the doctrine remains ambiguous as between sense data as sensory events and sense data as sensory qualities; and it remains vague as to the admissible ways of compounding. Moreover, the doctrine is unnecessarily and intolerably restrictive in the term-by-term critique which it imposes. More reasonably, and without yet exceeding the limits of what I have called radical reductionism, we may take full statements as our significant units – thus demanding that our statements as wholes be translatable into sense-datum language, but not that they be translatable term by term.

This emendation would unquestionably have been welcome to Locke and Hume and Tooke, but historically it had to await an important reorientation in semantics – the reorientation whereby the primary vehicle of meaning came to be seen no longer in the term but in the statement. This reorientation, seen in Bentham and Frege, underlies Russell's concept of incomplete symbols defined in use,[16] also it is implicit in the verification theory of meaning, since the objects of verification are statements.

Radical reductionism, conceived now with statements as units, set itself the task of specifying a sense-datum language and showing how to translate the rest of significant discourse, statement by statement, into it. Carnap embarked on this project in the *Aufbau*.

The language which Carnap adopted as his starting point was not a sense-datum language in the narrowest conceivable sense, for it included also the notations of logic, up through higher set theory. In effect it included the whole language of pure mathematics. The ontology implicit in it (that is, the range of values of its variables) embraced not only sensory events but classes, classes of classes, and so on. Empiricists there are who would boggle at such prodigality. Carnap's starting point is very parsimonious, however, in its extralogical or sensory part. In a series of constructions in which he exploits the resources of modern logic with much ingenuity, Carnap succeeds in defining a wide array of important additional sensory concepts which, but for his constructions, one would not have dreamed were definable on so slender a basis. He was the first empiricist who, not content with asserting the reducibility of science to terms of immediate experience, took serious steps toward carrying out the reduction.

If Carnap's starting point is satisfactory, still his constructions were, as he himself stressed, only a fragment of the full program. The construction of even the simplest statements about the physical world was left in a sketchy state. Carnap's suggestions on this subject were, despite their sketchiness, very suggestive. He explained spatio-temporal point-instants as quadruples of real numbers and envisaged assignment of sense qualities to point-instants according to certain canons. Roughly summarized, the plan was that qualities should be assigned to point-instants in such a way as to achieve the laziest world compatible with our experience. The principle of least action was to be our guide in constructing a world from experience.

Carnap did not seem to recognize, however, that his treatment of physical objects fell short of reduction not merely through sketchiness, but in principle. Statements of the form 'Quality q is at point-instant $x; y; z; t$' were, according to his canons, to be apportioned truth values in such a way as to maximize and minimize certain over-all features, and with growth of experience the truth values were to be progressively revised in the same spirit. I think this is a good schematization (deliberately over-simplified, to be sure) of what science really does; but it provides no indication, not even the sketchiest, of how a statement of the form 'Quality q is at $x; y; z; t$' could ever be translated into Carnap's initial language of sense data and logic. The connective 'is at' remains an added undefined connective; the canons counsel us in its use but not in its elimination.

Carnap seems to have appreciated this point afterward; for in his later writings he abandoned all notion of the translatability of statements about the physical world into statements about immediate experience. Reductionism in its radical form has long since ceased to figure in Carnap's philosophy.

But the dogma of reductionism has, in a subtler and more tenuous form, continued to influence the thought of empiricists. The notion lingers that to each statement, or each synthetic statement, there is associated a unique range of possible sensory events such that the occurrence of any of them would add to the likelihood of truth of the statement, and that there is associated also another unique range of possible sensory events whose occurrence would detract from that likelihood. This notion is of course implicit in the verification theory of meaning.

The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion, issuing essentially from Carnap's doctrine of the physical world in the *Aufbau*, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.[17]

The dogma of reductionism, even in its attenuated form, is intimately connected with the other dogma – that there is a cleavage between the analytic and the synthetic. We have found oursleves led, indeed, from the latter problem to the former through the verification theory of meaning. More directly, the one dogma clearly supports the other in this way: as long as it is taken to be significant in general to speak of the confirmation

and infirmation of a statement, it seems significant to speak also of a limiting kind of statement which is vacuously confirmed, *ipso facto*, come what may; and such a statement is analytic.

The two dogmas are, indeed, at root identical. We lately reflected that in general the truth of statements does obviously depend both upon language and upon extralinguistic fact; and we noted that this obvious circumstance carries in its train, not logically but all too naturally, a feeling that the truth of a statement is somehow analyzable into a linguistic component and a factual component. The factual component must, if we are empiricists, boil down to a range of confirmatory experiences, in the extreme case where the linguistic component is all that matters, a true statement is analytic. But I hope we are now impressed with how stubbornly the dsitinction between analytic and synthetic has resisted any straightforward drawing. I am impressed also, apart from prefabricated examples of black and white balls in an urn, with how baffling the problem has always been of arriving at any explicit theory of the empirical confirmation of a synthetic statement. My present suggestion is that it is nonsense, and the root of much nonsense, to speak of a linguistic component and a factual component in the truth of any individual statement. Taken collectively, science has its double dependence upon language and experience; but this duality is not significantly traceable into the statements of science taken one by one.

The idea of defining a symbol in use was, as remarked, an advance over the impossible term-by-term empiricism of Locke and Hume. The statement, rather than the term, came with Bentham to be recognized as the unit accountable to an empiricist critique. But what I am now urging is that even in taking the statement as unit we have drawn our grid too finely. The unit of empirical significance is the whole of science.

6. Empiricism without the dogmas

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjust-

ments in the interior of the field. Truth values have to be redistributed over some of our statements. Reëvaluation of some statements entails reëvaluation of others, because of their logical interconnections – the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reëvaluated one statement we must reëvaluate some others, which may be statements logically connected with the first or may be the statements of logical connections themselves. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reëvaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.

If this view is right, it is misleading to speak of the empirical content of an individual statement – especially if it is a statement at all remote from the experiential periphery of the field. Furthermore it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?

For vividness I have been speaking in terms of varying distances from a sensory periphery. Let me try now to clarify this notion without metaphor. Certain statements, though *about* physical objects and not sense experience, seem peculiarly germane to sense experience – and in a selective way: some statements to some experiences, others to others. Such statements, especially germane to particular experiences, I picture as near the periphery. But in this relation of "germaneness" I envisage nothing more than a loose association reflecting the relative likelihood, in practice, of our choosing one statement rather than another for revision in the event of recalcitrant experience. For example, we can imagine re-

calcitrant experiences to which we would surely be inclined to accommodate our system by reëvaluating just the statement that there are brick houses on Elm Street, together with related statements on the same topic. We can imagine other recalcitrant experiences to which we would be inclined to accommodate our system by reëvaluating just the statement that there are no centaurs, along with kindred statements. A recalcitrant experience can, I have urged, be accommodated by any of various alternative reëvaluations in various alternative quarters of the total system; but, in the cases which we are now imagining, our natural tendency to disturb the total system as little as possible would lead us to focus our revisions upon these specific statements concerning brick houses or centaurs. These statements are felt, therefore, to have a sharper empirical reference than highly theoretical statements of physics or logic or ontology. The latter statements may be thought of as relatively centrally located within the total network, meaning merely that little preferential connection with any particular sense data obtrudes itself.

As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries – not by definition in terms of experience, but simply as irreducible posits[18] comparable, epistemologically, to the gods of Homer. For my part I do, qua lay physicist, believe in physical objects and not in Homer's gods; and I consider it a scientific error to believe otherwise. But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind. Both sorts of entities enter our conception only as cultural posits. The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience.

Positing does not stop with macroscopic physical objects. Objects at the atomic level are posited to make the laws of macroscopic objects, and ultimately the laws of experience, simpler and more manageable; and we need not expect or demand full definition of atomic and subatomic entities in terms of macroscopic ones, any more than definition of macroscopic things in terms of sense data. Science is continuation of common sense, and it continues the common-sense expedient of swelling ontology to simplify theory.

Physical objects, small and large, are not the only posits. Forces are another example; and indeed we are told nowadays that the boundary between energy and matter is obsolete. Moreover, the abstract entities which are the substance of mathematics – ultimately classes and classes of classes and so on up – are another posit in the same spirit. Epistemologically these are myths on the same footing with physical objects and gods, neither better nor worse except for differences in the degree to which they expedite our dealings with sense experiences.

The over-all algebra of rational and irrational numbers is underdetermined by the algebra of rational numbers, but is smoother and more convenient; and it includes the algebra of rational numbers as a jagged or gerrymandered part.[19] Total science, mathematical and natural and human, is similarly but more extremely underdetermined by experience. The edge of the system must be kept squared with experience; the rest, with all its elaborate myths or fictions, has as its objective the simplicity of laws.

Ontological questions, under this view, are on a par with questions of natural science.[20] Consider the question whether to countenance classes as entities. This, as I have argued elsewhere,[21] is the question whether to quantify with respect to variables which take classes as values. Now Carnap (1950a) has maintained that is a question not of matters of fact but of choosing a convenient language form, a convenient conceptual scheme or framework for science. With this I agree, but only on the proviso that the same be conceded regarding scientific hypotheses generally. Carnap (1950a, p. 32n) has recognized that he is able to preserve a double standard for ontological questions and scientific hypotheses only by assuming an absolute distinction between the analytic and the synthetic; and I need not say again that this is a distinction which I reject.[22]

The issue over there being classes seems more a question of convenient conceptual scheme; the issue over there being centaurs, or brick houses on Elm Street, seems more a question of fact. But I have been urging that this difference is only one of degree, and that it turns upon our vaguely pragmatic inclination to adjust one strand of the fabric of science rather than another in accommodating some particular recalcitrant experience. Conservatism figures in such choices, and so does the quest for simplicity.

Carnap, Lewis, and others take a pragmatic stand on the question of

choosing between language forms, scientific frameworks; but their pragmatism leaves off at the imagined boundary between the analytic and the synthetic. In repudiating such a boundary I espouse a more thorough pragmatism. Each man is given a scientific heritage plus a continuing barrage of sensory stimulation; and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic.

NOTES

* Reprinted by permission of the publishers from Willard Van Orman Quine, *From a Logical Point of View*, Harvard University Press, Cambridge, Mass., Copyright © 1953, 1961 by the President and Fellows of Harvard College. First published in *The Philosophical Review* **60** (1951).

[1] See *From a Logical Point of View (FLPV)*, p. 9.

[2] See *FLPV*, p. 10, and pp. 107–115.

[3] See *FLPV*, pp. 11f, and pp. 48f.

[4] Carnap (1947), pp. 9ff; (1950b), pp. 70ff.

[5] According to an important variant sense of 'definition', the relation preserved may be the weaker relation of mere agreement in reference; see *FLPV*, p. 132. But definition in this sense is better ignored in the present connection, being irrelevant to the question of synonymy.

[6] Cf. Lewis (1918), p. 373.

[7] This is cognitive synonymy in a primary, broad sense. Carnap (1947, pp. 56ff) and Lewis (1946, pp. 83ff) have suggested how, once this notion is at hand, a narrower sense of cognitive synonymy which is preferable for some purposes can in turn be derived. But this special ramification of concept-building lies aside from the present purposes and must not be confused with the broad sort of cognitive synonymy here concerned.

[8] Pp. 81ff, *FLPV*, contain a description of just such a language except that there happens there to be just one predicate, the two-place predicate 'e'.

[9] See *FLPV*, pp. 5–8; also pp. 85f, 166f.

[10] See *FLPV*, p. 87.

[11] On such devices see also Essay VIII, 'Reference and Modality', in *FLPV*.

[12] This is the substance of Quine, *121.

[13] The 'if and only if' itself is intended in the truth functional sense. See Carnap (1947), p. 14.

[14] The foregoing paragraph was not part of the present essay as originally published. It was prompted by Martin (see Bibliography), as was the end of Essay VII, 'Notes on the Theory of Reference', in *FLPV*.

[15] The doctrine can indeed be formulated with terms rather than statements as the units. Thus Lewis describes the meaning of a term as "*a criterion in mind*, by reference to which one is able to apply or refuse to apply the expression in question in the case of presented, or imagined, things or situations" (1946, p. 133). – For an instructive account of the vicissitudes of the verification theory of meaning, centered however on the question of meaning*fulness* rather than synonymy and analyticity, see Hempel.

[16] See *FLPV*, p. 6.
[17] This doctrine was well argued by Duhem, pp. 303–328. Or see Lowinger, pp. 132–140.
[18] Cf. *FLPV*, pp. 17f.
[19] Cf. p. 18, *FLPV*.
[20] "L'ontologie fait corps avec la science elle-même et ne peut en être separée". Meyerson, p. 439.
[21] *FLPV*, pp. 12f; pp. 102ff.
[22] For an effective expression of further misgivings over this distinction, see White.

BIBLIOGRAPHY

Carnap, Rudolf: *Meaning and Necessity*, University of Chicago Press, Chicago, 1947.
Carnap, Rudolf: 'Empiricism, Semantics, and Ontology', *Revue Internationale de Philosophie* **4** (1950a), 20–40.
Carnap, Rudolf: *Logical Foundations of Probability*, University of Chicago Press, Chicago, 1950b.
Duhem, Pierre: *La Théorie Physique: Son Objet et Sa Structure*, Paris, 1906.
Hempel, C. G.: 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie* **4** (1950), 41–63.
Hempel, C. G.: 'The Concept of Cognitive Significance: A Reconsideration', *Proceedings of American Academy of Arts and Sciences* **80** (1951), 61–77.
Lewis, C. I.: *A Survey of Symbolic Logic*, Berkeley, 1918.
Lewis, C. I.: *An Analysis of Knowledge and Valuation*, Open Court, Le Salle, III., 1946.
Lowinger, Armand: *The Methodology of Pierre Duhem*, Columbia University Press, New York, 1941.
Martin, R. M.: 'On "Analytic"', *Philosophical Studies* **3** (1952), 42–47.
Meyerson, Émile: *Identité et Realité*, Paris, 1908; 4th ed., 1932.
Quine, W. V.: *Mathematical Logic*, Norton, New York, 1940; Harvard University Press, Cambridge, 1947; rev. ed., Harvard University Press, Cambridge, 1951.
White, Morton: 'The Analytic and the Synthetic: An Untenable Dualism', in Sidney Hook (ed.), *John Dewey: Philosopher of Science and Freedom*, Dial Press, New York, 1950, pp. 316–330.

CARL G. HEMPEL

EMPIRICIST CRITERIA OF COGNITIVE SIGNIFICANCE: PROBLEMS AND CHANGES*

1. THE GENERAL EMPIRICIST CONCEPTION OF COGNITIVE AND EMPIRICAL SIGNIFICANCE**

It is a basic principle of contemporary empiricism that a sentence makes a cognitively significant assertion, and thus can be said to be either true or false, if and only if either (1) it is analytic or contradictory – in which case it is said to have purely logical meaning or significance – or else (2) it is capable, at least potentially, of test by experiential evidence – in which case it is said to have empirical meaning or significance. The basic tenet of this principle, and especially of its second part, the so-called testability criterion of empirical meaning (or better: meaningfulness), is not peculiar to empiricism alone: it is characteristic also of contemporary operationism, and in a sense of pragmatism as well; for the pragmatist maxim that a difference must make a difference to be a difference may well be construed as insisting that a verbal difference between two sentences must make a difference in experiential implications if it is to reflect a difference in meaning.

How this general conception of cognitively significant discourse led to the rejection, as devoid of logical and empirical meaning, of various formulations in speculative metaphysics, and even of certain hypotheses offered within empirical science, is too well known to require recounting. I think that the general intent of the empiricist criterion of meaning is basically sound, and that notwithstanding much oversimplification in its use, its critical application has been, on the whole, enlightening and salutary. I feel less confident, however, about the possibility of restating the general idea in the form of precise and general criteria which establish sharp dividing lines (a) between statements of purely logical and statements of empirical significance, and (b) between those sentences which do have cognitive significance and those which do not.

In the present paper, I propose to reconsider these distinctions as conceived in recent empiricism, and to point out some of the difficulties they

present. The discussion will concern mainly the second of the two distinctions; in regard to the first, I shall limit myself to a few brief remarks.

2. The earlier testability criteria of meaning and their shortcomings

Let us note first that any general criterion of cognitive significance will have to meet certain requirements if it is to be at all acceptable. Of these, we note one, which we shall consider here as expressing a necessary, though by no means sufficient, *condition of adequacy* for criteria of cognitive significance.

(A) If under a given criterion of cognitive significance, a sentence N is non-significant, then so must be all truth-functional compound sentences in which N occurs nonvacuously as a component. For if N cannot be significantly assigned a truth value, then it is impossible to assign truth values to the compound sentences containing N; hence, they should be qualified as nonsignificant as well.

We note two corollaries of requirement (A):

(A1) If under a given criterion of cognitive significance, a sentence S is nonsignificant, then so must be its negation, $\sim S$.

(A2) If under a given criterion of cognitive significance, a sentence N is nonsignificant, then so must be any conjunction $N \cdot S$ and any disjunction $N \vee S$, no matter whether S is significant under the given criterion or not.

We now turn to the initial attempts made in recent empiricism to establish general criteria of cognitive significance. Those attempts were governed by the consideration that a sentence, to make an empirical assertion must be capable of being borne out by, or conflicting with, phenomena which are potentially capable of being directly observed. Sentences describing such potentially observable phenomena – no matter whether the latter do actually occur or not – may be called observation sentences. More specifically, an *observation sentence* might be construed as a sentence – no matter whether true or false – which asserts or denies

that a specified object, or group of objects, of macroscopic size has a particular *observable characteristic*, i.e., a characteristic whose presence or absence can, under favorable circumstances, be ascertained by direct observation.[1]

The task of setting up criteria of empirical significance is thus transformed into the problem of characterizing in a precise manner the relationship which obtains between a hypothesis and one or more observation sentences whenever the phenomena described by the latter either confirm or disconfirm the hypothesis in question. The ability of a given sentence to enter into that relationship to some set of observation sentences would then characterize its testability-in-principle, and thus its empirical significance. Let us now briefly examine the major attempts that have been made to obtain criteria of significance in this manner.

One of the earliest criteria is expressed in the so-called *verifiability requirement*. According to it, a sentence is empirically significant if and only if it is not analytic and is capable, at least in principle, of complete verification by observational evidence; i.e., if observational evidence can be described which, if actually obtained, would conclusively establish the truth of the sentence.[2] With the help of the concept of observation sentence, we can restate this requirement as follows: A sentence S has empirical meaning if and only if it is possible to indicate a finite set of observation sentences, $O_1, O_2 \ldots, O_n$, such that if these are true, then S is necessarily true, too. As stated, however, this condition is satisfied also if S is an analytic sentence or if the given observation sentences are logically incompatible with each other. By the following formulation, we rule these cases out and at the same time express the intended criterion more precisely:

2.1. Requirement of complete verifiability in principle. A sentence has empirical meaning if and only if it is not analytic and follows logically from some finite and logically consistent class of observation sentences.[3] These observation sentences need not be true, for what the criterion is to explicate is testability by "potentially observable phenomena", or testability "in principle".

In accordance with the general conception of cognitive significance outlined earlier, a sentence will now be classified as cognitively significant if either it is analytic or contradictory, or it satisfies the verifiability requirement.

This criterion, however, has several serious defects. One of them has been noted by several writers:

(a) Let us assume that the properties of being a stork and of being red-legged are both observable characteristics, and that the former does not logically entail the latter. Then the sentence

($S1$) All storks are red-legged,

is neither analytic nor contradictory; and clearly, it is not deducible from a finite set of observation sentences. Hence, under the contemplated criterion, ($S1$) is devoid of empirical significance; and so are all other sentences purporting to express universal regularities or general laws. And since sentences of this type constitute an integral part of scientific theories, the verifiability requirement must be regarded as overly restrictive in this respect.

Similarly, the criterion disqualifies all sentences such as 'For any substance there exists some solvent', which contain both universal and existential quantifiers (i.e., occurrences of the terms 'all' and 'some' or their equivalents); for no sentences of this kind can be logically deduced from any finite set of observation sentences.

Two further defects of the verifiability requirement do not seem to have been widely noticed:

(b) As is readily seen, the negation of ($S1$)

($\sim S1$) There exists at least one stork that is not red-legged

is deducible from any two observation sentences of the type 'a is a stork' and 'a is not red-legged'. Hence, ($\sim S1$) is cognitively significant under our criterion, but ($S1$) is not, and this constitutes a violation of condition (A1).

(c) Let S be a sentence which does, and N a sentence which does not satisfy the verifiability requirement. Then S is deducible from some set of observation sentences; hence, by a familiar rule of logic, $S \vee N$ is deducible from the same set, and therefore cognitively significant according to our criterion. This violates condition (A2) above.[4]

Strictly analogous considerations apply to an alternative criterion, which makes complete falsifiability in principle the defining characteristic of empirical significance. Let us formulate this criterion as follows:

2.2. REQUIREMENT OF COMPLETE FALSIFIABILITY IN PRINCIPLE. A sentence has empirical meaning if and only if its negation is not analytic and follows logically from some finite logically consistent class of observation sentences.

This criterion qualifies a sentence as empirically meaningful if its negation satisfies the requirement of complete verifiability; as it is to be expected, it is therefore inadequate on similar grounds as the latter:

(a) It denies cognitive significance to purely existential hypotheses, such as 'There exists at least one unicorn', and all sentences whose formulation calls for mixed – i.e., universal and existential – quantification, such as 'For every compound there exists some solvent', for none of these can possibly be conclusively falsified by a finite number of observation sentences.

(b) If 'P' is an observation predicate, then the assertion that all things have the property P is qualified as significant, but its negation, being equivalent to a purely existential hypothesis, is disqualified [cf. (a)]. Hence, criterion (2.2.) gives rise to the same dilemma as (2.1.).

(c) If a sentence S is completely falsifiable whereas N is a sentence which is not, then their conjunction, $S \cdot N$ (i.e., the expression obtained by connecting the two sentences by the word 'and') is completely falsifiable; for if the negation of S is entailed by a class of observation sentences, then the negation of $S \cdot N$ is, *a fortiori*, entailed by the same class. Thus, the criterion allows empirical significance to many sentences which an adequate empiricist criterion should rule out, such as 'All swans are white and the absolute is perfect'.

In sum, then, interpretations of the testability criterion in terms of complete verifiability or of complete falsifiability are inadequate because they are overly restrictive in one direction and overly inclusive in another, and because both of them violate the fundamental requirement A.

Several attempts have been made to avoid these difficulties by construing the testability criterion as demanding merely a partial and possibly indirect confirmability of empirical hypotheses by observational evidence.

A formulation suggested by Ayer[5] is characteristic of these attempts to set up a clear and sufficiently comprehensive criterion of confirmability. It states, in effect, that a sentence S has empirical import if from S in conjunction with suitable subsidiary hypotheses it is possible to derive

observation sentences which are not derivable from the subsidiary hypotheses alone.

This condition is suggested by a closer consideration of the logical structure of scientific testing; but it is much too liberal as it stands. Indeed, as Ayer himself has pointed out in the second edition of his book, *Language, Truth, and Logic*,[6] his criterion allows empirical import to any sentence whatever. Thus, e.g., if S is the sentence 'The absolute is perfect', it suffices to choose as a subsidiary hypothesis the sentence 'If the absolute is perfect then this apple is red' in order to make possible the deduction of the observation sentence 'This apple is red', which clearly does not follow from the subsidiary hypothesis alone.

To meet this objection, Ayer proposed a modified version of his testability criterion. In effect, the modification restricts the subsidiary hypotheses mentioned in the previous version to sentences which either are analytic or can independently be shown to be testable in the sense of the modified criterion.[7]

But it can readily be shown that this new criterion, like the requirement of complete falsifiability, allows empirical significance to any conjunction $S \cdot N$, where S satisfies Ayer's criterion while N is a sentence such as 'The absolute is perfect', which is to be disqualified by that criterion. Indeed, whatever consequences can be deduced from S with the help of permissible subsidiary hypotheses can also be deduced from $S \cdot N$ by means of the same subsidiary hypotheses; and as Ayer's new criterion is formulated essentially in terms of the deducibility of a certain type of consequence from the given sentence, it countenances $S \cdot N$ together with S. Another difficulty has been pointed out by Church, who has shown[8] that if there are any three observation sentences none of which alone entails any of the others, then it follows for any sentence S whatsoever that either it or its denial has empirical import according to Ayer's revised criterion.

All the criteria considered so far attempt to explicate the concept of empirical significance by specifying certain logical connections which must obtain between a significant sentence and suitable observation sentences. It seems now that this type of approach offers little hope for the attainment of precise criteria of meaningfulness: this conclusion is suggested by the preceding survey of some representative attempts, and it receives additional support from certain further considerations, some of which will be presented in the following sections.

3. Characterization of Significant Sentences by Criteria for their Constituent Terms

An alternative procedure suggests itself which again seems to reflect well the general viewpoint of empiricism: It might be possible to characterize cognitively significant sentences by certain conditions which their constituent terms have to satisfy. Specifically, it would seem reasonable to say that all extralogical terms[9] in a significant sentence must have experiential reference, and that therefore their meanings must be capable of explication by reference to observables exclusively.[10] In order to exhibit certain analogies between this approach and the previous one, we adopt the following terminological conventions:

Any term that may occur in a cognitively significant sentence will be called a *cognitively significant term*. Furthermore, we shall understand by an *observation term* any term which either (a) is an *observation predicate*, i.e., signifies some observable characteristic (as do the terms 'blue', 'warm', 'soft', 'coincident with', 'of greater apparent brightness than') or (b) names some physical object of macroscopic size (as do the terms 'the needle of this instrument', 'the Moon', 'Krakatoa Volcano', 'Greenwich, England', 'Julius Caesar').

Now while the testability criteria of meaning aimed at characterizing the cognitively significant sentences by means of certain inferential connections in which they must stand to some observation sentences, the alternative approach under consideration would instead try to specify the vocabulary that may be used in forming significant sentences. This vocabulary, the class of significant terms, would be characterized by the condition that each of its elements is either a logical term or else a term with empirical significance; in the latter case, it has to stand in certain definitional or explicative connections to some observation terms. This approach certainly avoids any violations of our earlier conditions of adequacy. Thus, e.g., if S is a significant sentence, i.e., contains cognitively significant terms only, then so is its denial, since the denial sign, and its verbal equivalents, belong to the vocabulary of logic and are thus significant. Again, if N is a sentence containing a non-significant term, then so is any compound sentence which contains N.

But this is not sufficient, of course. Rather, we shall now have to consider a crucial question analogous to that raised by the previous approach:

Precisely how are the logical connections between empirically significant terms and observation terms to be construed if an adequate criterion of cognitive significance is to result? Let us consider some possibilities.

3.1. The simplest criterion that suggests itself might be called the *requirement of definability*. It would demand that any term with empirical significance must be explicitly definable by means of observation terms.

This criterion would seem to accord well with the maxim of operationism that all significant terms of empirical science must be introduced by operational definitions. However, the requirement of definability is vastly too restrictive, for many important terms of scientific and even prescientific discourse cannot be explicitly defined by means of observation terms.

In fact, as Carnap[11] has pointed out, an attempt to provide explicit definitions in terms of observables encounters serious difficulties as soon as disposition terms, such as 'soluble', 'malleable', 'electric conductor', etc., have to be accounted for; and many of these occur even on the prescientific level of discourse.

Consider, for example, the word 'fragile'. One might try to define it by saying that an object x is fragile if and only if it satisfies the following condition: If at any time t the object is sharply struck, then it breaks at that time. But if the statement connectives in this phrasing are construed truth-functionally, so that the definition can be symbolized by

$$\text{(D)} \qquad Fx \equiv (t)\,(Sxt \supset Bxt),$$

then the predicate 'F' thus defined does not have the intended meaning. For let a be any object which is not fragile (e.g., a raindrop or a rubber band), but which happens not to be sharply struck at any time throughout its existence. Then 'Sat' is false and hence '$Sat \supset Bat$' is true for all values of 't'; consequently, 'Fa' is true though a is not fragile.

To remedy this defect, one might construe the phrase 'if...then...' in the original definiens as having a more restrictive meaning than the truth-functional conditional. This meaning might be suggested by the subjunctive phrasing 'If x were to be sharply struck at any time t, then x would break at t.' But a satisfactory elaboration of this construal would require a clarification of the meaning and the logic of counterfactual and subjunctive conditionals, which is a thorny problem.[12]

An alternative procedure was suggested by Carnap in his theory of reduction sentences.[13] These are sentences which, unlike definitions, specify the meaning of a term only conditionally or partially. The term 'fragile', for example, might be introduced by the following reduction sentence:

(R) $(x)(t)\,[Sxt \supset (Fx \equiv Bxt)]$

which specifies that if x is sharply struck at any time t, then x is fragile if and only if x breaks at t.

Our earlier difficulty is now avoided, for if a is a nonfragile object that is never sharply struck, then that expression in R which follows the quantifiers is true of a; but this does not imply that 'Fa' is true. But the reduction sentence R specifies the meaning of 'F' only for application to those objects which meet the 'test condition' of being sharply struck at some time; for these it states that fragility then amounts to breaking. For objects that fail to meet the test condition, the meaning of 'F' is left undetermined. In this sense, reduction sentences have the character of partial or conditional definitions.

Reduction sentences provide a satisfactory interpretation of the experiential import of a large class of disposition terms and permit a more adequate formulation of so-called operational definitions, which, in general, are not complete definitions at all. These considerations suggest a greatly liberalized alternative to the requirement of definability:

3.2. *The Requirement of Reducibility*

Every term with empirical significance must be capable of introduction, on the basis of observation terms, through chains of reduction sentences.

This requirement is characteristic of the liberalized versions of positivism and physicalism which, since about 1936, have superseded the older, overly narrow conception of a full definability of all terms of empirical science by means of observables,[14] and it avoids many of the shortcomings of the latter. Yet, reduction sentences do not seem to offer an adequate means for the introduction of the central terms of advanced scientific theories, often referred to as theoretical constructs. This is indicated by the following considerations: A chain of reduction sentences provides a necessary and a sufficient condition for the applicability of the term it introduces. (When the two conditions coincide, the chain is tantamount to an explicit definition.) But now take, for example, the concept of length

as used in classical physical theory. Here, the length in centimeters of the distance between two points may assume any positive real number as its value; yet it is clearly impossible to formulate, by means of observation terms, a sufficient condition for the applicability of such expressions as 'having a length of $\sqrt{2}$ cm' and 'having a length of $\sqrt{2}+10^{-100}$ cm'; for such conditions would provide a possibility for discrimination, in observational terms, between two lengths which differ by only 10^{-100} cm.[15]

It would be ill-advised to argue that for this reason, we ought to permit only such values of the magnitude, length, as permit the statement of sufficient conditions in terms of observables. For this would rule out, among others, all irrational numbers and would prevent us from assigning, to the diagonal of a square with sides of length 1, the length $\sqrt{2}$, which is required by Euclidean geometry. Hence, the principles of Euclidean geometry would not be universally applicable in physics. Similarly, the principles of the calculus would become inapplicable, and the system of scientific theory as we know it today would be reduced to a clumsy, unmanageable torso. This, then, is no way of meeting the difficulty. Rather, we shall have to analyze more closely the function of constructs in scientific theories, with a view to obtaining through such an analysis a more adequate characterization of cognitively significant terms.

Theoretical constructs occur in the formulation of scientific theories. These may be conceived of, in their advanced stages, as being stated in the form of deductively developed axiomatized systems. Classical mechanics, or Euclidean or some Non-Euclidean form of geometry in physical interpretation, present examples of such systems. The extralogical terms used in a theory of this kind may be divided, in familiar manner, into primitive or basic terms, which are not defined within the theory, and defined terms, which are explicitly defined by means of the primitives. Thus, e.g., in Hilbert's axiomatization of Euclidean geometry, the terms 'point', 'straight line', 'between' are among the primitives, while 'line segment', 'angle', 'triangle', 'length' are among the defined terms. The basic and the defined terms together with the terms of logic constitute the vocabulary out of which all the sentences of the theory are constructed. The latter are divided, in an axiomatic presentation, into primitive statements (also called postulates or basic statements) which, in the theory, are not derived from any other statements, and derived ones, which are obtained by logical deduction from the primitive statements.

From its primitive terms and sentences, an axiomatized theory can be developed by means of purely formal principles of definition and deduction, without any consideration of the empirical significance of its extra-logical terms. Indeed, this is the standard procedure employed in the axiomatic development of uninterpreted mathematical theories such as those of abstract groups or rings or lattices, or any form of pure (i.e., noninterpreted) geometry.

However, a deductively developed system of this sort can constitute a scientific theory only if it has received an empirical interpretation[16] which renders it relevant to the phenomena of our experience. Such interpretation is given by assigning a meaning, in terms of observables, to certain terms or sentences of the formalized theory. Frequently, an interpretation is given not for the primitive terms or statements but rather for some of the terms definable by means of the primitives, or for some of the sentences deducible from the postulates.[17] Furthermore, interpretation may amount to only a partial assignment of meaning. Thus, e.g., the rules for the measurement of length by means of a standard rod may be considered as providing a *partial* empirical interpretation for the term 'the length, in centimeters, of interval *i*', or alternatively, for some sentences of the form 'the length of interval *i* is *r* centimeters'. For the method is applicable only to intervals of a certain medium size, and even for the latter it does not constitute a full interpretation since the use of a standard rod does not constitute the only way of determining length: various alternative procedures are available involving the measurement of other magnitudes which are connected, by general laws, with the length that is to be determined.

This last observation, concerning the possibility of an indirect measurement of length by virtue of certain laws, suggests an important reminder. It is not correct to speak, as is often done, of 'the experiential meaning' of a term or a sentence in isolation. In the language of science, and for similar reasons even in pre-scientific discourse, a single statement usually has no experiential implications. A single sentence in a scientific theory does not, as a rule, entail any observation sentences; consequences asserting the occurrence of certain observable phenomena can be derived from it only by conjoining it with a set of other, subsidiary, hypotheses. Of the latter, some will usually be observation sentences, others will be previously accepted theoretical statements. Thus, e.g., the relativistic theory of the deflection of light rays in the gravitational field of the sun entails asser-

tions about observable phenomena only if it is conjoined with a considerable body of astronomical and optical theory as well as a large number of specific statements about the instruments used in those observations of solar eclipses which serve to test the hypothesis in question.

Hence, the phrase, 'the experiential meaning of expression E' is elliptical: What a given expression 'means' in regard to potential empirical data is relative to two factors, namely:

(I) *the linguistic framework* L to which the expression belongs. Its rules determine, in particular, what sentences – observational or otherwise – may be inferred from a given statement or class of statements;

(II) the theoretical context in which the expression occurs, i.e., the class of those statements in L which are available as subsidiary hypotheses.

Thus, the sentence formulating Newton's law of gravitation has no experiential meaning by itself; but when used in a language whose logical apparatus permits the development of the calculus, and when combined with a suitable system of other hypotheses – including sentences which connect some of the theoretical terms with observation terms and thus establish a partial interpretation – then it has a bearing on observable phenomena in a large variety of fields. Analogous considerations are applicable to the term 'gravitational field', for example. It can be considered as having experiential meaning only within the context of a theory, which must be at least partially interpreted; and the experiential meaning of the term – as expressed, say, in the form of operational criteria for its application – will depend again on the theoretical system at hand, and on the logical characteristics of the language within which it is formulated.

4. Cognitive significance as a characteristic of interpreted systems

The preceding considerations point to the conclusion that a satisfactory criterion of cognitive significance cannot be reached through the second avenue of approach here considered, namely by means of specific requirements for the terms which make up significant sentences. This result accords with a general characteristic of scientific (and, in principle, even

pre-scientific) theorizing: Theory formation and concept formation go hand in hand; neither can be carried on successfully in isolation from the other.

If, therefore, cognitive significance can be attributed to anything, then only to entire theoretical systems formulated in a language with a well-determined structure. And the decisive mark of cognitive significance in such a system appears to be the existence of an interpretation for it in terms of observables. Such an interpretation might be formulated, for example, by means of conditional or biconditional sentences connecting nonobservational terms of the system with observation terms in the given language; the latter as well as the connecting sentences may or may not belong to the theoretical system.

But the requirement of partial interpretation is extremely liberal; it is satisfied, for example, by the system consisting of contemporary physical theory combined with some set of principles of speculative metaphysics, even if the latter have no empirical interpretation at all. Within the total system, these metaphysical principles play the role of what K. Reach and also O. Neurath liked to call *isolated sentences*: They are neither purely formal truths or falsehoods, demonstrable or refutable by means of the logical rules of the given language system; nor do they have any experiential bearing; i.e., their omission from the theoretical system would have no effect on its explanatory and predictive power in regard to potentially observable phenomena (i.e., the kind of phenomena described by observation sentences). Should we not, therefore, require that a cognitively significant system contain no isolated sentences? The following criterion suggests itself:

(4.1.) A theoretical system is cognitively significant if and only if it is partially interpreted to at least such an extent that none of its primitive sentences is isolated.

But this requirement may bar from a theoretical system certain sentences which might well be viewed as permissible and indeed desirable. By way of a simple illustration, let us assume that our theoretical system T contains the primitive sentence

$(S1)$ $(x)\,[P_1x \supset (Qx \equiv P_2x)]$,

where 'P_1' and 'P_2' are observation predicates in the given language L, while 'Q' functions in T somewhat in the manner of a theoretical construct and occurs in only one primitive sentence of T, namely ($S1$). Now ($S1$) is not a truth or falsehood of formal logic; and furthermore, if ($S1$) is omitted from the set of primitive sentences of T, then the resulting system, T', possesses exactly the same systematic, i.e., explanatory and predictive, power as T. Our contemplated criterion would therefore qualify ($S1$) as an isolated sentence which has to be eliminated – excised by means of Occam's razor, as it were – if the theoretical system at hand is to be cognitively significant.

But it is possible to take a much more liberal view of ($S1$) by treating it as a partial definition for the theoretical term 'Q'. Thus conceived, ($S1$) specifies that in all cases where the observable characteristic P_1 is present, 'Q' is applicable if and only if the observable characteristic P_2 is present as well. In fact, ($S1$) is an instance of those partial, or conditional, definitions which Carnap calls bilateral reduction sentences. These sentences are explicitly qualified by Carnap as analytic (though not, of course, as truths of formal logic), essentially on the ground that all their consequences which are expressible by means of observation predicates (and logical terms) alone are truths of formal logic.[18]

Let us pursue this line of thought a little further. This will lead us to some observations on analytic sentences and then back to the question of the adequacy of (4.1.).

Suppose that we add to our system T the further sentence

$$(S2) \qquad (x)\,[P_3x \supset (Qx \equiv P_4x)],$$

where 'P_3', 'P_4' are additional observation predicates. Then, on the view that "every bilateral reduction sentence is analytic",[19] ($S2$) would be analytic as well as ($S1$). Yet, the two sentences jointly entail non-analytic consequences which are expressible in terms of observation predicates alone, such as[20]

$$(Q) \qquad (x)\,[\sim(P_1x \cdot P_2x \cdot Px_3 \cdot \sim P_4x) \cdot \sim(P_1x \cdot \sim P_2x \cdot P_3x \cdot P_4x)].$$

But one would hardly want to admit the consequence that the conjunction of two analytic sentences may be synthetic. Hence if the concept of analyticity can be applied at all to the sentences of interpreted de-

ductive systems, then it will have to be relativized with respect to the theoretical context at hand. Thus, e.g., ($S1$) might be qualified as analytic relative to the system T, whose remaining postulates do not contain the term 'Q', but as synthetic relative to the system T enriched by ($S2$). Strictly speaking, the concept of analyticity has to be relativized also in regard to the rules of the language at hand, for the latter determine what observational or other consequences are entailed by a given sentence. This need for at least a twofold relativization of the concept of analyticity was almost to be expected in view of those considerations which required the same twofold relativization for the concept of experiential meaning of a sentence.

If, on the other hand, we decide not to permit ($S1$) in the role of a partial definition and instead reject it as an isolated sentence, then we are led to an analogous conclusion: Whether a sentence is isolated or not will depend on the linguistic frame and on the theoretical context at hand: While ($S1$) is isolated relative to T (and the language in which both are formulated), it acquires definite experiential implications when T is enlarged by ($S2$).

Thus we find, on the level of interpreted theoretical systems, a peculiar rapprochement, and partial fusion, of some of the problems pertaining to the concepts of cognitive significance and of analyticity: Both concepts need to be relativized; and a large class of sentences may be viewed, apparently with equal right, as analytic in a given context, or as isolated, or nonsignificant, in respect to it.

In addition to barring, as isolated in a given context, certain sentences which could just as well be construed as partial definitions, the criterion (4.1.) has another serious defect. Of two logically equivalent formulations of a theoretical system it may qualify one as significant while barring the other as containing an isolated sentence among its primitives. For assume that a certain theoretical system $T1$ contains among its primitive sentences S', S'',... exactly one, S', which is isolated. Then $T1$ is not significant under (4.1.). But now consider the theoretical system $T2$ obtained from $T1$ by replacing the two first primitive sentences, S', S'', by one, namely their conjunction. Then, under our assumptions, none of the primitive sentences of $T2$ is isolated, and $T2$, though equivalent to $T1$, is qualified as significant by (4.1.). In order to do justice to the intent of (4.1.), we would therefore have to lay down the following stricter requirement:

(4.2.) A theoretical system is cognitively significant if and only if it is partially interpreted to such an extent that in no system equivalent to it at least one primitive sentence is isolated.

Let us apply this requirement to some theoretical system whose postulates include the two sentences ($S1$) and ($S2$) considered before, and whose other postulates do not contain 'Q' at all. Since the sentences ($S1$) and ($S2$) together entail the sentence O, the set consisting of ($S1$) and ($S2$) is logically equivalent to the set consisting of ($S1$), ($S2$) and O. Hence, if we replace the former set by the latter, we obtain a theoretical system equivalent to the given one. In this new system, both ($S1$) and ($S2$) are isolated since, as can be shown, their removal does not affect the explanatory and predictive power of the system in reference to observable phenomena. To put it intuitively, the systematic power of ($S1$) and ($S2$) is the same as that of O. Hence, the original system is disqualified by (4.2.). From the viewpoint of a strictly sensationalist positivism as perhaps envisaged by Mach, this result might be hailed as a sound repudiation of theories making reference to fictitious entities, and as a strict insistence on theories couched exclusively in terms of observables. But from a contemporary vantage point, we shall have to say that such a procedure overlooks or misjudges the important function of constructs in scientific theory: The history of scientific endeavor shows that if we wish to arrive at precise, comprehensive, and well-confirmed general laws, we have to rise above the level of direct observation. The phenomena directly accessible to our experience are not connected by general laws of great scope and rigor. Theoretical constructs are needed for the formulation of such higher-level laws. One of the most important functions of a well-chosen construct is its potential ability to serve as a constituent in ever new general connections that may be discovered; and to such connections we would blind ourselves if we insisted on banning from scientific theories all those terms and sentences which could be 'dispensed with' in the sense indicated in (4.2.). In following such a narrowly phenomenalistic or positivistic course, we would deprive ourselves of the tremendous fertility of theoretical constructs, and we would often render the formal structure of the expurgated theory clumsy and inefficient.

Criterion (4.2.), then, must be abandoned, and considerations such as those outlined in this paper seem to lend strong support to the conjecture

that no adequate alternative to it can be found; i.e., that it is not possible to formulate general and precise criteria which would separate those partially interpreted systems whose isolated sentences might be said to have a significant function from those in which the isolated sentences are, so to speak, mere useless appendages.

We concluded earlier that cognitive significance in the sense intended by recent empiricism and operationism can at best be attributed to sentences forming a theoretical system, and perhaps rather to such systems as wholes. Now, rather than try to replace (4.2.) by some alternative, we will have to recognize further that cognitive significance in a system is a matter of degree: Significant systems range from those whose entire extralogical vocabulary consists of observation terms, through theories whose formulation relies heavily on theoretical constructs, on to systems with hardly any bearing on potential empirical findings. Instead of dichotomizing this array into significant and non-significant systems it would seem less arbitrary and more promising to appraise or compare different theoretical systems in regard to such characteristics as these:

(a) the clarity and precision with which the theories are formulated, and with which the logical relationships of their elements to each other and to expressions couched in observational terms have been made explicit;

(b) the systematic, i.e., explanatory and predictive, power of the systems in regard to observable phenomena;

(c) the formal simplicity of the theoretical system with which a certain systematic power is attained;

(d) the extent to which the theories have been confirmed by experiential evidence.

Many of the speculative philosophical approaches to cosmology, biology, or history, for example, would make a poor showing on practically all of these counts and would thus prove no matches to available rival theories, or would be recognized as so unpromising as not to warrant further study or development.

If the procedure here suggested is to be carried out in detail, so as to become applicable also in less obvious cases, then it will be necessary, of course, to develop general standards, and theories pertaining to them, for the appraisal and comparison of theoretical systems in the various respects just mentioned. To what extent this can be done with rigor and precision cannot well be judged in advance. In recent years, a considerable amount

of work has been done towards a definition and theory of the concept of degree of confirmation, or logical probability, of a theoretical system;[21] and several contributions have been made towards the clarification of some of the other ideas referred to above.[22] The continuation of this research represents a challenge for further constructive work in the logical and methodological analysis of scientific knowledge.

NOTES

** This essay combines, with certain omissions and some other changes, the contents of two articles: 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie* (1950), 41–63; and 'The Concept of Cognitive Significance: A Reconsideration', *Proceedings of the American Academy of Arts and Sciences* **80** (1951), 61–77. This material is reprinted with kind permission of the Director of *Revue Internationale de Philosophie* and of the American Academy of Arts and Sciences.

[1] Observation sentences of this kind belong to what Carnap has called the thing-language, cf., e.g. (1938), pp. 52–53. That they are adequate to formulate the data which serve as the basis for empirical tests is clear in particular for the intersubjective testing procedures used in science as well as in large areas of empirical inquiry on the common-sense level. In epistemological discussions, it is frequently assumed that the ultimate evidence for beliefs about empirical matters consists in perceptions and sensations whose description calls for a phenomenalistic type of language. The specific problems connected with the phenomenalistic approach cannot be discussed here; but it should be mentioned that at any rate all the critical considerations presented in this article in regard to the testability criterion are applicable, *mutatis mutandis*, to the case of a phenomenalistic basis as well.

[2] Originally, the permissible evidence was meant to be restricted to what is observable by the speaker and perhaps his fellow beings during their life times. Thus construed, the criterion rules out, as cognitively meaningless, all statements about the distant future or the remote past, as has been pointed out, among others, by Ayer (1946), Chapter I; by Pap (1949), Chapter 13, esp. pp. 333ff.; and by Russell (1948), pp. 445–47. This difficulty is avoided, however, if we permit the evidence to consist of any finite set of "logically possible observation data", each of them formulated in an observation sentence. Thus, e.g., the sentence S_1, "The tongue of the largest dinosaur in New York's Museum of Natural History was blue or black" is completely verifiable in our sense; for it is a logical consequence of the sentence S_2, "The tongue of the largest dinosaur in New York's Museum of Natural History was blue"; and this is an observation sentence, in the sense just indicated.

And if the concept of *verifiability in principle* and the more general concept of *confirmability in principle*, which will be considered later, are construed as referring to *logically possible evidence* as expressed by observation sentences, then it follows similarly that the class of statements which are verifiable, or at least confirmable, in principle include such assertions as that the planet Neptune and the Antarctic Continent existed before they were discovered, and that atomic warfare, if not checked, will lead to the extermination of this planet. The objections which Russell (1948), pp. 445 and

447, raises against the verifiability criterion by reference to those examples do not apply therefore if the criterion is understood in the manner here suggested. Incidentally, statements of the kind mentioned by Russell, which are not actually verifiable by any human being, were explicitly recognized as cognitively significant already by Schlick (1936), Part V, who argued that the impossibility of verifying them was 'merely empirical'. The characterization of verifiability with the help of the concept of observation sentence as suggested here might serve as a more explicit and rigorous statement of that conception.

[3] As has frequently been emphasized in the empiricist literature, the term 'verifiability' is to indicate, of course, the conceivability, or better, the logical possibility, of evidence of an observational kind which, if actually encountered, would constitute conclusive evidence for the given sentence; it is not intended to mean the technical possibility of performing the tests needed to obtain such evidence, and even less the possibility of actually finding directly observable phenomena which constitute conclusive evidence for that sentence – which would be tantamount to the actual existence of such evidence and would thus imply the truth of the given sentence. Analogous remarks apply to the terms 'falsifiability' and 'confirmability'. This point has clearly been disregarded in some critical discussions of the verifiability criterion. Thus, e.g., Russell (1948), p. 448 construes verifiability as the actual existence of a set of conclusively verifiying occurrences. This conception, which has never been advocated by any logical empiricist, must naturally turn out to be inadequate since according to it the empirical meaningfulness of a sentence could not be established without gathering empirical evidence, and moreover enough of it to permit a conclusive proof of the sentence in question! It is not surprising, therefore, that his extraordinary interpretation of verifiability leads Russell to the conclusion: "In fact, that a proposition is verifiable is itself not verifiable" (*l.c.*). Actually, under the empiricist interpretation of complete verifiability, any statement asserting the verifiability of some sentence *S* whose text is quoted, is either analytic or contradictory; for the decision whether there exists a class of observation sentences which entail *S*, i.e., whether such observation sentences can be formulated, no matter whether they are true or false – that decision is a purely logical matter.

[4] The arguments here adduced against the verifiability criterion also prove the inadequacy of a view closely related to it, namely that two sentences have the same cognitive significance if any set of observation sentences which would verify one of them would also verify the other, and conversely. Thus, e.g., under this criterion, any two general laws would have to be assigned the same cognitive significance, for no general law is verified by any set of observation sentences. The view just referred to must be clearly distinguished from a position which Russell examines in his critical discussion of the positivistic meaning criterion. It is "the theory that two propositions whose verified consequences are identical have the same significance" (1948), p. 448. This view is untenable indeed, for what consequences of a statement have actually been verified at a given time is obviously a matter of historical accident which cannot possibly serve to establish identity of cognitive significance. But I am not aware that any logical empiricist ever subscribed to that 'theory'.

[5] (1936, 1946), Chapter I. The case against the requirements of verifiability and of falsifiability, and in favor of a requirement of partial confirmability and disconfirmability, is very clearly presented also by Pap (1949), Chapter 13.

[6] (1946), 2nd ed., pp. 11–12.

[7] This restriction is expressed in recursive form and involves no vicious circle. For the full statement of Ayer's criterion, see Ayer (1946), p. 13.

[8] Church (1949). An alternative criterion recently suggested by O'Connor (1950) as a revision of Ayer's formulations is subject to a slight variant of Church's stricture: It can be shown that if there are three observation sentences none of which entails any of the others, and if S is any noncompound sentence, then either S or $\sim S$ is significant under O'Connor's criterion.

[9] An extralogical term is one that does not belong to the specific vocabulary of logic. The following phrases, and those definable by means of them, are typical examples of logical terms: 'not', 'or', 'if... then', 'all', 'some', '... is an element of class...'. Whether it is possible to make a sharp theoretical distinction between logical and extra-logical terms is a controversial issue related to the problem of discriminating between analytic and synthetic sentences. For the purpose at hand, we may simply assume that the logical vocabulary is given by enumeration.

[10] For a detailed exposition and critical discussion of this idea, see H. Feigl's stimulating and enlightening article (1950).

[11] Cf. (1936–37), especially Section 7.

[12] On this subject, see for example Langford (1941); Lewis (1946), pp. 210–30; Chisholm (1946); Goodman (1947); Reichenbach (1947), Chapter VIII; Hempel and Oppenheim (1948), Part III; Popper (1949); and especially Goodman's further analysis (1955).

[13] Cf. Carnap, *loc. cit.* Note 11. For a brief elementary presentation of the main idea, see Carnap (1938), Part III. The sentence R here formulated for the predicate 'F' illustrates only the simplest type of reduction sentence, the so-called bilateral reduction sentence.

[14] Cf. the analysis in Carnap (1936–37), especially Section 15; also see the briefer presentation of the liberalized point of view in Carnap (1938).

[15] (Added in 1964.) This is not strictly correct. For a more circumspect statement, see Note 12 in 'A Logical Appraisal of Operationism' and the fuller discussion in Section 7 of the essay 'The Theoretician's Dilemma'. Both of these pieces are reprinted in *Aspects of Scientific Explanation.*

[16] The interpretation of formal theories has been studied extensively by Reichenbach, especially in his pioneer analyses of space and time in classical and in relativistic physics. He describes such interpretation as the establishment of *coordinating definitions* (Zuordnungsdefinitionen) for certain terms of the formal theory. See, for example, Reichenbach (1928). More recently, Northrop [cf.(1947), Chapters VII, and also the detailed study of the use of deductively formulated theories in science, *ibid.*, Chapters IV, V, VI] and H. Margenau [cf., for example (1935)] have discussed certain aspects of this process under the title of *epistemic correlation.*

[17] A somewhat fuller account of this type of interpretation may be found in Carnap (1939), Section 24. The articles by Spence (1944) and by MacCorquodale and Meehl (1948) provide enlightening illustrations of the use of theoretical constructs in a field outside that of the physical sciences, and of the difficulties encountered in an attempt to analyze in detail their function and interpretation.

[18] Cf. Carnap (1936–37), especially Sections 8 and 10.

[19] Carnap (1936–37), p. 452.

[20] The sentence O is what Carnap calls the *representative sentence* of the couple consisting of the sentences ($S1$) and ($S2$); see (1936–37), pp. 450–53.

[21] Cf., for example, Carnap (1945)1 and (1945)2, and especially (1950). Also see Helmer and Oppenheim (1945).

[22] On simplicity, cf. especially Popper (1935), Chapter V; Reichenbach (1938), Section 42; Goodman (1949)1, (1949)2, (1950); on explanatory and predictive power, cf. Hempel and Oppenheim (1948), Part IV.

BIBLIOGRAPHY

Ayer, A. J.: *Language, Truth and Logic*, London, 1936; 2nd ed. 1946.

Carnap, R.: 'Testability and Meaning', *Philosophy of Science* **3** (1936) and **4** (1937).

Carnap, R.: 'Logical Foundations of the Unity of Science', in *International Encyclopedia of Unified Science* **I**, 1; Chicago, 1938.

Carnap, R.: *Foundations of Logic and Mathematics*, Chicago, 1939.

Carnap, R.: 'On Inductive Logic', *Philosophy of Science* **12** (1945). Referred to as (1945) 1 in this article.

Carnap, R.: 'The Two Concepts of Probability', *Philosophy and Phenomenological Research* **5** (1945). Referred to as (1945) 2 in this article.

Carnap, R.: *Logical Foundations of Probability*, Chicago, 1950.

Chisholm, R. M.: 'The Contrary-to-Fact Conditional', *Mind* **55** (1946).

Church, A.: 'Review of Ayer (1946)', *The Journal of Symbolic Logic* **14** (1949), 52–53.

Feigl, H.: 'Existential Hypotheses: Realistic vs. Phenomenalistic Interpretations', *Philosophy of Science* **17** (1950).

Goodman, N.: 'The Problem of Counterfactual Conditionals', *The Journal of Philosophy* **44** (1947).

Goodman, N.: 'The Logical Simplicity of Predicates', *The Journal of Symbolic Logic* **14** (1949). Referred to as (1949) 1 in this article.

Goodman, N.: 'Some Reflections on the Theory of Systems', *Philosophy and Phenomenological Research* **9** (1949). Referred to as (1949) 2 in this article.

Goodman, N.: 'An Improvement in the Theory of Simplicity', *The Journal of Symbolic Logic* **15** (1950).

Goodman, N.: *Fact, Fiction, and Forecast*, Cambridge, Massachusetts, 1955.

Helmer, O. and Oppenheim, P.: 'A Syntactical Definition of Probability and of Degree of Confirmation', *The Journal of Symbolic Logic* **10** (1945).

Hempel, C. G. and Oppenheim, P.: 'Studies in the Logic of Explanation', *Philosophy of Science* **15** (1948).

Langford, C. H.: Review in *The Journal of Symbolic Logic* **6** (1941), 67–68.

Lewis, C. I.: *An Analysis of Knowledge and Valuation*, La Salle, Ill., 1946.

MacCorquodale, K. and Meehl, P. E.: 'On a Distinction Between Hypothetical Constructs and Intervening Variables', *Psychological Review* **55** (1948).

Margenau, H.: 'Methodology of Modern Physics', *Philosophy of Science* **2** (1935).

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

O'Connor, D. J.: 'Some Consequences of Professor A. J. Ayer's Verification Principle', *Analysis* **10** (1950).

Pap, A.: *Elements of Analytic Philosophy*, New York, 1949.

Popper, K.: *Logik der Forschung*, Wien, 1935.

Popper, K.: 'A Note on Natural Laws and So-Called "Contrary-to-Fact Conditionals"', *Mind* **58** (1949).

Reichenbach, H.: *Philosophie der Raum-Zeit-Lehre*, Berlin, 1928.

Reichenbach, H.: *Elements of Symbolic Logic*, New York, 1947.

Russell, B.: *Human Knowledge*, New York, 1948.

Schlick, M.: 'Meaning and Verification', *Philosophical Review* **45** (1936). Also reprinted in Feigl, H. and W. Sellars (eds.), *Readings in Philosophical Analysis*, New York, 1949.

Spence, Kenneth W.: 'The Nature of Theory Construction in Contemporary Psychology', *Psychological Review* **51** (1944).

POSTSCRIPT (1964) ON

COGNITIVE SIGNIFICANCE

The preceding essay is a conflation of two articles: 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie*, No. 11 (1950), and 'The Concept of Cognitive Significance: A Reconsideration', *Proceedings of the American Academy of Arts and Sciences* **80** (1951). In combining the two, I omitted particularly some parts of the first article, which had been largely superseded by the second one;[1] I also made a few minor changes in the remaining text. Some of the general problems raised in the combined essay are pursued further in *Aspects of Scientific Explanation*, especially in 'The Theoretician's Dilemma'. In this Postscript, I propose simply to note some second thoughts concerning particular points in the preceding essay.

(i) The objections 2.1(c) and 2.2(c) against the requirements of complete verifiability and of complete falsifiability are, I think, of questionable force. For $S \vee N$ can properly be said to be entailed by S, and S in turn by $S \cdot N$, only if N as well as S is a declarative sentence and thus is either true or false. But if the criterion of cognitive significance is understood to delimit the class of sentences which make significant assertions, and which are thus either true or false, then the sentence N invoked in the objections is not declarative, and neither are $S \vee N$ or $S \cdot N$; hence the alleged inferences from $S \cdot N$ to S and from S to $S \vee N$ are inadmissible.[2]

My objection retains its force, however, against the use of falsifiability, not as a criterion of significance, but as a "criterion of demarcation". This use would draw a dividing line "between the statements, or systems of statements, of the empirical sciences, and all other statements – whether they are of a religious or of a metaphysical character, or simply pseudo-scientific".[3] For the argument 2.2(c) shows that the conjunction of a scientific statement S with a nonscientific statement N is falsifiable and thus qualifies as a scientific statement; and this would defeat the intended purpose of the criterion of demarcation.

(ii) My assertion, in 2.1(a) and 2.2(a), that the requirements of verifiability and of falsifiability would rule out *all* hypotheses of mixed

quantificational form is false. Consider the hypothesis 'All ravens are black and something is white', or, in symbolic notation

$$(x)\,(Rx \supset Bx)\cdot(\exists y)\;Wy,$$

which is equivalent to

$$(x)\,(\exists y)\;[(Rx \supset Bx)\cdot Wy].$$

This sentence satisfies the falsifiability requirement because it implies the purely universal hypothesis '$(x)\,(Rx \supset Bx)$', which would be falsified, for example, by the following set of observation sentences: {'Ra', '$\sim Ba$'}. Similarly, the sentence

$$(\exists x)\,(y)\,(Rx \vee Wy)$$

is verifiable since it is implied, for example, by 'Ra'.

The essential point of the objection remains unaffected, however: Many scientific hypotheses of mixed quantificational form are neither verifiable nor falsifiable; these would therefore be disqualified by the requirement of verifiability as well as by that of falsifiability; and if the latter is used as a criterion of demarcation rather than of significance, it excludes those hypotheses from the class of scientific statements. These consequences are unacceptable.

(iii) An even stronger criticism of the criteria of verifiability and of falsifiability results from condition (A1), which is stated early in Section 2, and which demands in effect that any acceptable criterion of significance which admits a sentence as significant must also admit its negation. That this condition must be met is clear, for since a significant sentence is one that is either true or false, its negation can be held nonsignificant only on pain of violating a fundamental principle of logic. And even if the falsifiability criterion is used as a criterion of demarcation rather than of cognitive significance, satisfaction of (A1) seems imperative. Otherwise, a scientist reporting that he had succeeded in refuting a scientific hypothesis S of universal form would be making a nonscientific statement if he were to say: "Hence, it is not the case that S holds", for this statement would not be falsifiable. More generally, formally valid deductive logical inference would often lead from scientific premises to nonscientific conclusions – e.g., from '$Ra\cdot \sim Ba$' to '$(\exists x)\,(Rx\cdot \sim Bx)$'; and, surely, this is intolerable.

But when the requirement of verifiability, or that of falsifiability, is combined with condition (A1), then a sentence qualifies as cognitively significant just in case it and its negation are verifiable, or just in case it and its negation are falsifiable. These two criteria now demand the same thing of a significant sentence, namely, that it be both verifiable and falsifiable. This characterization admits, besides all truth-functional compounds of observation sentences, also certain sentences containing quantifiers. For example, '$Pa \vee (x)\, Qx$' is verifiable by 'Pa' and falsifiable by {'$\sim Pa$', '$\sim Qb$'}; and as is readily seen, '$Pa \cdot (\exists x)\, Qx$' equally meets the combined requirement. But this requirement excludes all strictly general hypotheses, i.e., those containing essential occurrences of quantifiers but not of individual constants; such as '$(x)\,(Rx \supset Bx)$', '$(x)\,(\exists y)\,(Rxy \supset Sxy)$', and so forth. Again, this consequence is surely unacceptable, no matter whether the criterion is meant to delimit the class of significant sentences or the class of statements of empirical science.

NOTES

[1] The basic ideas presented in the earlier articles and in the present conflated version are penetratingly examined by I. Scheffler in *The Anatomy of Inquiry*, New York, 1963. Part II of his book deals in detail with the concept of cognitive significance.

[2] I owe this correction to graduate students who put forth the above criticism in one of my seminars. The same point has recently been stated very clearly by D. Rynin in 'Vindication of L*G*C*L*P*S*T*V*M' *Proceedings and Addresses of the American Philosophical Association* **30** (1957); see especially pp, 57–58.

[3] K. R. Popper, 'Philosophy of Science: A Personal Report', In C. A. Mace (ed.), *British Philosophy in the Mid-Century*, London, 1957, pp. 155–91; quotations from pp. 163, 162.

KARL R. POPPER

SOME FUNDAMENTAL PROBLEMS IN THE LOGIC OF SCIENTIFIC DISCOVERY*

A scientist, whether theorist or experimenter, puts forward statements, or systems of statements, and tests them step by step. In the field of the empirical sciences, more particularly, he constructs hypotheses, or systems of theories, and tests them against experience by observation and experiment.

I suggest that it is the task of the logic of scientific discovery, or the logic of knowledge, to give a logical analysis of this procedure; that is, to analyse the method of the empirical sciences.

But what are these 'methods of the empirical sciences'? And what do we call 'empirical science'?

1. THE PROBLEM OF INDUCTION

According to a widely accepted view – to be opposed in this book – the empirical sciences can be characterized by the fact that they use '*inductive methods*', as they are called. According to this view, the logic of scientific discovery would be identical with inductive logic, i.e. with the logical analysis of these inductive methods.

It is usual to call an inference 'inductive' if it passes from *singular statements* (sometimes also called 'particular' statements), such as accounts of the results of observations or experiments, to *universal statements*, such as hypotheses or theories.

Now it is far from obvious, from a logical point of view, that we are justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false: no matter how many instances of white swans we may have observed, this does not justify the conclusion that *all* swans are white.

The question whether inductive inferences are justified, or under what conditions, is known as *the problem of induction.*

The problem of induction may also be formulated as the question of how to establish the truth of universal statements which are based

on experience, such as the hypotheses and theoretical systems of the empirical sciences. For many people believe that the truth of these universal statements is '*known by experience*'; yet it is clear that an account of an experience – of an observation or the result of an experiment – can in the first place be only a singular statement and not a universal one. Accordingly, people who say of a universal statement that we know its truth from experience usually mean that the truth of this universal statement can somehow be reduced to the truth of singular ones, and that these singular ones are known by experience to be true; which amounts to saying that the universal statement is based on inductive inference. Thus to ask whether there are natural laws known to be true appears to be only another way of asking whether inductive inferences are logically justified.

Yet if we want to find a way of justifying inductive inferences, we must first of all try to establish a *principle of induction.* A principle of induction would be a statement with the help of which we could put inductive inferences into a logically acceptable form. In the eyes of the upholders of inductive logic, a principle of induction is of supreme importance for scientific method: "... this principle", says Reichenbach,

> determines the truth of scientific theories. To eliminate it from science would mean nothing less than to deprive science of the power to decide the truth or falsity of its theories. Without it, clearly, science would no longer have the right to distinguish its theories from the fanciful and arbitrary creations of the poet's mind.[1]

Now this principle of induction cannot be a purely logical truth like a tautology or an analytic statement. Indeed, if there were such a thing as a purely logical principle of induction, there would be no problem of induction; for in this case, all inductive inferences would have to be regarded as purely logical or tautological transformations, just like inferences in deductive logic. Thus the principle of induction must be a synthetic statement; that is, a statement whose negation is not self-contradictory but logically possible. So the question arises why such a principle should be accepted at all, and how we can justify its acceptance on rational grounds.

Some who believe in inductive logic are anxious to point out, with Reichenbach, that "the principle of induction is unreservedly accepted by the whole of science and that no man can seriously doubt this principle in everyday life either".[2] Yet even supposing this were the case – for

after all, 'the whole of science' might err – I should still contend that a principle of induction is superfluous, and that it must lead to logical inconsistencies.

That inconsistencies may easily arise in connection with the principle of induction should have been clear from the work of Hume;*[1] also, that they can be avoided, if at all, only with difficulty. For the principle of induction must be a universal statement in its turn. Thus if we try to regard its truth as known from experience, then the very same problems which occasioned its introduction will arise all over again. To justify it, we should have to employ inductive inferences; and to justify these we should have to assume an inductive principle of a higher order; and so on. Thus the attempt to base the principle of induction on experience breaks down, since it must lead to an infinite regress.

Kant tried to force his way out of this difficulty by taking the principle of induction (which he formulated as the 'principle of universal causation') to be '*a priori* valid'. But I do not think that his ingenious attempt to provide an *a priori* justification for synthetic statements was successful.

My own view is that the various difficulties of inductive logic here sketched are insurmountable. So also, I fear, are those inherent in the doctrine, so widely current today, that inductive inference, although not 'strictly valid', *can attain some degree of 'reliability' or of 'probability'*. According to this doctrine, inductive inferences are 'probable inferences'.[3] "We have described", says Reichenbach,

> the principle of induction as the means whereby science decides upon truth. To be more exact, we should say that it serves to decide upon probability. For it is not given to science to reach either truth or falsity... but scientific statements can only attain continuous degrees of probability whose unattainable upper and lower limits are truth and falsity.[4]

At this stage I can disregard the fact that the believers in inductive logic entertain an idea of probability that I shall later reject as highly unsuitable for their own purposes (see Section 80, below). I can do so because the difficulties mentioned are not even touched by an appeal to probability. For if a certain degree of probability is to be assigned to statements based on inductive inference, then this will have to be justified by invoking a new principle of induction, appropriately modified. And this new principle in its turn will have to be justified, and so on. Nothing is gained, moreover, if the principle of induction, in its turn, is taken not as

'true' but only as 'probable'. In short, like every other form of inductive logic, the logic of probable inference, or 'probability logic', leads either to an infinite regress, or to the doctrine of *apriorism.*[*2]

The theory to be developed in the following pages stands directly opposed to all attempts to operate with the ideas of inductive logic. It might be described as the theory of *the deductive method of testing*, or as the view that a hypothesis can only be empirically *tested* – and only *after* it has been advanced.

Before I can elaborate this view (which might be called 'deductivism', in contrast to 'inductivism'[5]) I must first make clear the distinction between the *psychology of knowledge* which deals with empirical facts, and the *logic of knowledge* which is concerned only with logical relations. For the belief in inductive logic is largely due to a confusion of psychological problems with epistemological ones. It may be worth noticing, by the way, that this confusion spells trouble not only for the logic of knowledge but for its psychology as well.

2. Elimination of psychologism

I said above that the work of the scientist consists in putting forward and testing theories.

The initial stage, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor to be susceptible of it. The question how it happens that a new idea occurs to a man – whether it is a musical theme, a dramatic conflict, or a scientific theory – may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. This latter is concerned not with *questions of fact* (Kant's *quid facti?*), but only with questions of *justification or validity* (Kant's *quid juris?*). Its questions are of the following kind. Can a statement be justified? And if so, how? Is it testable? Is it logically dependent on certain other statements? Or does it perhaps contradict them? In order that a statement may be logically examined in this way, it must already have been presented to us. Someone must have formulated it, and submitted it to logical examination.

Accordingly I shall distinguish sharply between the process of conceiving a new idea, and the methods and results of examining it logically. As to the task of the logic of knowledge – in contradistinction to the

psychology of knowledge – I shall proceed on the assumption that it consists solely in investigating the methods employed in those systematic tests to which every new idea must be subjected if it is to be seriously entertained.

Some might object that it would be more to the purpose to regard it as the business of epistemology to produce what has been called a '*rational reconstruction*' of the steps that have led the scientist to a discovery – to the finding of some new truth. But the question is: what, precisely, do we want to reconstruct? If it is the processes involved in the stimulation and release of an inspiration which are to be reconstructed, then I should refuse to take it as the task of the logic of knowledge. Such processes are the concern of empirical psychology but hardly of logic. It is another matter if we want to reconstruct rationally the *subsequent tests* whereby the inspiration may be discovered to be a discovery, or become known to be knowledge. In so far as the scientist critically judges, alters, or rejects his own inspiration we may, if we like, regard the methodological analysis undertaken here as a kind of 'rational reconstruction' of the corresponding thought-processes. But this reconstruction would not describe these processes as they actually happen: it can give only a logical skeleton of the procedure of testing. Still, this is perhaps all that is meant by those who speak of a 'rational reconstruction' of the ways in which we gain knowledge.

It so happens that my arguments in this book are quite independent of this problem. However, my view of the matter, for what it is worth, is that there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every discovery contains 'an irrational element', or 'a creative intuition', in Bergson's sense. In a similar way Einstein speaks of "...the search for those highly universal...laws from which a picture of the world can be obtained by pure deduction. There is no logical path", he says, "leading to these...laws. They can only be reached by intuition, based upon something like an intellectual love ('*Einfühlung*') of the objects of experience".[1]

3. Deductive testing of theories

According to the view that will be put forward here, the method of

critically testing theories, and selecting them according to the results of tests, always proceeds on the following lines. From a new idea, put up tentatively, and not yet justified in any way – an anticipation, a hypothesis, a theoretical system, or what you will – conclusions are drawn by means of logical deduction. These conclusions are then compared with one another and with other relevant statements, so as to find what logical relations (such as equivalence, derivability, compatibility, or incompatibility) exist between them.

We may if we like distinguish four different lines along which the testing of a theory could be carried out. First there is the logical comparison of the conclusions among themselves, by which the internal consistency of the system is tested. Secondly, there is the investigation of the logical form of the theory, with the object of determing whether it has the character of an empirical or scientific theory, or whether it is, for example, tautological. Thirdly, there is the comparison with other theories, chiefly with the aim of determining whether the theory would constitute a scientific advance should it survive our various tests. And finally, there is the testing of the theory by way of empirical applications of the conclusions which can be derived from it.

The purpose of this last kind of test is to find out how far the new consequences of the theory – whatever may be new in what it asserts – stand up to the demands of practice, whether raised by purely scientific experiments, or by practical technological applications. Here too the procedure of testing turns out to be deductive. With the help of other statements, previously accepted, certain singular statements – which we may call 'predictions' – are deduced from the theory; especially predictions that are easily testable or applicable. From among these statements, those are selected which are not derivable from the current theory, and more especially those which the current theory contradicts. Next we seek a decision as regards these (and other) derived statements by comparing them with the results of practical applications and experiments. If this decision is positive, that is, if the singular conclusions turn out to be acceptable, or *verified*, then the theory has, for the time being, passed its test: we have found no reason to discard it. But if the decision is negative, or in other words, if the conclusions have been *falsified*, then their falsification also falsifies the theory from which they were logically deduced.

It should be noticed that a positive decision can only temporarily support the theory, for subsequent negative decisions may always overthrow it. So long as a theory withstands detailed and severe tests and is not superseded by another theory in the course of scientific progress, we may say that it has 'proved its mettle' or that it is '*corroborated*'.*[1]

Nothing resembling inductive logic appears in the procedure here outlined. I never assume that we can argue from the truth of singular statements to the truth of theories. I never assume that by force of 'verified' conclusions, theories can be established as 'true', or even as merely 'probable'.

In this book I intend to give a more detailed analysis of the methods of deductive testing. And I shall attempt to show that, within the framework of this analysis, all the problems can be dealt with that are usually called '*epistemological*'. Those problems, more especially, to which inductive logic gives rise, can be eliminated without creating new ones in their place.

4. The problem of demarcation

Of the many objections which are likely to be raised against the view here advanced, the most serious is perhaps the following. In rejecting the method of induction, it may be said, I deprive empirical science of what appears to be its most important characteristic; and this means that I remove the barriers which separate science from metaphysical speculation. My reply to this objection is that my main reason for rejecting inductive logic is precisely that *it does not provide a suitable distinguishing mark* of the empirical, non-metaphysical, character of a theoretical system; or in other words, that *it does not provide a suitable 'criterion of demarcation'*.

The problem of finding a criterion which would enable us to distinguish between the empirical sciences on the one hand, and mathematics and logic as well as 'metaphysical' systems on the other, I call the *problem of demarcation*.[1]

This problem was known to Hume who attempted to solve it.[2] With Kant it became the central problem of the theory of knowledge. If, following Kant, we call the problem of induction 'Hume's problem', we might call the problem of demarcation 'Kant's problem'.

Of these two problems – the source of nearly all the other problems of the theory of knowledge – the problem of demarcation is, I think, the more fundamental. Indeed, the main reason why epistemologists with empiricist leanings tend to pin their faith to the 'method of induction' seems to be their belief that this method alone can provide a suitable criterion of demarcation. This applies especially to those empricists who follow the flag of 'positivism'.

The older positivists wished to admit, as scientific or legitimate, only those *concepts* (or notions or ideas) which were, as they put it, 'derived from experience'; those concepts, that is, which they believed to be logically reducible to elements of sense-experience, such as sensations (or sense-data), impressions, perceptions, visual or auditory memories, and so forth. Modern positivists are apt to see more clearly that science is not a system of concepts but rather a system of *statements*.*[1] Accordingly, they wish to admit, as scientific or legitimate, only those statements which are reducible to elementary (or 'atomic') statements of experience – to 'judgments of perception' or 'atomic propositions' or 'protocol-sentences' or what not.*[2] It is clear that the implied criterion of demarcation is identical with the demand for an inductive logic.

Since I reject inductive logic I must also reject all these attempts to solve the problem of demarcation. With this rejection, the problem of demarcation gains in importance for the present inquiry. Finding an acceptable criterion of demarcation must be a crucial task for any epistemology which does not accept inductive logic.

Positivists usually interpret the problem of demarcation in a *naturalistic* way; they interpret it as if it were a problem of natural science. Instead of taking it as their task to propose a suitable convention, they believe they have to discover a difference, existing in the nature of things, as it were, between empirical science on the one hand and metaphysics on the other. They are constantly trying to prove that metaphysics by its very nature is nothing but nonsensical twaddle – 'sophistry and illusion', as Hume says, which we should 'commit to the flames'.*[3]

If by the words 'nonsensical' or 'meaningless' we wish to express no more, by definition, than 'not belonging to empirical science', then the characterization of metaphysics as meaningless nonsense would be trivial; for metaphysics has usually been defined as non-empirical. But of course, the positivists believe they can say much more about meta-

physics than that some of its statements are non-empirical. The words 'meaningless' or 'nonsensical' convey, and are meant to convey, a derogatory evaluation; and there is no doubt that what the positivists really want to achieve is not so much a successful demarcation as the final overthrow[3] and the annihilation of metaphysics. However this may be, we find that each time the positivists tried to say more clearly what 'meaningful' meant, the attempt led to the same result – to a definition of 'meaningful sentence' (in contradistinction to 'meaningless pseudo-sentence') which simply reiterated the criterion of demarcation of their *inductive logic.*

This 'shows itself' very clearly in the case of Wittgenstein, according to whom every meaningful proposition must be *logically reducible*[4] to elementary (or atomic) propositions, which he characterizes as descriptions or 'pictures of reality'[5] (a characterization, by the way, which is to cover all meaningful propositions). We may see from this that Wittgenstein's criterion of meaningfulness coincides with the inductivists' criterion of demarcation, provided we replace their words 'scientific' or 'legitimate' by 'meaningful'. And it is precisely over the problem of induction that this attempt to solve the problem of demarcation comes to grief: positivists, in their anxiety to annihilate metaphysics, annihilate natural science along with it. For scientific laws, too, cannot be logically reduced to elementary statements of experience. If consistently applied, Wittgenstein's criterion of meaningfulness rejects as meaningless those natural laws the search for which, as Einstein says,[6] is "the supreme task of the physicist": they can never be accepted as genuine or legitimate statements. This view, which tries to unmask the problem of induction as an empty pseudoproblem, has been expressed by Schlick*[4] in the following words:

> The problem of induction consists in asking for a logical justification of *universal statements* about reality... We recognize, with Hume, that there is no such logical justification: there can be none, simply because *they are not genuine* statements.[7]

This shows how the inductivist criterion of demarcation fails to draw a dividing line between scientific and metaphysical systems, and why it must accord them equal status; for the verdict of the positivist dogma of meaning is that both are systems of meaningless pseudo-statements. Thus instead of eradicating metaphysics from the empirical sciences, positivism leads to the invasion of metaphysics into the scientific realm.[8]

In contrast to these anti-metaphysical stratagems – anti-metaphysical in intention, that is – my business, as I see it, is not to bring about the overthrow of metaphysics. It is, rather, to formulate a suitable characterization of empirical science, or to define the concepts 'empirical science' and 'metaphysics' in such a way that we shall be able to say of a given system of statements whether or not its closer study is the concern of empirical science.

My criterion of demarcation will accordingly have to be regarded as a *proposal for an agreement or convention.* As to the suitability of any such convention opinions may differ; and a reasonable discussion of these questions is only possible between parties having some purpose in common. The choice of that purpose must, of course, be ultimately a matter of decision, going beyond rational argument.*[5]

Thus anyone who envisages a system of absolutely certain, irrevocably true statements[9] as the end and purpose of science will certainly reject the proposals I shall make here. And so will those who see 'the essence of science ... in its dignity', which they think resides in its 'wholeness' and its 'real truth and essentiality'.[10] They will hardly be ready to grant this dignity to modern theoretical physics in which I and others see the most complete realization to date of what I call 'empirical science'.

The aims of science which I have in mind are different. I do not try to justify them, however, by representing them as the true or the essential aims of science. This would only distort the issue, and it would mean a relapse into positivist dogmatism. There is only *one* way, as far as I can see, of arguing rationally in support of my proposals. This is to analyse their logical consequences: to point out their fertility – their power to elucidate the problems of the theory of knowledge.

Thus I freely admit that in arriving at my proposals I have been guided, in the last analysis, by value judgments and predilections. But I hope that my proposals may be acceptable to those who value not only logical rigour but also freedom from dogmatism; who seek practical applicability, but are even more attracted by the adventure of science, and by discoveries which again and again confront us with new and unexpected questions, challenging us to try out new and hitherto undreamed-of answers.

The fact that value judgments influence my proposals does not mean that I am making the mistake of which I have accused the positivists –

that of trying to kill metaphysics by calling it names. I do not even go so far as to assert that metaphysics has no value for empirical science. For it cannot be denied that along with metaphysical ideas which have obstructed the advance of science there have been others – such as speculative atomism – which have aided it. And looking at the matter from the psychological angle, I am inclined to think that scientific discovery is impossible without faith in ideas which are of a purely speculative kind, and sometimes even quite hazy; a faith which is completely unwarranted from the point of view of science, and which, to that extent, is 'metaphysical'.[11]

Yet having issued all these warnings, I still take it to be the first task of the logic of knowledge to put forward a *concept of empirical science*, in order to make linguistic usage, now somewhat uncertain, as definite as possible, and in order to draw a clear line of demarcation between science and metaphysical ideas – even though these ideas may have furthered the advance of science throughout its history.

5. Experience as a method

The task of formulating an acceptable definition of the idea of empirical science is not without its difficulties. Some of these arise from *the fact that there must be many theoretical systems* with a logical structure very similar to the one which at any particular time is the accepted system of empirical science. This situation is sometimes described by saying that there are a great many – presumably an infinite number – of 'logically possible worlds'. Yet the system called 'empirical science' is intended to represent only *one* world: the 'real world' or the 'world of our experience'.*[1]

In order to make this idea a little more precise, we may distinguish three requirements which our empirical theoretical system will have to satisfy. First, it must be *synthetic*, so that it may represent a noncontradictory, a *possible* world. Secondly, it must satisfy the criterion of demarcation (cf. Sections 6 and 21), i.e. it must not be metaphysical, but must represent a world of possible *experience*. Thirdly, it must be a system distinguished in some way from other such systems as the one which represents *our* world of experience.

But how is the system that represents our world of experience to be distinguished? The answer is: by the fact that it has been submitted to

tests, and has stood up to tests. This means that it is to be distinguished by applying to it that deductive method which it is my aim to analyse, and to describe.

'Experience', on this view, appears as a distinctive *method* whereby one theoretical system may be distinguished from others; so that empirical science seems to be characterized not only by its logical form but, in addition, by its distinctive *method.* (This, of course, is also the view of the inductivists, who try to characterize empirical science by its use of the inductive method.)

The theory of knowledge whose task is the analysis of the method or procedure peculiar to empirical science, may accordingly be described as a theory of the empirical method – *a theory of what is usually called experience.*

6. Falsifiability as a Criterion of Demarcation

The criterion of demarcation inherent in inductive logic – that is, the positivistic dogma of meaning – is equivalent to the requirement that all the statements of empirical science (or all 'meaningful' statements) must be capable of being finally decided, with respect to their truth *and* falsity; we shall say that they must be '*conclusively decidable*'. This means that their form must be such that *to verify them and to falsify them* must both be logically possible. Thus Schlick says: "... a genuine statement must be capable of *conclusive verification*"[1]; and Waismann says still more clearly: "If there is no possible way to *determine whether a statement is true* then that statement has no meaning whatsoever. For the meaning of a statement is the method of its verification."[2]

Now in my view there is no such thing as induction.*[1] Thus inference to theories, from singular statements which are 'verified by experience' (whatever that may mean), is logically inadmissible. Theories are, therefore, *never* empirically verifiable. If we wish to avoid the positivist's mistake of eliminating, by our criterion of demarcation, the theoretical systems of natural science,*[2] then we must choose a criterion which allows us to admit to the domain of empirical science even statements which cannot be verified.

But I shall certainly admit a system as empirical or scientific only if it is capable of being *tested* by experience. These considerations suggest

that not the *verifiability* but the *falsifiability* of a system is to be taken as a criterion of demarcation.*[3] In other words: I shall not require of a scientific system that it shall be capable of being singled out, once and for all, in a positive sense; but I shall require that its logical form shall be such that it can be singled out, by means of empirical tests, in a negative sense: *it must be possible for an empirical scientific system to be refuted by experience.*[3]

(Thus the statement, 'It will rain or not rain here tomorrow' will not be regarded as empirical, simply because it cannot be refuted; whereas the statement, 'It will rain here tomorrow' will be regarded as empirical.)

Various objections might be raised against the criterion of demarcation here proposed. In the first place, it may well seem somewhat wrong-headed to suggest that science, which is supposed to give us positive information, should be characterized as satisfying a negative requirement such as refutability. However, I shall show, in Sections 31 to 46, that this objection has little weight, since the amount of positive information about the world which is conveyed by a scientific statement is the greater the more likely it is to clash, because of its logical character, with possible singular statements. (Not for nothing do we call the laws of nature 'laws': the more they prohibit the more they say.)

Again, the attempt might be made to turn against me my own criticism of the inductivist criterion of demarcation; for it might seem that objections can be raised against falsifiability as a criterion of demarcation similar to those which I myself raised against verifiability.

This attack would not disturb me. My proposal is based upon an *asymmetry* between verifiability and falsifiability; an asymmetry which results from the logical form of universal statements.*[4] For these are never derivable from singular statements, but can be contradicted by singular statements. Consequently it is possible by means of purely deductive inferences (with the help of the *modus tollens* of classical logic) to argue from the truth of singular statements to the falsity of universal statements. Such an argument to the falsity of universal statements is the only strictly deductive kind of inference that proceeds, as it were, in the 'inductive direction'; that is, from singular to universal statements.

A third objection may seem more serious. It might be said that even if the asymmetry is admitted, it is still impossible, for various reasons, that any theoretical system should ever be conclusively falsified. For it is

always possible to find some way of evading falsification, for example by introducing *ad hoc* an auxiliary hypothesis, or by changing *ad hoc* a definition. It is even possible without logical inconsistency to adopt the position of simply refusing to acknowledge any falsifying experience whatsoever. Admittedly, scientists do not usually proceed in this way, but logically such procedure is possible; and this fact, it might be claimed, makes the logical value of my proposed criterion of demarcation dubious, to say the least.

I must admit the justice of this criticism; but I need not therefore withdraw my proposal to adopt falsifiability as a criterion of demarcation. For I am going to propose (in Sections 20f.) that the *empirical method* shall be characterized as a method that excludes precisely those ways of evading falsification which, as my imaginary critic rightly insists, are logically admissible. According to my proposal, what characterizes the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested. Its aim is not to save the lives of untenable systems but, on the contrary, to select the one which is by comparison the fittest, by exposing them all to the fiercest struggle for survival.

The proposed criterion of demarcation also leads us to a solution of Hume's problem of induction – of the problem of the validity of natural laws. The root of this problem is the apparent contradiction between what may be called 'the fundamental thesis of empiricism' – the thesis that experience alone can decide upon the truth or falsity of scientific statements – and Hume's realization of the inadmissibility of inductive arguments. This contradiction arises only if it is assumed that all empirical scientific statements must be 'conclusively decidable', i.e. that their verification and their falsification must both in principle be possible. If we renounce this requirement and admit as empirical also statements which are decidable in one sense only – unilaterally decidable and, more especially, falsifiable – and which may be tested by systematic attempts to falsify them, the contradiction disappears: the method of falsification presupposes no inductive inference, but only the tautological transformations of deductive logic whose validity is not in dispute.[4]

7. The problem of the 'empirical basis'

If falsifiability is to be at all applicable as a criterion of demarcation,

then singular statements must be available which can serve as premisses in falsifying inferences. Our criterion therefore appears only to shift the problem – to lead us back from the question of the empirical character of theories to the question of the empirical character of singular statements.

Yet even so, something has been gained. For in the practice of scientific research, demarcation is sometimes of immediate urgency in connection with theoretical systems, whereas in connection with singular statements, doubts as to their empirical character rarely arise. It is true that errors of observation occur and give rise to false singular statements, but the scientist scarcely ever has occasion to describe a singular statement as non-empirical or metaphysical.

Problems of the empirical basis – that is, problems concerning the empirical character of singular statements, and how they are tested – thus play a part within the logic of science that differs somewhat from that played by most of the other problems which will concern us. For most of these stand in close relation to the *practice* of research, whilst the problem of the empirical basis belongs almost exclusively to the *theory* of knowledge. I shall have to deal with them, however, since they have given rise to many obscurities. This is especially true of the relation between *perceptual experiences* and *basic statements*. (What I call a 'basic statement' or a 'basic proposition' is a statement which can serve as a premise in an empirical falsification; in brief, a statement of a singular fact.)

Perceptual experiences have often been regarded as providing a kind of justification for basic statements. It was held that these statements are 'based upon' these experiences; that their truth becomes 'manifest by inspection' through these experiences; or that it is made 'evident' by these experiences, etc. All these expressions exhibit the perfectly sound tendency to emphasize the close connection between the basic statements and our perceptual experiences. Yet it was also rightly felt that *statements can be logically justified only by statements*. Thus the connection between the perceptions and the statements remained obscure, and was described by correspondingly obscure expressions which elucidated nothing, but slurred over the difficulties or, at best, adumbrated them through metaphors.

Here too a solution can be found, I believe, if we clearly separate the

psychological from the logical and methodological aspects of the problem. We must distinguish between, on the one hand, *our subjective experiences or our feelings of conviction*, which can never justify any statement (though they can be made the subject of psychological investigation) and, on the other hand, the *objective logical relations* subsisting among the various systems of scientific statements, and within each of them.

The problems of the empirical basis will be discussed in some detail in Sections 25 to 30. For the present I had better turn to the problem of scientific objectivity, since the terms 'objective' and 'subjective' which I have just used are in need of elucidation.

8. Scientific objectivity and subjective conviction

The words 'objective' and 'subjective' are philosophical terms heavily burdened with a heritage of contradictory usages and of inconclusive and interminable discussions.

My use of the terms 'objective' and 'subjective' is not unlike Kant's. He uses the word 'objective' to indicate that scientific knowledge should be *justifiable*, independently of anybody's whim: a justification is 'objective' if in principle it can be tested and understood by anybody. "If something is valid", he writes, "for anybody in possession of his reason, then its grounds are objective and sufficient".[1]

Now I hold that scientific theories are never fully justifiable or verifiable, but that they are nevertheless testable. I shall therefore say that the *objectivity* of scientific statements lies in the fact that they can be *intersubjectively tested.**[1]

The word 'subjective' is applied by Kant to our feelings of conviction (of varying degrees).[2] To examine how these come about is the business of psychology. They may arise, for example, "in accordance with the laws of association".[3] Objective reasons too may serve as "subjective *causes* of judging",[4] in so far as we may reflect upon these reasons, and become convinced of their cogency.

Kant was perhaps the first to realize that the objectivity of scientific statements is closely connected with the construction of theories – with the use of hypotheses and universal statements. Only when certain events recur in accordance with rules or regularities, as is the case with repeatable experiments, can our observations be tested – in principle – by anyone.

We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated 'coincidence', but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable.[5]

Every experimental physicist knows those surprising and inexplicable apparent 'effects' which can perhaps even be reproduced in his laboratory for some time, but which finally disappear without trace. Of course, no physicist would say in such a case that he had made a scientific discovery (though he might try to rearrange his experiments so as to make the effect reproducible). Indeed the scientifically significant *physical effect* may be defined as that which can be regularly reproduced by anyone who carries out the appropriate experiment in the way prescribed. No serious physicist would offer for publication, as a scientific discovery, any such 'occult effect', as I propose to call it – one for whose reproduction he could give no instructions. The 'discovery' would be only too soon rejected as chimerical, simply because attempts to test it would lead to negative results.[6] (It follows that any controversy over the question whether events which are in principle unrepeatable and unique ever do occur cannot be decided by science: it would be a metaphysical controversy.)

We may now return to a point made in the previous section: to my thesis that a subjective experience, or a feeling of conviction, can never justify a scientific statement, and that within science it can play no part but that of the subject of an empirical (a psychological) inquiry. No matter how intense a feeling of conviction it may be, it can never justify a statement. Thus I can be utterly convinced of the truth of a statement; certain of the evidence of my perceptions; overwhelmed by the intensity of my experience: every doubt may seem to me absurd. But does this afford the slightest reason for science to accept my statement? Can any statement be justified by the fact that K.R.P. is utterly convinced of its truth? The answer is, 'No'; and any other answer would be incompatible with the idea of scientific objectivity. Even the fact, for me so firmly established, that I am experiencing this feeling of conviction, cannot appear within the field of objective science except in the form of a *psychological hypothesis* which, of course, calls for inter-subjective testing: from the conjecture that I have this feeling of conviction the psychologist

may deduce, with the help of psychological and other theories, certain predictions about my behaviour; and these may be confirmed or refuted in the course of experimental tests. But from the epistemological point of view, it is quite irrelevant whether my feeling of conviction was strong or weak; whether it came from a strong or even irresistible impression of indubitable certainty (or 'self-evidence'), or merely from a doubtful surmise. None of this has any bearing on the question of how scientific statements can be justified.

Considerations like these do not of course provide an answer to the problem of the empirical basis. But at least they help us to see its main difficulty. In demanding objectivity for basic statements as well as for other scientific statements, we deprive ourselves of any logical means by which we might have hoped to reduce the truth of scientific statements to our experiences. Moreover we debar ourselves from granting any favoured status to statements which represent experiences, such as those statements which describe our perceptions (and which are sometimes called 'protocol sentences'). They can occur in science only as psychological statements; and this means, as hypotheses of a kind whose standards of inter-subjective testing (considering the present state of psychology) are certainly not very high.

Whatever may be our eventual answer to the question of the empirical basis, one thing must be clear: if we adhere to our demand that scientific statements must be objective, then those statements which belong to the empirical basis of science must also be objective, i.e. inter-subjectively testable. Yet inter-subjective testability always implies that from the statements which are to be tested, other testable statements can be deduced. Thus if the basic statements in their turn are to be inter-subjectively testable, *there can be no ultimate statements in science*: there can be no statements in science which cannot be tested, and therefore none which cannot in principle be refuted, by falsifying some of the conclusions which can be deduced from them.

We thus arrive at the following view. Systems of theories are tested by deducing from them statements of a lesser level of universality. These statements in their turn, since they are to be inter-subjectively testable, must be testable in like manner – and so *ad infinitum*.

It might be thought that this view leads to an infinite regress, and that it is therefore untenable. In Section I, when criticizing induction, I raised

the objection that it may lead to an infinite regress; and it might well appear to the reader now that the very same objection can be urged against that procedure of deductive testing which I myself advocate. However, this is not so. The deductive method of testing cannot establish or justify the statements which are being tested; nor is it intended to do so. Thus there is no danger of an infinite regress. But it must be admitted that the situation to which I have drawn attention – testability *ad infinitum* and the absence of ultimate statements which are not in need of tests – does create a problem. For, clearly, tests cannot in fact be carried on *ad infinitum*: sooner or later we have to stop. Without discussing this problem here in detail, I only wish to point out that the fact that the tests cannot go on for ever does not clash with my demand that every scientific statement must be testable. For I do not demand that every scientific statement must *have in fact been tested* before it is accepted. I only demand that every such statement must be *capable* of being tested; or in other words, I refuse to accept the view that there are statements in science which we have, resignedly, to accept as true merely because it does not seem possible, for logical reasons, to test them.

APPENDIX

... A system such as classical mechanics may be 'scientific' to any degree you like; but those who uphold it dogmatically – believing, perhaps, that it is their business to defend such a successful system against criticism as long as it is not *conclusively disproved* – are adopting the very reverse of that critical attitude which in my view is the proper one for the scientist. In point of fact, no conclusive disproof of a theory can ever be produced; for it is always possible to say that the experimental results are not reliable, or that the discrepancies which are asserted to exist between the experimental results and the theory are only apparent and that they will disappear with the advance of our understanding. (In the struggle against Einstein, both these arguments were often used in support of Newtonian mechanics, and similar arguments abound in the field of the social sciences.) If you insist on strict proof (or strict disproof*[1]) in the empirical sciences, you will never benefit from experience, and never learn from it how wrong you are.

If therefore we characterize empirical science merely by the formal

or logical structure of its statements, we shall not be able to exclude from it that prevalent form of metaphysics which results from elevating an obsolete scientific theory into an incontrovertible truth.

Such are my reasons for proposing that empirical science should be characterized by its methods: by our manner of dealing with scientific systems: by what we do with them and what we do to them. Thus I shall try to establish the rules, or if you will the norms, by which the scientist is guided when he is engaged in research or in discovery, in the sense here understood.

... The falsifying mode of inference here referred to – the way in which the falsification of a conclusion entails the falsification of the system from which it is derived – is the *modus tollens* of classical logic....

By means of this mode of inference we falsify *the whole system* (the theory as well as the initial conditions) which was required for the deduction of the statement *p*, i.e. of the falsifying statement. Thus it cannot be asserted of any one statement of the system that it is, or is not, specifically upset by the falsification. Only if *p* is *independent* of some part of the system can we say that this part is not involved in the falsification.[1] With this is connected the following possibility: we may, in some cases, perhaps in consideration of the *levels of universality*, attribute the falsification to some definite hypothesis – for instance to a newly introduced hypothesis. This may happen if a well-corroborated theory, and one which continues to be further corroborated, has been deductively explained by a new hypothesis of a higher level. The attempt will have to be made to test this new hypothesis by means of some of its consequences which have not yet been tested. If any of these are falsified, then we may well attribute the falsification to the new hypothesis alone. We shall then seek, in its stead, other high-level generalizations, but we shall not feel obliged to regard the old system, of lesser generality, as having been falsified.

NOTES

* Chapter 1 of *The Logic of Scientific Discovery*, and, as an appendix, selections from pp. 50, 76, 77; copyright © 1959 by Karl Raimund Popper. Published by Basic Books, New York, 1959, seventh English impression Hutchinson of London, 1974. Reprinted by permission. This book is a translation by the author (with the assistance of Julius Freed and Lan Freed) of *Logik der Forschung*, first published in Vienna in 1935. The

starred notes were added to the translation by the author. The various appendices, the Postscript, and the other numbered sections referred to in the footnotes may be found in this translation.

SECTION 1

[1] H. Reichenbach, *Erkenntnis* **1** (1930), 186 (cf. also p. 64 f.).
[2] Reichenbach, *ibid.*, p. 67.
*[1] The decisive passages from Hume are quoted in appendix *vii, text to Notes 4, 5, and 6; see also Note 2 to Section 81, below.
[3] Cf. J. M. Keynes, *A Treatise on Probability* (1921); O. Külpe, *Vorlesungen über Logic* (ed. by Selz, 1923); Reichenbach (who uses the term 'probability implications'), 'Axiomatik der Wahrscheinlichkeitrechnung', *Mathem. Zeitschr.* **34** (1932); and in many other places.
[4] Reichenbach, *Erkenntnis* **1** (1930), 186.
*[2] See also Chapter x, below, especially Note 2 to Section 81, and Chapter *ii of the *Postscript* for a fuller statement of this criticism.
[5] Liebig (in *Induktion und Deduktion*, 1865) was probably the first to reject the inductive method from the standpoint of natural science; his attack is directed against Bacon. Duhem (in *La Théorie physique, son objet et sa structure*, 1906; English translation by P. P. Wiener: *The Aim and Structure of Physical Theory*, Princeton, 1954) held pronounced deductivist views. (*But there are also inductivist views to be found in Duhem's book, for example in the third chapter, Part One, where we are told that only experiment, induction, and generalization have produced Descartes's law of diffraction; cf. the English translation, p. 455.) See also V. Kraft, *Die Grundformen der Wissenschaftlichen Methoden*, 1925; and Carnap, *Erkenntnis* **2** (1932), 440.

SECTION 2

[1] Address on Max Planck's 60th birthday. The passage quoted begins with the words, "The supreme task of the physicist is to search for those universal laws...", etc. (quoted from A. Einstein, *Mein Weltbild*, 1934, p. 168; English translation by A. Harris: *The World as I see It*, 1935, p. 125). Similar ideas are found earlier in Liebig, *op. cit.*; cf. also Mach, *Principien der Wärmelehre* (1896), p. 443ff. *The German word '*Einfühlung*' is difficult to translate. Harris translates: "sympathetic understanding of experience".

SECTION 3

*[1] For this term, see Note *1 before Section 79, and Section *[29] of my *Postscript*.

SECTION 4

[1] With this (and also with Sections 1 to 6 and 13 to 24) cf. my note: *Erkenntnis* **3** (1933), 426; *It is now here reprinted, in translation, as appendix*.
[2] Cf. the last sentence of his *Enquiry Concerning Human Understanding*. *With the next paragraph, compare for example the quotation from Reichenbach in the text to Note 1, Section 1.
*[1] When I wrote this paragraph I overrated the 'modern positivists', as I now see. I should have remembered that *in this respect* the promising beginning of Wittgenstein's *Tractatus* – "The world is the totality of facts, not of things" – was cancelled by its end which denounced the man who "had given no meaning to certain signs in his propositions". See also my *Open Society and its Enemies*, Chapter II, Section ii, and Chapter *i of my *Postscript*, especially Sections *11 (Note 5), *24 (the last five paragraphs), and *25.

*2 Nothing depends on names, of course. When I invented the new name 'basic statement' (or 'basic proposition'; see below, Sections 7 and 28) I did so only because I needed a term *not* burdened with the connotation of a perception statement. But unfortunately it was soon adopted by others, and used to convey precisely the kind of meaning which I wished to avoid. Cf. also my *Postscript*, *29.

*3 Hume thus condemned his own *Enquiry* on its last page, just as later Wittgenstein condemned his own *Tractatus* on its last page. (See Note 2 to Section 10.)

3 Carnap, *Erkenntnis* **2** (1932), 219 ff. Earlier Mill had used the word 'meaningless' in a similar way, *no doubt under the influence of Comte; cf. Comte's *Early Essays on Social Philosophy*, ed. by H. D. Hutton, 1911, p. 223. See also my *Open Society*, Note 51 to Chapter 11.

4 Wittgenstein, *Tractatus Logico-Philosophicus* (1918 and 1922), Proposition 5. *As this was written in 1934, I am dealing here of course *only* with the *Tractatus*.

5 Wittgenstein, *op. cit.*, Proposition 4.01; 4.03; 2.221.

6 Cf. Note 1 to Section 2.

*4 The idea of treating scientific laws as pseudo-propositions – thus solving the problem of induction – was attributed by Schlick to Wittgenstein. (Cf. my *Open Society*, Notes 46 and 51 f. to Chapter II.) But it is really much older. It is part of the instrumentalist tradition which can be traced back to Berkeley, and further. (See for example my paper 'Three Views Concerning Human Knowledge', in *Contemporary British Philosophy*, 1956; and 'A Note on Berkeley as a Precursor of Mach', in *The British Journal for the Philosophy of Science* iv, 4, (1953), 26 ff., now in my *Conjectures and Refutations*, 1959. Further references in Note *I before Section 12 (p. 59). The problem is also treated in my *Postscript*, Sections *11 to *14, and *19 to *26.)

7 Schlick, *Naturwissenschaften* **19** (1931), 156. (The italics are mine.) Regarding natural laws Schlick writes (p. 151), "It has often been remarked that, strictly, we can never speak of an absolute verification of a law, since we always, so to speak, tacitly make the reservation that it may be modified in the light of further experience. If I may add, by way of parenthesis", Schlick continues, "a few words on the logical situation, the above-mentioned fact means that a natural law, in principle, does not have the logical character of a statement, but is, rather, a prescription for the formation of statements." *('Formation' no doubt was meant to include transformation or derivation.) Schlick attributed this theory to a personal communication of Wittgenstein's. See also Section *12 of my *Postscript*.

8 Cf. Section 78 (for example Note 1). *See also my *Open Society*, Notes 46, 51, and 52 to Chapter II, and my paper 'The Demarcation between Science and Metaphysics', contributed in January 1955 to the planned Carnap volume of the *Library of Living Philosophers*, ed. by P. A. Schilpp.

*5 I believe that a reasonable discussion is always possible between parties interested in truth, and ready to pay attention to each other. (Cf. my *Open Society*, Chapter 24.)

9 This is Dingler's view; cf. Note 1 to Section 19.

10 This is the view of O. Spaan (*Kategorienlehre*, 1924).

11 Cf. also Planck, *Positivismus und reale Aussenwelt* (1931) and Einstein, 'Die Religiosität der Forschung', in *Mein Weltbild* (1934), p. 43; English translation by A. Harris: *The World as I See It* (1935), p. 23 ff. *See also Section 85, and my *Postscript*.

SECTION 5

*1 Cf. appendix *x.

SECTION 6

[1] Schlick, *Naturwissenschaften* **19** (1931), 150.
[2] Waismann, *Erkenntnis* **1** (1930), 229.
*[1] I am not, of course, here considering so-called 'mathematical induction'; what I am denying is that there is such a thing as induction in the so-called 'inductive sciences'; that there are either 'inductive procedures' or 'inductive inferences'.
*[2] In his *Logical Syntax* (1937, p. 321 f.) Carnap admitted that this was a mistake (with a reference to my criticism); and he did so even more fully in 'Testability and Meaning', recognizing the fact that universal laws are not only 'convenient' for science but even 'essential' (*Philosophy of Science* **4** (1937), 27). But in his inductivist *Logical Foundations of Probability* (1950), he returns to a position very like the one here criticized: finding that universal laws have zero probability (p. 511), he is compelled to say (p. 575) that though they need not be expelled from science, science can very well do without them.
*[3] Note that I suggest falsifiability as a criterion of demarcation, but *not of meaning*. Note, moreover, that I have already (Section 4) sharply criticized the use of the idea of meaning as a criterion of demarcation, and that I attack the dogma of meaning again, even more sharply, in Section 9. It is therefore a sheer myth (though any number of refutations of my theory have been based upon this myth) that I ever proposed falsifiability as a criterion of meaning. Falsifiability separates two kinds of perfectly meaningful statements: the falsifiable and the non-falsifiable. It draws a line inside meaningful language, not around it. See also Appendix *i, and Chapter *i of my *Postscript*, especially Sections *17 and *19.
[3] Related ideas are to be found, for example, in Frank, *Die Kausalität und ihre Grenzen* (1931), Ch. I, Section 10 (p. 15 f); Dubislav, *Die Definition* (3rd edition 1931), p. 100 f. (Cf. also Note 1 to Section 4, above.).
*[4] This asymmetry is now more fully discussed in Section *22 of my *Postscript*.
[4] For this see also my paper mentioned in note 1 to section 4, *now here reprinted as appendix *i; and my *Postscript*, esp. Section *2.

SECTION 8

[1] *Kritik der reinen Vernunft*, Methodenlehre, 2. Haupstück, 3. Abschnitt (2nd edition, p. 848; English Translation by N. Kemp Smith, 1933: *Critique of Pure Reason*, The Transcendental Doctrine of Method, Chapter ii, Section 3, p. 645).
*[1] I have since generalized this formulation; for inter-subjective *testing* is merely a very important aspect of the more general idea of inter-subjective *criticism*, or in other words, of the idea of mutual rational control by critical discussion. This more general idea, discussed at some length in my *Open Society and Its Enemies*, Chapters 23 and 24, and in my *Poverty of Historicism*, Section 32, is also discussed in my *Postscript*, especially in Chapters *i, *ii, and *vi.
[2] *Ibid.*
[3] Cf. *Kritik der reinen Vernunft*, Transcendentale Elementarlehre Section 19 (2nd edition, p. 142; English translation by N. Kemp Smith, 1933: *Critique of Pure Reason*, Transcendental Doctrine of Elements, Section 19, p. 159).
[4] Cf. *Kritik der reinen Vernunft*, Methodenlehre, 2. Haupstück, 3. Abschnitt (2nd edition, p. 849; English translation, Chapter ii, Section 3, p. 646).
[5] Kant realized that from the required objectivity of scientific statements it follows that they must be at any time inter-subjectively testable, and that they must therefore have the form of universal laws or theories. He formulated this discovery somewhat

obscurely by his "principle of temporal succession according to the law of causality" (which principle he believed that he could prove *a priori* by employing the reasoning here indicated). I do not postulate any such principle (cf. Section 12); but I agree that scientific statements, since they must be inter-subjectively testable, must always have the character of universal hypotheses.* See also Note *1 to Section 22.

[6] In the literature of physics there are to be found some instances of reports, by serious investigators, of the occurrence of effects which could not be reproduced, since further tests led to negative results. A well-known example from recent times is the unexplained positive result of Michelson's experiment observed by Miller (1921–1926) at Mount Wilson, after he himself (as well as Morley) had previously reproduced Michelson's negative result. But since later tests again gave negative results it is now customary to regard these latter as decisive, and to explain Miller's divergent result as "due to unknown sources of error". *See also Section 22, especially Note *1.

Appendix

*[1] I have now here added in brackets the words 'or strict disproof' to the text (a) because they are clearly implied by what is said immediately before ("no conclusive disproof of a theory can ever be produced"), and (b) because I have been constantly misinterpreted as upholding a criterion (and moreover one of *meaning* rather than of *demarcation*) based upon a doctrine of 'complete' or 'conclusive' falsifiability.

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

KARL R. POPPER

BACKGROUND KNOWLEDGE AND SCIENTIFIC GROWTH*

People involved in a fruitful critical discussion of a problem often rely, if only unconsciously, upon two things: the acceptance by all parties of the common aim of getting at the truth, or at least nearer to the truth, and a considerable amount of common background knowledge. This does not mean that either of these two things is an indispensible basis of every discussion, or that these two things are themselves '*a priori*', and cannot be critically discussed in their turn. It only means that criticism never starts from nothing, even though every one of its starting points *may* be challenged, one at a time, in the course of the critical debate.

Yet though every one of our assumptions may be challenged, it is quite impracticable to challenge all of them at the same time. Thus all criticism must be piecemeal (as against the holistic view of Duhem and of Quine); which is only another way of saying that the fundamental maxim of every critical discussion is that we should stick to our problem, and that we should subdivide it, if practicable, and try to solve no more than one problem at a time, although we may, of course, always proceed to a subsidiary problem, or replace our problem by a better one.

While discussing a problem we always accept (if only temporarily) all kinds of things as *unproblematic*: they constitute for the time being, and for the discussion of this particular problem, what I call our *background knowledge*. Few parts of this background knowledge will appear to us in all contexts as absolutely unproblematic, and any particular part of it *may* be challenged at any time, especially if we suspect that its uncritical acceptance may be responsible for some of our difficulties. But almost all of the vast amount of background knowledge which we constantly use in any informal discussion will, for practical reasons, necessarily remain unquestioned; and the misguided attempt to question it all – that is to say, *to start from scratch* – can easily lead to the breakdown of a critical debate. (Were we to start the race where Adam started, I know of no reason why we should get any further than Adam did.)

The fact that, as a rule, we are at any given moment taking a vast amount of traditional knowledge for granted (for almost all our knowledge is traditional) creates no difficulty for the falsificationist or fallibilist. For he does not *accept* this background knowledge; neither as established nor as fairly certain, nor yet as probable. He knows that even its tentative acceptance is risky, and stresses that every bit of it is open to criticism, even though only in a piecemeal way. We can never be certain that we shall challenge the right bit; but since our quest is not for certainty, this does not matter. It will be noticed that this remark contains my answer to Quine's holistic view of empirical tests; a view which Quine formulates (with reference to Duhem), by asserting that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.[1] Now it has to be admitted that we can often test only a large chunk of a theoretical system, and sometimes perhaps only the whole system, and that, in these cases, it is sheer guesswork which of its ingredients should be held responsible for any falsification; a point which I have tried to emphasize – also with reference to Duhem – for a long time past. Though this argument may turn a verificationist into a sceptic, it does not affect those who hold that all our theories are guesses anyway.

This shows that the holistic view of tests, even if it were true, would not create a serious difficulty for the fallibilist and falsificationist. On the other hand, it should be said that the holistic argument goes much too far. It is possible in quite a few cases to find which hypothesis is responsible for the refutation; or in other words, which part, or group of hypotheses, was necessary for the derivation of the refuted prediction. The fact that such logical dependencies may be discovered is established by the practice of *independence proofs* of axiomatized systems; proofs which show that certain axioms of an axiomatic system cannot be derived from the rest. The more simple of these proofs consist in the construction, or rather in the discovery, of a *model* – a set of things, relations, operations, or functions – which satisfies all the axioms except the *one* whose independence is to be shown: for this one axiom – and therefore for the theory as a whole – the model constitutes a counter example.

Now let us say that we have an axiomatized theoretical system, for example of physics, which allows us to predict that certain things do not happen, and that we discover a counter example. There is no reason what-

ever why this counter example may not be found to satisfy most of our axioms or even all our axioms except one whose independence would be thus established. This shows that the holistic dogma of the 'global' character of all tests or counter examples is untenable. And it explains why, even without axiomatizing our physical theory, we may well have an inkling of what has gone wrong with our system.

This, incidentally, speaks in favour of operating, in physics, with highly analysed theoretical systems – that is, with systems which, even though they may fuse all the hypotheses into one, allow us to separate various groups of hypotheses, each of which may become an object of refutation by counter examples. (An excellent recent example is the rejection, in atomic theory, of the law of parity; another is the rejection of the law of commutation for conjugate variables, prior to their interpretation as matrices, and to the statistical interpretation of these matrices.)

NOTES

[1] See W. V. Quine, *From a Logical Point of View*, 1953, p. 41.

ADOLF GRÜNBAUM*

THE DUHEMIAN ARGUMENT**

ABSTRACT. This paper offers a refutation of P. Duhem's thesis that the *falsifiability* of an isolated empirical hypothesis H as an *explanans* is *unavoidably inconclusive*. Its central contentions are the following:

(1) No general features of the logic of falsifiability can assure, for every isolated empirical hypotheses H and independently of the domain to which it pertains, that H can always be preserved as an *explanans* of any empirical findings O whatever by some modification of the auxiliary assumptions A in conjunction with which H functions as an *explanans*. For Duhem *cannot* guarantee on any general logical grounds the deducibility of O from an *explanans* constituted by the conjunction of H and some revised *non*-trivial version R of A: the existence of the required set R of collateral assumptions must be demonstrated for each particular case.

(2) The categorical form of the Duhemian thesis is not only a *non-sequitur* but actually false. This is shown by adducing the testing of physical geometry as a counterexample to Duhem in the form of a rebuttal to A. Einstein's geometrical articulation of Duhem's thesis.

(3) The possibility of a quasi *a priori* choice of a physical geometry in the sense of Duhem must be clearly *distinguished* from the feasibility of a conventional adoption of such a geometry in the sense of H. Poincaré. And the legitimacy of the latter cannot be invoked to save the Duhemian thesis from refutation by the foregoing considerations.

In a recent paper published in this journal, Herburt [6] expounds and endorses the Duhemian argument but contests its being adduced by Quine to repudiate the distinction between analytic and synthetic statements.

The present paper is intended to report the result of an investigation by the author which shows that the categorical form of Duhem's contention, viz. that the falsification of part of an *explanans* is always unavoidably inconclusive, is untenable. In particular, it turns out that it is one thing to maintain with Herburt that "in every empirical test a certain number of statements of various types is involved" [6, p. 109] but quite another to conclude in Duhemian fashion, as he does, that "in principle, it is possible... to maintain any particular empirical statement, whatever the data of experience, provided we make appropriate changes in the system of hypotheses which is put to test" [6, p. 108].

We must distinguish the following two forms of Duhem's thesis:

(i) the logic of every disconfirmation, no less than of every confirmation, of a presumably empirical hypothesis H is such as to *involve at some*

stage or other an entire network of interwoven hypotheses in which H is ingredient rather than the separate testing of the component H,

(ii) *No one constituent hypothesis H* can ever be *extricated* from the ever-present web of collateral assumptions so as to be *open* to *decisive refutation* by the evidence as part of an *explanans* of that evidence, just as no such isolation is achievable for purposes of verification. This conclusion becomes apparent by a consideration of the two parts of the schema of unavoidably inconclusive falsifiability, which are:

(a) it is an elementary fact of *deductive* logic that if certain observational consequences O are entailed by the *conjunction* of H and a set A of auxiliary assumptions, then the *failure* of O to materialize entails *not* the falsity of H by itself but only the weaker conclusion that H and A cannot *both* be true; the falsifiability of H is therefore *inconclusive* in the sense that the *falsity* of H is *not deductively inferable* from the premiss $[(H \cdot A) \rightarrow O] \cdot \sim O$,

(b) the actual observational findings O′, which are incompatible with O, *allow* that H be true while A is false, because they always permit the theorist to preserve H with impunity as a part of the *explanans* of O′ by so modifying A that the *conjunction* of H and the *revised* version A′ of A does explain (entail) O′. This preservability of H is to be understood as a retainability *in principle* and does *not* depend on the ability of scientists to propound the required set A′ of collateral assumptions at any given time.

Thus, there is an ingression of a kind of *a priori* choice into physical theory: at the price of suitable compensatory modifications in the remainder of the theory, *any one* of its *component* hypotheses H may be retained in the face of seemingly contrary empirical findings as an *explanans* of these very findings. And this quasi *a priori* preservability of H is sanctioned by the far-reaching theoretical *ambiguity* and flexibility of the logical constraints imposed by the observational evidence.[1]

Let us now consider the two parts (a) and (b) of the schema which the *stronger* form (ii) of the Duhemian thesis claims to be the *universal paradigm* of the logic of falsifiability in empirical science. Clearly, part (a) is valid, being a *modus tollens* argument in which the antecedent of the conditional premise is a conjunction. But part (a) utilizes the *de facto* findings O′ only to the extent that they are *incompatible* with the observational expectations O derived from the conjunction of H and A. And part (a) is *not at all sufficient to show that the falsifiability of H as part of an explanans of the actual empirical facts O′ is unavoidably inconclusive.*

For neither part (a) nor other general logical considerations can *guarantee* the deducibility of O′ from an *explanans* constitued by the conjunction of H and some *non-trivial* revised set A′ of the auxiliary assumptions which is logically incompatible with A *under the hypothesis H.*[2]

How then does Duhem propose to assure that there exists such a *non-trivial* set A′ for any one component hypothesis H *independently* of the domain of empirical science to which H pertains? It would seem that such assurance *cannot* be given on general logical grounds at all but that the existence of the required set A′ needs *separate* and *concrete* demonstration for each particular case. In short, even in contexts to which part (a) of the Duhemian schema is applicable – which is *not* true for *all* contexts, as we shall see – neither the premiss

$$[(H \cdot A) \rightarrow O] \cdot \sim O,$$

nor other general logical considerations entail that

$$(\exists A')\,[(H \cdot A') \rightarrow O'],$$

where A′ is non-trivial in the sense of Note 2. And hence Duhem's thesis that the falsifiability of an *explanans* H is unavoidably inconclusive is a *non-sequitur.*

That the Duhemian thesis is not only a *non-sequitur* but actually false is borne out, as we shall now see, by the case of testing the hypothesis that a certain *physical geometry* holds, a case of conclusive falsifiability which yields an important *counterexample* to Duhem's stronger thesis concerning falsifiability but which does justify the *weaker* form (i) of his thesis.

I have given a detailed logical analysis of the highly ramified issue of the *empirical credentials* of physical geometry (and chronometry) in other recent publications [3, 4], and I shall confine the present discussion to summarizing those results of that analysis which serve the following objectives of this paper:

1. To substantiate *geometrically* my claim that

(1) by denying the feasibility of conclusive falsification, the Duhemian schema is a serious *misrepresentation* of the actual logical situation characterizing an important class of cases of falsifiability of a purported explanans, and that

(2) the plausibility of Duhem's thesis derives from the false supposition that part (a) of the schema *is* always applicable *and* that its formal validity guarantees the applicability of part (b) of the schema.

The geometrical substantiation of my claim will make apparent the incorrectness of Herburt's contention that part (a) of the schema "establishes conclusively the cogency of Duhem's claim" [6, p. 107].

2. To clarify the important distinction between the *a priori* choice of a physical geometry in the sense of *Duhem* and the conventional choice of such a geometry in the sense of *Poincaré*. An appreciation of this distinction will counteract the prevalent misunderstanding, apparently shared by Herburt [6, p. 105], that Poincare's conventionalist conception of geometry is to be construed as an espousal of the Duhemian argument as applied to the special case of physical geometry.

1. Physical geometry as a counterexample to the Duhemian thesis

Since Duhem's argument was *articulated* and endorsed by Einstein a decade ago in regard to the epistemological status of physical geometry, I shall summarize my critique of Einstein's geometrical version of that argument.

Physical geometry is usually conceived as the system of metric relations exhibited by transported solid bodies *independently of their particular chemical composition*. On this conception, the criterion of congruence can be furnished by a transported solid body for the purpose of determining the geometry by measurement, only if the computational application of suitable 'corrections' (or, ideally, appropriate shielding) has assured rigidity in the sense of essentially eliminating inhomogeneous thermal, elastic, electromagnetic and other perturbational influences. For these influences are 'deforming' in the sense of producing changes of *varying degree* in different kinds of materials. Since the existence of perturbational influences thus issues in a dependence of the coincidence behavior of transported solid rods on the latter's *chemical composition*, and since physical geometry is concerned with the behavior common to all solids apart from their substance-specific idiosyncrasies, the discounting of idiosyncratic distortions is an essential aspect of the logic of physical geometry. The demand for the computational *elimination* of such distortions as a prerequisite to the experimental determination of the geometry has a thermodynamic counterpart: the requirement of a means for measuring temperature which does not yield the discordant results pro-

duced by expansion thermometers at other than fixed points when different thermometric substances are employed. This thermometric need is fulfilled successfully by Kelvin's thermodynamic scale of temperature. But attention to the implementation of the corresponding prerequisite of physical geometry has led Einstein to impugn the empirical status of that geometry. He considers the case in which congruence has been defined by the diverse kinds of transported solid measuring rods *as corrected for their respective idiosyncratic distortions* with a view to then making an empirical determination of the prevailing geometry. And Einstein's thesis is that the very logic of computing these corrections precludes that the geometry itself be accessible to experimental ascertainment *in isolation from* other physical regularities. Specifically, he states his case in the form of a dialogue [1, pp. 676–678] in which he attributes his own Duhemian view to Poincaré for reasons that will become clear later on and offers that view in opposition to Hans Reichenbach's conception [10, 11]. But I submit that Poincaré's text will *not* bear Einstein's interpretation. For in speaking of the variations which solids exhibit under distorting influences, Poincaré says "we neglect these variations in laying the foundations of geometry, because, besides their being very slight, they are irregular and consequently seem to us accidental". [8, p. 76] I am therefore taking the liberty of replacing the name 'Poincaré' in Einstein's dialogue by the term 'Duhem and Einstein'. *With this modification*, the dialogue reads as follows [1, pp. 676–678]:

Duhem and Einstein: The empirically given bodies are not rigid, and consequently can not be used for the embodiment of geometric intervals. Therefore, the theorems of geometry are not verifiable.

Reichenbach: I admit that there are no bodies which can be *immediately* adduced for the 'real definition' [i.e. physical definition] of the interval. Nevertheless, this real definition can be achieved by taking the thermal volume-dependence, elasticity, electro- and magneto-striction, etc., into consideration. That this is really and without contradiction possible, classical physics has surely demonstrated.

Duhem and Einstein: In gaining the real definition improved by yourself you have made use of physical laws, the formulation of which presupposes (in this case) Euclidean geometry. The verification, of which you have spoken, refers, therefore, not merely to geometry but to the entire system of physical laws which constitute its foundation. An examination of geometry by itself is consequently not thinkable. – Why should it consequently not be entirely up to me to choose geometry according to my own convenience (i.e., Euclidean) and to fit the remaining (in the usual sense 'physical') laws to this choice in such manner that there can arise no contradiction of the whole with experience?

Einstein is making two major points here:

(1) In obtaining a physical geometry by giving a physical interpretation of the postulates of a formal geometric axiom system, the specification of the physical meaning of such theoretical terms as 'congruent', 'length', or 'distance' is *not* at all simply a matter of giving an operational definition in the strict sense. Instead, what has been variously called a 'rule of correspondence' (Margenau and Carnap), a 'coordinative definition' (Reichenbach), an 'epistemic correlation' (Northrop) or a 'dictionary' (N. R. Campbell) is provided here *through the mediation of hypotheses and laws* which are *collateral* to the geometric theory whose physical meaning is being specified. Einstein's point that the physical meaning of congruence is given by the transported rod *as corrected theoretically* for idiosyncratic distortions is an illuminating one and has an abundance of analogues throughout physical theory, thus showing, incidentally, that strictly operational definitions are a rather simplified and limiting species of rules of correspondence [cf. 4, §2, Section (ii)].

(2) Einstein's second claim, which is the cardinal one for our purposes, is that the role of collateral theory in the physical definition of congruence is such as to issue in the following *circularity*, from which there is no escape, he maintains, short of acknowledging the existence of an *a priori* element *in the sense of the Duhemian ambiguity*: the rigid body is not even defined without first *decreeing* the validity of Euclidean geometry (or of some other particular geometry). For *before* the *corrected* rod can be used to make an empirical determination of the *de facto* geometry, the required corrections must be computed via laws, such as those of elasticity, which involve Euclideanly-calculated areas and volumes. But clearly the warrant for thus introducing Euclidean geometry *at this stage* cannot be empirical.

If Einstein's Duhemian thesis were to prove correct, then it would have to be acknowledged that there is a sense in which physical geometry *itself* does not provide a geometric characterization of physical reality. For by this characterization we understand the articulation of the system of relations obtaining between bodies and transported solid rods quite apart from their substance-specific distortions. And to the extent to which physical geometry is *a priori* in the sense of the Duhemian ambiguity, there is an ingression of *a priori* elements into physical theory to take the place of distinctively geometric gaps in our knowledge of the physical world.

I now wish to set forth my doubts regarding the soundness of Einstein's contention. And I shall do so in two parts the first of which deals with the special case in which effectively no deforming influences are present in a certain region whose geometry is to be ascertained.

(i) If we are confronted with the problem of the falsifiability of the geometry ascribed to a region which is effectively free from deforming influences, then the *correctional* physical laws play no role as auxiliary assumptions, and the latter reduce to the claim that the region in question is, in fact, effectively *free* from deforming influences. And *if* such freedom can be affirmed *without* presupposing collateral theory, then the geometry alone rather than only a wider theory in which it is ingredient will be falsifiable. The question is therefore whether freedom from deforming influences can be asserted and ascertained independently of (sophisticated) collateral theory. My answer to this question is: Yes. For quite independently of the conceptual elaboration of such physical magnitudes as temperature, whose constancy would characterize a region free from deforming influences, the absence of perturbations is certifiable for the region as follows: two solid rods of very different chemical constitution which coincide at one place in the region will also coincide everywhere else in it independently of their paths of transport. Accordingly, the absence of deforming influences is ascertainable *independently* of any assumptions as to the geometry and of other (sophisticated) collateral theory.

Let us now employ our earlier notation and denote the geometry by 'H' and the assertion concerning the freedom from perturbations by 'A'. Then, once we have laid down the congruence definition and the remaining semantical rules, the physical geometry H becomes conclusively falsifiable as an *explanans* of the posited empirical findings O′. For the actual logical situation is characterized *not* by part (a) of the Duhemian schema but instead by the schema

$$[\{(H \cdot A) \rightarrow O\} \cdot \sim O \cdot A] \rightarrow \sim H.$$

It will be noted that we identified the H of the Duhemian schema with the geometry. But since a geometric theory, at least in its synthetic form, can be axiomatized as a conjunction of logically independent postulates, a particular axiomatization of H could be decomposed logically into various sets of component subhypotheses. Thus, for example, the hypoth-

esis of Euclidean geometry could be stated, if we wished, as the conjunction of two parts consisting respectively of the Euclidean parallel postulate and the postulates of absolute geometry. And the hypothesis of hyperbolic geometry could be stated in the form of a conjunction of absolute geometry and the hyperbolic parallel postulate.

In view of the logically-compound character of a geometric hypothesis, Professor Grover Maxwell has suggested that the Duhemian thesis may be tenable in this context if we construe it as pertaining *not* to the falsifiability of a geometry as a whole but to the falsifiability of its component subhypotheses in any given axiomatization. There are two ways in which this proposed interpretation might be understood: (1) as an assertion that *any one component sub*hypothesis eludes conclusive refutation on the grounds that the empirical findings can falsify the set of axioms only as a whole, or (2) in any given axiomatization of a physical geometry there exists *at least one component sub*hypothesis which eludes conclusive refutation.

The first version of the proposed interpretation will not bear examination. For suppose that H is the hypothesis of Euclidean geometry and that we consider absolute geometry as one of its subhypotheses and the Euclidean parallel postulate as the other. If now the empirical findings were to show that the geometry is hyperbolic, then indeed absolute geometry would have eluded refutation. But if, on the other hand, the prevailing geometry were to turn out to be spherical, then the mere replacement of the Euclidean parallel postulate by the spherical one could not possibly save absolute geometry from refutation. For absolute geometry alone is logically incompatible with spherical geometry and hence with the posited empirical findings.

If one were to read Duhem as per the very cautious *second* version of Maxwell's proposed interpretation, then our analysis of the logic of testing the geometry of a *perturbation-free* region could *not* be adduced as having furnished a counter-example to so mild a form of Duhemism. And the question of the validity of this highly attenuated version is thus left open by our analysis without detriment to that analysis.

We now turn to the critique of Einstein's Duhemian argument as applied to the empirical determination of the geometry of a region which *is* subject to deforming influences.

(ii) There can be no question that when deforming influences *are*

present, the laws used to make the corrections for deformations involve areas and volumes in a fundamental way (e.g. in the definitions of the elastic stresses and strains) and that this involvement presupposes a geometry, as is evident from the area and volume formulae of differential geometry, which contain the square root of the determinant of the components g_{ik} of the metric tensor. Thus, the empirical determination of the geometry involves the joint assumption of a geometry and of certain collateral hypotheses. But we see already that this assumption cannot be adequately represented by the conjunction H·A of the Duhemian schema where H represents the geometry.

Now suppose that we begin with a set of Euclideanly-formulated physical laws P_0 in correcting for the distortions induced by perturbations and then use the thus Euclideanly-corrected congruence standard for *empirically* exploring the geometry of space by determining the metric tensor. *The initial stipulational affirmation of the Euclidean geometry G_0 in the physical laws P_0 used to compute the corrections in no way assures that the geometry obtained by the corrected rods will be Euclidean.* If it is *non-*Euclidean, then the question is: what will be involved in Einstein's fitting of the physical laws to preserve Euclideanism and avoid a contradiction of the theoretical system with experience? Will the adjustments in P_0 necessitated by the retention of Euclideanism entail merely a change in the dependence of the length assigned to the transported rod on such *non-positional* parameters as temperature, pressure, magnetic field etc.? Or could the putative empirical findings compel that the length of the transported rod be likewise made a non-constant function of its *position* and *orientation* as independent variables in order to square the coincidence findings with the requirement of Euclideanism? The temporal variability of distorting influences and the possibility of obtaining non-Euclidean results by measurements carried out in a spatial region uniformly characterized by standard conditions of temperature, pressure, electric and magnetic field strength etc. show it to be *quite doubtful* that the preservation of Euclideanism could always be accomplished short of introducing *the dependence of the rod's length on the independent variables of position and orientation.* Thus, in order to retain Euclideanism, it may be necessary to *remetrize* entirely apart from any consideration of idiosyncratic distortions and even after correcting for these in some way or other. But this kind of remetrization, though entirely admissible in *other*

contexts, does *not* provide the requisite support for Einstein's Duhemian thesis. For it is the avowed onus of that thesis to show that the geometry *by itself* cannot be held to be empirical even when, with Reichenbach, we have sought to assure its empirical character by choosing and then adhering to the customary (standard) definition of congruence, *which excludes resorting to such remetrization.*

Thus, there may well obtain observational findings O′, expressed in terms of a particular definition of congruence (e.g., the *customary* one), which are such that there does *not* exist any *non*-trivial set A′ of auxiliary assumptions capable of preserving the Euclidean H in the face of O′. And this result alone suffices to invalidate the Einsteinian version of Duhem's thesis to the effect that any geometry, such as Euclid's, can be preserved in the face of any experimental findings which are expressed in terms of the customary definition of congruence.

But what of the possibility of actually *extricating* the unique underlying geometry (to within experimental accuracy) from the network of hypotheses which enter into the testing procedure? Elsewhere [3, 4], I have given two methods which can determine the unique underlying geometry in the case of a space of constant curvature, it being an open question whether these methods can also be generalized to cover the case of a space of variable curvature, and, if not, whether there is another method which succeeds in that case.[3]

It might appear that my geometric counterexample to the Duhemian thesis of unavoidably inconclusive falsifiability of an *explanans* is vulnerable to the following criticism:

> To be sure, Einstein's geometric articulation of that thesis does *not* leave room for saving it by resorting to a remetrization in the sense of making the length of the rod *vary* with position or orientation even *after* it has been corrected for idiosyncratic distortions. But why saddle the Duhemian thesis as such with a restriction peculiar to Einstein's particular version of it? And thus why not allow Duhem to save his thesis by countenancing those *alterations in the congruence definition* which are *remetrizations*?

My reply is that to deny the Duhemian the invocation of such an alteration of the congruence definition *in this context* is *not* a matter of gratuitously requiring him to justify his thesis within the confines of Einstein's particular version of that thesis; instead, the imposition of this restriction is entirely legitimate here, and the Duhemian could hardly wish to reject it as unwarranted. For it is of the essence of Duhem's con-

tention that H (in this case: Euclidean geometry) can always be preserved *not* by tampering with the *semantical rules* (interpretive sentences) linking H to the observational base but rather by availing oneself of the alleged *inductive latitude* afforded by the ambiguity of the experimental evidence to do the following: (a) leave the factual commitments of H *unaltered* by retaining both the statement of H and the semantical rules linking its terms to the observational base, and (b) replace the set A by a set A′ of auxiliary assumptions *differing in factual content* from A such that A and A′ are logically incompatible under the hypothesis H. Now, the factual content of a geometrical hypothesis can be *changed* either by preserving the original statement of the hypothesis while changing one or more of the semantical rules or by keeping all of the semantical rules intact and suitably changing the statement of the hypothesis [3, pp. 213–214]. We can see, therefore, that the retention of a Euclidean H by the device of changing through remetrization the semantical rule governing the meaning of 'congruent' (for line segments) effects a retention not of the *factual commitments* of the original Euclidean H but only of its *linguistic trappings*. That the thus 'preserved' Euclidean H actually *repudiates* the factual commitments of the *original* one is clear from the following: the *original* Euclidean H had asserted that the coincidence behavior common to all kinds of solid rods is Euclidean, *if* such transported rods are taken as the physical realization of congruent intervals; but the Euclidean H which survived the confrontation with the posited empirical findings only by dint of a *remetrization* is predicated on a *denial* of the very assertion that was made by the original Euclidean H, which it was to 'preserve'.

Hence, the confines within which the Duhemian must make good his claim of the preservability of a Euclidean H do *not* admit of the kind of change in the congruence definition which alone would render his claim tenable under the assumed empirical conditions. Accordingly, the geometrical critique of Duhem's thesis given in this paper does *not* depend for its validity on restrictions peculiar to Einstein's version of it.

Even apart from the fact that Duhem's thesis precludes resorting to an alternative metrization to save it from refutation in our geometrical context, the very feasibility of alternative metrizations is vouchsafed *not* by any general Duhemian considerations pertaining to the logic of falsifiability but by a property peculiar to the subject matter of geometry (and chronometry): the latitude for *convention* in the ascription of the spatial

(or temporal) *equality* relation to intervals in the continuous manifolds of physical space (or time) [3, 4].

2. The distinction between the conventional adoption of a physical geometry in the sense of H. Poincaré and its quasi 'a priori' choice in the sense of P. Duhem

The key to the difference between the geometric conventionalism of H. Poincaré and the geometrical form of the conventionalism of P. Duhem is furnished by the distinction just used to rebut the objection that Duhem is being unfairly saddled with liabilities incurred by Einstein: the distinction between preserving a particular geometry (e.g. the Euclidean one) by a remetrizational change in the congruence definition, on the one hand, and intending to retain a particular geometry *without* change in that definition (or in other semantical rules) by an alteration of the factual content of the auxiliary assumptions, on the other. More specifically, the Duhemian conception envisions scope for alternative geometric accounts of a given body of evidence only to the extent that these geometries are associated with alternative sets of correctional physical laws. On the other hand, the range of alternative geometric descriptions of given evidence affirmed by Poincaré is far wider and rests on very different grounds: instead of involving the Duhemian *inductive* latitude, Poincaré bases the possibility of giving *either* a Euclidean *or* a non-Euclidean description of the same spatio-physical facts on alternative metrizability. For Poincaré tells us [8, pp. 66–80] *that quite apart from any considerations of substance-specific distorting influences and even after correcting for these in some way or other*, we are at liberty to define congruence – and thereby to fix the geometry appropriate to the given facts – either by calling the solid rod equal to itself everywhere or by making its length vary in a specified way with its position and orientation. Thus, whereas Duhem's affirmation of the retainability of Euclidean geometry in the face of any observational evidence is *inductive*, the preservability of that geometry asserted by Poincaré is *remetrizational*: Poincaré's conventionalist claim regarding geometry is that if the customary definition of congruence on the basis of the coincidence behavior common to all kinds of solid rods does not assure a particular geometric description of the facts, then such a description can be guaranteed remetrizationally, i.e., by merely choosing an

appropriately different noncustomary congruence definition which makes the length of every kind of solid rod a specified *non*-constant function of the *independent* variables of position and orientation.

Duhem's and Poincaré's differing conceptions of the retainability of a given metric geometry issue in two correspondingly different views as to the *interdependence* of geometry and of the remainder of physics. The special optical case of Poincaré's much-discussed and widely-misunderstood statement of the possibility of always giving a Euclidean description of any results of stellar parallax measurements [8, p. 81] will serve to articulate the difference between the *linguistic* interdependence of geometry and optics espoused by Poincaré and their *inductive* (epistemological) interdependence as championed by Duhem.

Poincaré's point is that if the paths of light rays are geodesics on a particular definition of congruence, and if these paths are found parallactically to sustain *non*-Euclidean relations on that metrization, then we need only choose a different definition of congruence such that these *same* paths will no longer be geodesics and that the geodesics of the newly chosen congruence are Euclideanly related. From the standpoint of synthetic geometry, the latter choice effects only a *renaming* of optical and other paths and thus is merely a *recasting of the same factual content in Euclidean language rather than a revision of the extra-linguistic content of optical and other laws*. For an alternative metrization affects only the *language* in which the facts of optics and the coincidence behavior of a transported rod are described: the two geometric descriptions respectively associated with two alternative metrizations are *alternative representations of the same factual content*, and so are the two sets of optical laws corresponding to these geometries. Accordingly, Poincaré is affirming a *linguistic* interdependence of the geometric theory of rigid solids and the optical theory of light rays.

On the other hand, the attempt to explain certain parallactic data by different geometries which constitute alternatives in the inductive sense of Duhem would presumably issue in the following alternative between two theoretical systems, each of which comprises a geometry G and an optics O:

(a) G_E: the geometry of the rigid body geodesics is Euclidean, and

O_1: the paths of light rays do *not* coincide with these geodesics but form a non-Euclidean system,

or

(b) $G_{non\cdot E}$: the geodesics of the rigid body congruence are *not* a Euclidean system, and

O_2: the paths of light rays *do* coincide with these geodesics, and thus they form a non-Euclidean system.

We saw that the physically-interpreted alternative geometries associated with two (or more) different metrizations in the sense of Poincaré have precisely the same total factual content, as do the corresponding two sets of optical laws. By contrast, in the Duhemian account, G_E and $G_{non\cdot E}$ not only *differ* in factual content but are logically incompatible, and so are O_1 and O_2. And on the latter conception, there is sameness of factual content *in regard to the assumed parallactic data* only between the *combined* systems formed by the two conjunctions (G_E and O_1) and ($G_{non\cdot E}$ and O_2). Thus, the need for the combined system of G and O to yield the empirical facts, coupled with the avowed epistemological (inductive) *inseparability* of G and O lead the Duhemian to conceive of their *interdependence* as *inductive* (epistemological).

Hence whereas Duhem (and Einstein) construe the interdependence of G and O inductively such that the geometry by itself is *not* accessible to empirical test, Poincaré's conception of their interdependence allows for an empirical determination of G *by itself, if* we have *renounced* recourse to an alternative metrization in which the length of the rod is held to vary with its position or orientation.

It would seem that it was Poincaré's discussion of the interdependence of optics and geometry by reference to stellar parallax measurements which led Einstein and others to regard him as a proponent of the Duhemian thesis. But this interpretation appears untenable not only in the light of the immediate context of Poincaré's discussion of his astronomical example, but also, as I have shown elsewhere [2, 4], upon taking account of the remainder of his writings.

NOTES

* The author is indebted to the National Science Foundation for the support of research and wishes to acknowledge the benefit of discussions with Dr Grover Maxwell

and other fellow-participants in the 1959 summer sessions of the Minnesota Center for Philosophy of Science.

[1] Cf. P. Duhem, *The Aim and Structure of Physical Theory* (tr. by P. P. Wiener), Princeton, 1954, Part II, Chapter VI [Selection 1 in this collection – Ed.], esp. pp. 183–190. Duhem's explicit disavowal of both decisive falsifiability and crucial verifiability of an *explanans* will *not* bear K. R. Popper's reading of him [9, p. 78]: Popper, who is an exponent of decisive falsifiability [9], misinterprets Duhem as allowing that tests of a hypothesis may be decisively *falsifying* and as denying only that they may be crucially *verifying*. Notwithstanding Popper's *exegetical* error, we shall find presently that his thesis of the feasibility of decisively falsifying tests can be buttressed by a telling counterexample to Duhem's categorical denial of that thesis.

A defense of the claim that *isolated* parts of physical theory can be *confirmed* is outlined by H. Feigl in his 'Confirmability and Confirmation', *Revue Internationale de Philosophie* V (1951), 268–279, which is reprinted in P. P. Wiener (ed.), *Readings in Philosophy of Science*, New York, 1953, esp. pp. 528–529.

[2] The requirement of *non-triviality* of A′ requires clarification. If one were to allow O′ itself, for example, to qualify as a set A′, then, of course, O′ could be deduced trivially, and H would not even be needed in the *explanans*. Hence a *necessary* condition for the *non-triviality* of A′ is that H be required in addition to A′ for the deduction of the *explanandum*. But, as N. Rescher has pointed out to me, this necessary condition is not also sufficient. For it *fails* to rule out an A′ of the trivial form $\sim H \vee O'$ (or $H \supset O'$) from which O′ could *not* be deduced without H.

The unavailability of a formal sufficient condition for non-triviality is not, however, damaging to the critique of Duhem presented in this paper. For surely Duhem's illustrations from the history of physics as well as the whole tenor of his writing indicate that *he intends his thesis to stand or fall on the existence of the kind of A′ which we would all recognize as non-trivial in any given case*. Any endeavor to save Duhem's thesis from refutation by invoking the kind of A′ which no scientist would accept as admissible would turn Duhem's thesis into a most unenlightening triviality that no one would wish to contest. Thus, I have no intention whatever of denying the following compound formal claim: if H and A jointly entail O, the falsity of O does not entail the falsity of H, *and* there will always be *some kind of A′* which, in conjunction with H, will entail O′.

[3] Whatever the answer to this open question, I have argued in another publication [5] that it is wholly misconceived to suppose with J. Maritain [7] that there are *supra*-scientific philosophical means for ascertaining the underlying geometry, if *scientific* procedures do not succeed in unraveling it.

BIBLIOGRAPHY

[1] Einstein, A.: 'Reply to Criticisms', in *Albert Einstein: Philosopher-Scientist* (ed. by Schilpp, P. A.), Evanston, 1949, pp. 665–688.

[2] Grünbaum, A.: 'Carnap's Views on the Foundations of Geometry', in *The Philosophy of Rudolf Carnap* (ed. by Schilpp, P. A.), New York, 1963.

[3] Grünbaum, A.: 'Conventionalism in Geometry', in *The Axiomatic Method* (ed. by Henkin, L., Suppes, P., and Tarski, A.), Amsterdam, 1959, pp. 204–222.

[4] Grünbaum, A.: 'Geometry , Chronometry and Empiricism', in *Minnesota Studies in the Philosophy of Science* (ed. by Feigl, H. and Maxwell, G.), Vol. III, Minneapolis, 1962.

[5] Grünbaum, A.: 'The *A Priori* in Physical Theory', to appear in the Proceedings of the Symposium on the Nature of Physical Knowledge, held at the summer 1959 meeting of the American Physical Society, Milwaukee, Wisconsin.
[6] Herburt, G. K.: 'The Analytic and the Synthetic', *Philosophy of Science* **26** (1959), 104–113.
[7] Maritain, J.: *The Degrees of Knowledge*, New York, 1959, pp. 165–173.
[8] Poincaré, H.: *The Foundations of Science*, Lancaster, 1946.
[9] Popper, K. R.: *The Logic of Scientific Discovery*, London, 1959.
[10] Reichenbach, H.: 'The Philosophical Significance of the Theory of Relativity', in *Albert Einstein: Philosopher-Scientist* (ed. by Schilpp, P. A.), Evanston, 1949, pp. 287–311.
[11] Reichenbach, H.: *The Philosophy of Space and Time*, New York, 1958, Ch. I, §§3–8 incl.

WILLARD VAN ORMAN QUINE

A COMMENT ON GRÜNBAUM'S CLAIM

[*Editor's note*: Quine and Grünbaum have authorized publication of the following letter in which Quine gives his view of Grünbaum's critique as the latter appeared in 'The Falsifiability of Theories: Total or Partial? A Contemporary Evaluation of the Duhem–Quine Thesis', later published in M. Wartofsky (ed.), *Boston Studies in the Philosophy of Science*, Vol. I Reidel, Dordrecht, 1963).]

Professor Adolf Grünbaum
Department of Philosophy
University of Pittsburgh
Pittsburgh 13, Pennsylvania

June 1, 1962

Dear Professor Grünbaum:

I have read your paper on the falsifiability of theories with interest. Your claim that the Duhem–Quine thesis, as you call it, is untenable if taken nontrivially, strikes me as persuasive. Certainly it is carefully argued.

For my own part I would say that the thesis as I have used it *is* probably trivial. I haven't advanced it as an interesting thesis as such. I bring it in only in the course of arguing against such notions as that the empirical content of sentences can in general be sorted out distributively, sentence by sentence, or that the understanding of a term can be segregated from collateral information regarding the object. For such purposes I am not concerned even to avoid the trivial extreme of sustaining a law by changing a meaning; for the cleavage between meaning and fact is part of what, in such contexts, I am questioning.

Actually my holism is not as extreme as those brief vague paragraphs at the end of "Two dogmas of empiricism" are bound to sound. See sections 1–3 and 7–10 of *Word and Object*.

Sincerely yours,
W. V. Quine

THOMAS S. KUHN

SCIENTIFIC REVOLUTIONS AS CHANGES OF WORLD VIEW*

Examining the record of past research from the vantage of contemporary historiography, the historian of science may be tempted to exclaim that when paradigms change, the world itself changes with them. Led by a new paradigm, scientists adopt new instruments and look in new places. Even more important, during revolutions scientists see new and different things when looking with familiar instruments in places they have looked before. It is rather as if the professional community had been suddenly transported to another planet where familiar objects are seen in a different light and are joined by unfamiliar ones as well. Of course, nothing of quite that sort does occur: there is no geographical transplantation; outside the laboratory everyday affairs usually continue as before. Nevertheless, paradigm changes do cause scientists to see the world of their research-engagement differently. In so far as their only recourse to that world is through what they see and do, we may want to say that after a revolution scientists are responding to a different world.

It is as elementary prototypes for these transformations of the scientist's world that the familiar demonstrations of a switch in visual gestalt prove so suggestive. What were ducks in the scientist's world before the revolution are rabbits afterwards. The man who first saw the exterior of the box from above later sees its interior from below. Transformations like these, though usually more gradual and almost always irreversible, are common concomitants of scientific training. Looking at a contour map, the student sees lines on paper, the cartographer a picture of a terrain. Looking at a bubble-chamber photograph, the student sees confused and broken lines, the physicist a record of familiar subnuclear events. Only after a number of such transformations of vision does the student become an inhabitant of the scientist's world, seeing what the scientist sees and responding as the scientist does. The world that the student then enters is not, however, fixed once and for all by the nature of the environment, on the one hand, and of science, on the other. Rather, it is determined jointly by the environment and the particular normal-scientific tradition

that the student has been trained to pursue. Therefore, at times of revolution, when the normal-scientific tradition changes, the scientist's perception of his environment must be re-educated – in some familiar situations he must learn to see a new gestalt. After he has done so the world of his research will seem, here and there, incommensurable with the one he had inhabited before. That is another reason why schools guided by different paradigms are always slightly at cross-purposes.

In their most usual form, of course, gestalt experiments illustrate only the nature of perceptual transformations. They tell us nothing about the role of paradigms or of previously assimilated experience in the process of perception. But on that point there is a rich body of psychological literature, much of it stemming from the pioneering work of the Hanover Institute. An experimental subject who puts on goggles fitted with inverting lenses initially sees the entire world upside down. At the start his perceptual apparatus functions as it had been trained to function in the absence of the goggles, and the result is extreme disorientation, an acute personal crisis. But after the subject has begun to learn to deal with his new world, his entire visual field flips over, usually after an intervening period in which vision is simply confused. Thereafter, objects are again seen as they had been before the goggles were put on. The assimilation of a previously anomalous visual field has reacted upon and changed the field itself.[1] Literally as well as metaphorically, the man accustomed to inverting lenses has undergone a revolutionary transformation of vision.

The subjects of the anomalous playing-card experiment discussed in Section VI experienced a quite similar transformation. Until taught by prolonged exposure that the universe contained anomalous cards, they saw only the types of cards for which previous experience had equipped them. Yet once experience had provided the requisite additional categories, they were able to see all anomalous cards on the first inspection long enough to permit any identification at all. Still other experiments demonstrate that the perceived size, color, and so on, of experimentally displayed objects also varies with the subject's previous training and experience.[2] Surveying the rich experimental literature from which these examples are drawn makes one suspect that something like a paradigm is prerequisite to perception itself. What a man sees depends both upon what he looks at and also upon what his previous visual-conceptual experience has taught him to see. In the absence of such training there

can only be, in William James's phrase, "a bloomin' buzzin' confusion".

In recent years several of those concerned with the history of science have found the sorts of experiments described above immensely suggestive. N. R. Hanson, in particular, has used gestalt demonstrations to elaborate some of the same consequences of scientific belief that concern me here.[3] Other colleagues have repeatedly noted that history of science would make better and more coherent sense if one could suppose that scientists occasionally experienced shifts of perception like those described above. Yet, though psychological experiments are suggestive, they cannot, in the nature of the case, be more than that. They do display characteristics of perception that *could* be central to scientific development, but they do not demonstrate that the careful and controlled observation exercised by the research scientist at all partakes of those characteristics. Furthermore, the very nature of these experiments makes any direct demonstration of that point impossible. If historical example is to make these psychological experiments seem relevant, we must first notice the sorts of evidence that we may and may not expect history to provide.

The subject of a gestalt demonstration knows that his perception has shifted because he can make it shift back and forth repeatedly while he holds the same book or piece of paper in his hands. Aware that nothing in his environment has changed, he directs his attention increasingly not to the figure (duck or rabbit) but to the lines on the paper he is looking at. Ultimately he may even learn to see those lines without seeing either of the figures, and he may then say (what he could not legitimately have said earlier) that it is these lines that he really sees but that he sees them alternately *as* a duck and *as* a rabbit. By the same token, the subject of the anomalous card experiment knows (or, more accurately, can be persuaded) that his perception must have shifted because an external authority, the experimenter, assures him that regardless of what he *saw*, he was *looking at* a black five of hearts all the time. In both these cases, as in all similar psychological experiments, the effectiveness of the demonstration depends upon its being analyzable in this way. Unless there were an external standard with respect to which a switch of vision could be demonstrated, no conclusion about alternate perceptual possibilities could be drawn.

With scientific observation, however, the situation is exactly reversed. The scientist can have no recourse above or beyond what he sees with

his eyes and instruments. If there were some higher authority by recourse to which his vision might be shown to have shifted, then that authority would itself become the source of his data, and the behavior of his vision would become a source of problems (as that of the experimental subject is for the psychologist). The same sorts of problems would arise if the scientist could switch back and forth like the subject of the gestalt experiments. The period during which light was "sometimes a wave and sometimes a particle" was a period of crisis – a period when something was wrong – and it ended only with the development of wave mechanics and the realization that light was a self-consistent entity different from both waves and particles. In the sciences, therefore, if perceptual switches accompany paradigm changes, we may not expect scientists to attest to these changes directly. Looking at the moon, the convert to Copernicanism does not say, "I used to see a planet, but now I see a satellite". That locution would imply a sense in which the Ptolemaic system had once been correct. Instead, a convert to the new astronomy says, "I once took the moon to be (or saw the moon as) a planet, but I was mistaken". That sort of statement does recur in the aftermath of scientific revolutions. If it ordinarily disguises a shift of scientific vision or some other mental transformation with the same effect, we may not expect direct testimony about that shift. Rather we must look for indirect and behavioral evidence that the scientist with a new paradigm sees differently from the way he had seen before.

Let us then return to the data and ask what sorts of transformations in the scientist's world the historian who believes in such changes can discover. Sir William Herschel's discovery of Uranus provides a first example and one that closely parallels the anomalous card experiment. On at least seventeen different occasions between 1690 and 1781, a number of astronomers, including several of Europe's most eminent observers, had seen a star in positions that we now suppose must have been occupied at the time by Uranus. One of the best observers in this group had actually seen the star on four successive nights in 1769 without noting the motion that could have suggested another identification. Herschel, when he first observed the same object twelve years later, did so with a much improved telescope of his own manufacture. As a result, he was able to notice an apparent disk-size that was at least unusual for stars. Something was awry, and he therefore postponed identification

pending futher scrutiny. That scrutiny disclosed Uranus' motion among the stars, and Herschel therefore announced that he had seen a new comet! Only several months later, after fruitless attempts to fit the observed motion to a cometary orbit, did Lexell suggest that the orbit was probably planetary.[4] When that suggestion was accepted, there were several fewer stars and one more planet in the world of the professional astronomer. A celestial body that had been observed off and on for almost a century was seen differently after 1781 because, like an anomalous playing card, it could no longer be fitted to the perceptual categories (star or comet) provided by the paradigm that had previously prevailed.

The shift of vision that enabled astronomers to see Uranus, the planet, does not, however, seem to have affected only the perception of that previously observed object. Its consequences were more far-reaching. Probably, though the evidence is equivocal, the minor paradigm change forced by Herschel helped to prepare astronomers for the rapid discovery, after 1801, of the numerous minor planets or asteroids. Because of their small size, these did not display the anomalous magnification that had alerted Herschel. Nevertheless, astronomers prepared to find additional planets were able, with standard instruments, to identify twenty of them in the first fifty years of the nineteenth century.[5] The history of astronomy provides many other examples of paradigm-induced changes in scientific perception, some of them even less equivocal. Can it conceivably be an accident, for example, that Western astronomers first saw change in the previously immutable heavens during the half-century after Copernicus' new paradigm was first proposed? The Chinese, whose cosmological beliefs did not preclude celestial change, had recorded the appearance of many new stars in the heavens at a much earlier date. Also, even without the aid of a telescope, the Chinese had systematically recorded the appearance of sunspots centuries before these were seen by Galileo and his contemporaries.[6] Nor were sunspots and a new star the only examples of celestial change to emerge in the heavens of Western astronomy immediately after Copernicus. Using traditional instruments, some as simple as a piece of thread, late sixteenth-century astronomers repeatedly discovered that comets wandered at will through the space previously reserved for the immutable planets and stars.[7] The very ease and rapidity with which astronomers saw new things when looking at old objects with old instruments may make us wish to say that, after Copernicus,

astronomers lived in a different world. In any case, their research responded as though that were the case.

The preceding examples are selected from astronomy because reports of celestial observation are frequently delivered in a vocabulary consisting of relatively pure observation terms. Only in such reports can we hope to find anything like a full parallelism between the observations of scientists and those of the psychologist's experimental subjects. But we need not insist on so full a parallelism, and we have much to gain by relaxing our standard. If we can be content with the everyday use of the verb 'to see', we may quickly recognize that we have already encountered many other examples of the shifts in scientific perception that accompany paradigm change. The extended use of 'perception' and of 'seeing' will shortly require explicit defense, but let me first illustrate its application in practice.

Look again for a moment at two of our previous examples from the history of electricity. During the seventeenth century, when their research was guided by one or another effluvium theory, electricians repeatedly saw chaff particles rebound from, or fall off, the electrified bodies that had attracted them. At least that is what seventeenth-century observers said they saw, and we have no more reason to doubt their reports of perception than our own. Placed before the same apparatus, a modern observer would see electrostatic repulsion (rather than mechanical or gravitational rebounding), but historically, with one universally ignored exception, electrostatic repulsion was not seen as such until Hauksbee's large-scale apparatus had greatly magnified its effects. Repulsion after contact electrification was, however, only one of many new repulsive effects that Hauksbee saw. Through his researches, rather as in a gestalt switch, repulsion suddenly became *the* fundamental manifestation of electrification, and it was then attraction that needed to be explained.[8] The electrical phenomena visible in the early eighteenth century were both subtler and more varied than those seen by observers in the seventeenth century. Or again, after the assimilation of Franklin's paradigm, the electrician looking at a Leyden jar saw something different from what he had seen before. The device had become a condenser, for which neither the jar shape nor glass was required. Instead, the two conducting coatings – one of which had been no part of the original device – emerged to prominence. As both written discussions and pic-

torial representations gradually attest, two metal plates with a non-conductor between them had become the prototype for the class.[9] Simultaneously, other inductive effects received new descriptions, and still others were noted for the first time.

Shifts of this sort are not restricted to astronomy and electricity. We have already remarked some of the similar transformations of vision that can be drawn from the history of chemistry. Lavoisier, we said, saw oxygen where Priestley had seen dephlogisticated air and where others had seen nothing at all. In learning to see oxygen, however, Lavoisier also had to change his view of many other more familiar substances. He had, for example, to see a compound ore where Priestley and his contemporaries had seen an elementary earth, and there were other such changes besides. At the very least, as a result of discovering oxygen, Lavoisier saw nature differently. And in the absence of some recourse to that hypothetical fixed nature that he "saw differently", the principle of economy will urge us to say that after discovering oxygen Lavoisier worked in a different world.

I shall inquire in a moment about the possibility of avoiding this strange locution, but first we require an additional example of its use, this one deriving from one of the best known parts of the work of Galileo. Since remote antiquity most people have seen one or another heavy body swinging back and forth on a string or chain until it finally comes to rest. To the Aristotelians, who believed that a heavy body is moved by its own nature from a higher position to a state of natural rest at a lower one, the swinging body was simply falling with difficulty. Constrained by the chain, it could achieve rest at its low point only after a tortuous motion and a considerable time. Galileo, on the other hand, looking at the swinging body, saw a pendulum, a body that almost succeeded in repeating the same motion over and over again ad infinitum. And having seen that much, Galileo observed other properties of the pendulum as well and constructed many of the most significant and original parts of his new dynamics around them. From the properties of the pendulum, for example, Galileo derived his only full and sound arguments for the independence of weight and rate of fall, as well as for the relationship between vertical height and terminal velocity of motions down inclined planes.[10] All these natural phenomena he saw differently from the way they had been seen before.

Why did that shift of vision occur? Through Galileo's individual genius, of course. But note that genius does not here manifest itself in more accurate or objective observation of the swinging body. Descriptively, the Aristotelian perception is just as accurate. When Galileo reported that the pendulum's period was independent of amplitude for amplitudes as great as 90°, his view of the pendulum led him to see far more regularity than we can now discover there.[11] Rather, what seems to have been involved was the exploitation by genius of perceptual possibilities made available by a medieval paradigm shift. Galileo was not raised completely as an Aristotelian. On the contrary, he was trained to analyze motions in terms of the impetus theory, a late medieval paradigm which held that the continuing motion of a heavy body is due to an internal power implanted in it by the projector that initiated its motion. Jean Buridan and Nicole Oresme, the fourteenth-century scholastics who brought the impetus theory to its most perfect formulations, are the first men known to have seen in oscillatory motions any part of what Galileo saw there. Buridan describes the motion of a vibrating string as one in which impetus is first implanted when the string is struck; the impetus is next consumed in displacing the string against the resistance of its tension; tension then carries the string back, implanting increasing impetus until the mid-point of motion is reached; after that the impetus displaces the string in the opposite direction, again against the string's tension, and so on in a symmetric process that may continue indefinitely. Later in the century Oresme sketched a similar analysis of the swinging stone in what now appears as the first discussion of a pendulum.[12] His view is clearly very close to the one with which Galileo first approached the pendulum. At least in Oresme's case, and almost certainly in Galileo's as well, it was a view made possible by the transition from the original Aristotelian to the scholastic impetus paradigm for motion. Until that scholastic paradigm was invented, there were no pendulums, but only swinging stones, for the scientist to see. Pendulums were brought into existence by something very like a paradigm-induced gestalt switch.

Do we, however, really need to describe what separates Galileo from Aristotle, or Lavoisier from Priestley, as a transformation of vision? Did these men really *see* different things when *looking at* the same sorts of objects? Is there any legitimate sense in which we can say that they pursued their research in different worlds? Those questions can no longer

be postponed, for there is obviously another and far more usual way to describe all of the historical examples outlined above. Many readers will surely want to say that what changes with a paradigm is only the scientist's interpretation of observations that themselves are fixed once and for all by the nature of the environment and of the perceptual apparatus. On this view, Priestley and Lavoisier both saw oxygen, but they interpreted their observations differently; Aristotle and Galileo both saw pendulums, but they differed in their interpretations of what they both had seen.

Let me say at once that this very usual view of what occurs when scientists change their minds about fundamental matters can be neither all wrong nor a mere mistake. Rather it is an essential part of a philosophical paradigm initiated by Descartes and developed at the same time as Newtonian dynamics. That paradigm has served both science and philosophy well. Its exploitation, like that of dynamics itself, has been fruitful of a fundamental understanding that perhaps could not have been achieved in another way. But as the example of Newtonian dynamics also indicates, even the most striking past success provides no guarantee that crisis can be indefinitely postponed. Today research in parts of philosophy, psychology, linguistics, and even art history, all converge to suggest that the traditional paradigm is somehow askew. That failure to fit is also made increasingly apparent by the historical study of science to which most of our attention is necessarily directed here.

None of these crisis-promoting subjects has yet produced a viable alternate to the traditional epistemological paradigm, but they do begin to suggest what some of that paradigm's characteristics will be. I am, for example, acutely aware of the difficulties created by saying that when Aristotle and Galileo looked at swinging stones, the first saw constrained fall, the second a pendulum. The same difficulties are presented in an even more fundamental form by the opening sentences of this section: though the world does not change with a change of paradigm, the scientist afterward works in a different world. Nevertheless, I am convinced that we must learn to make sense of statements that at least resemble these. What occurs during a scientific revolution is not fully reducible to a reinterpretation of individual and stable data. In the first place, the data are not unequivocally stable. A pendulum is not a falling stone, nor is oxygen dephlogisticated air. Consequently, the data that scientists collect from these diverse objects are, as we shall shortly see, themselves different.

More important, the process by which either the individual or the community makes the transition from constrained fall to the pendulum or from dephlogisticated air to oxygen is not one that resembles interpretation. How could it do so in the absence of fixed data for the scientist to interpret? Rather than being an interpreter, the scientist who embraces a new paradigm is like the man wearing inverting lenses. Confronting the same constellation of objects as before and knowing that he does so, he nevertheless finds them transformed through and through in many of their details.

None of these remarks is intended to indicate that scientists do not characteristically interpret observations and data. On the contrary, Galileo interpreted observations on the pendulum, Aristotle observations on falling stones, Musschenbroek observations on a charge-filled bottle, and Franklin observations on a condenser. But each of these interpretations presupposed a paradigm. They were parts of normal science, an enterprise that, as we have already seen, aims to refine, extend, and articulate a paradigm that is already in existence. Section III provided many examples in which interpretation played a central role. Those examples typify the overwhelming majority of research. In each of them the scientist, by virtue of an accepted paradigm, knew what a datum was, what instruments might be used to retrieve it, and what concepts were relevant to its interpretation. Given a paradigm, interpretation of data is central to the enterprise that explores it.

But that interpretive enterprise – and this was the burden of the paragraph before last – can only articulate a paradigm, not correct it. Paradigms are not corrigible by normal science at all. Instead, as we have already seen, normal science ultimately leads only to the recognition of anomalies and to crises. And these are terminated, not by deliberation and interpretation, but by a relatively sudden and unstructured event like the gestalt switch. Scientists then often speak of the 'scales falling from the eyes' or of the 'lightning flash' that 'inundates' a previously obscure puzzle, enabling its components to be seen in a new way that for the first time permits its solution. On other occasions the relevant illumination comes in sleep.[13] No ordinary sense of the term 'interpretation' fits these flashes of intuition through which a new paradigm is born. Though such intuitions depend upon the experience, both anomalous and congruent, gained with the old paradigm, they are not logically or piecemeal linked

to particular items of that experience as an interpretation would be. Instead, they gather up large portions of that experience and transform them to the rather different bundle of experience that will thereafter be linked piecemeal to the new paradigm but not to the old.

To learn more about what these differences in experience can be, return for a moment to Aristotle, Galileo, and the pendulum. What data did the interaction of their different paradigms and their common environment make accessible to each of them? Seeing constrained fall, the Aristotelian would measure (or at least discuss – the Aristotelian seldom measured) the weight of the stone, the vertical height to which it had been raised, and the time required for it to achieve rest. Together with the resistance of the medium, these were the conceptual categories deployed by Aristotelian science when dealing with a falling body.[14] Normal research guided by them could not have produced the laws that Galileo discovered. It could only – and by another route it did – lead to the series of crises from which Galileo's view of the swinging stone emerged. As a result of those crises and of other intellectual changes besides, Galileo saw the swinging stone quite differently. Archimedes' work on floating bodies made the medium non-essential; the impetus theory rendered the motion symmetrical and enduring; and Neoplatonism directed Galileo's attention to the motion's circular form.[15] He therefore measured only weight, radius, angular displacement, and time per swing, which were precisely the data that could be interpreted to yield Galileo's laws for the pendulum. In the event, interpretation proved almost unnecessary. Given Galileo's paradigms, pendulum-like regularities were very nearly accessible to inspection. How else are we to account for Galileo's discovery that the bob's period is entirely independent of amplitude, a discovery that the normal science stemming from Galileo had to eradicate and that we are quite unable to document today. Regularities that could not have existed for an Aristotelian (and that are, in fact, nowhere precisely exemplified by nature) were consequences of immediate experience for the man who saw the swinging stone as Galileo did.

Perhaps that example is too fanciful since the Aristotelians recorded no discussions of swinging stones. On their paradigm it was an extraordinarily complex phenomenon. But the Aristotelians did discuss the simpler case, stones falling without uncommon constraints, and the same differences of vision are apparent there. Contemplating a falling stone,

Aristotle saw a change of state rather than a process. For him the relevant measures of a motion were therefore total distance covered and total time elapsed, parameters which yield what we should now call not speed but average speed.[16] Similarly, because the stone was impelled by its nature to reach its final resting point, Aristotle saw the relevant distance parameter at any instant during the motion as the distance *to* the final end point rather than as that *from* the origin of motion.[17] Those conceptual parameters underlie and give sense to most of his well-known 'laws of motion'. Partly through the impetus paradigm, however, and partly through a doctrine known as the latitude of forms, scholastic criticism changed this way of viewing motion. A stone moved by impetus gained more and more of it while receding from its starting point; distance from rather than distance to therefore became the revelant parameter. In addition, Aristotle's notion of speed was bifurcated by the scholastics into concepts that soon after Galileo became our average speed and instantaneous speed. But when seen through the paradigm of which these conceptions were a part, the falling stone, like the pendulum, exhibited its governing laws almost on inspection. Galileo was not one of the first men to suggest that stones fall with a uniformly accelerated motion.[18] Furthermore, he had developed his theorem on this subject together with many of its consequences before he experimented with an inclined plane. That theorem was another one of the network of new regularities accessible to genius in the world determined jointly by nature and by the paradigms upon which Galileo and his contemporaries had been raised. Living in that world, Galileo could still, when he chose, explain why Aristotle had seen what he did. Nevertheless, the immediate content of Galileo's experience with falling stones was not what Aristotle's had been.

It is, of course, by no means clear that we need be so concerned with 'immediate experience' – that is, with the perceptual features that a paradigm so highlights that they surrender their regularities almost upon inspection. Those features must obviously change with the scientist's commitments to paradigms, but they are far from what we ordinarily have in mind when we speak of the raw data or the brute experience from which scientific research is reputed to proceed. Perhaps immediate experience should be set aside as fluid, and we should discuss instead the concrete operations and measurements that the scientist performs in

his laboratory. Or perhaps the analysis should be carried further still from the immediately given. It might, for example, be conducted in terms of some neutral observation-language, perhaps one designed to conform to the retinal imprints that mediate what the scientist sees. Only in one of these ways can we hope to retrive a realm in which experience is again stable once and for all – in which the pendulum and constrained fall are not different perceptions but rather different interpretations of the unequivocal data provided by observation of a swinging stone.

But is sensory experience fixed and neutral? Are theories simply man-made interpretations of given data? The epistemological viewpoint that has most often guided Western philosophy for three centuries dictates an immediate and unequivocal, Yes! In the absence of a developed alternative, I find it impossible to relinquish entirely that viewpoint. Yet it no longer functions effectively, and the attempts to make it do so through the introduction of a neutral language of observations now seem to me hopeless.

The operations and measurements that a scientist undertakes in the laboratory are not 'the given' of experience but rather 'the collected with difficulty'. They are not what the scientist sees – at least not before his research is well advanced and his attention focused. Rather, they are concrete indices to the content of more elementary perceptions, and as such they are selected for the close scrutiny of normal research only because they promise opportunity for the fruitful elaboration of an accepted paradigm. Far more clearly than the immediate experience from which they in part derive, operations and measurements are paradigm-determined. Science does not deal in all possible laboratory manipulations. Instead, it selects those relevant to the juxtaposition of a paradigm with the immediate experience that that paradigm has partially determined. As a result, scientists with different paradigms engage in different concrete laboratory manipulations. The measurements to be performed on a pendulum are not the ones relevant to a case of constrained fall. Nor are the operations relevant for the elucidation of oxygen's properties uniformly the same as those required when investigating the characteristics of dephlogisticated air.

As for a pure observation-language, perhaps one will yet be devised. But three centuries after Descartes our hope for such an eventuality still depends exclusively upon a theory of perception and of the mind. And

modern psychological experimentation is rapidly proliferating phenomena with which that theory can scarcely deal. The duck-rabbit shows that two men with the same retinal impressions can see different things; the inverting lenses show that two men with different retinal impressions can see the same thing. Psychology supplies a great deal of other evidence to the same effect, and the doubts that derive from it are readily reinforced by the history of attempts to exhibit an actual language of observation. No current attempt to achieve that end has yet come close to a generally applicable language of pure percepts. And those attempts that come closest share one characteristic that strongly reinforces several of this essay's main theses. From the start they presuppose a paradigm, taken either from a current scientific theory or from some fraction of everyday discourse, and they then try to eliminate from it all non-logical and non-perceptual terms. In a few realms of discourse this effort has been carried very far and with fascinating results. There can be no question that efforts of this sort are worth pursuing. But their result is a language that – like those employed in the sciences – embodies a host of expectations about nature and fails to function the moment these expectations are violated. Nelson Goodman makes exactly this point in describing the aims of his *Structure of Appearance*: "It is fortunate that nothing more [than phenomena known to exist] is in question; for the notion of 'possible' cases, of cases that do not exist but might have existed, is far from clear".[19] No language thus restricted to reporting a world fully known in advance can produce mere neutral and objective reports on 'the given". Philosophical investigation has not yet provided even a hint of what a language able to do that would be like.

Under these circumstances we may at least suspect that scientists are right in principle as well as in practice when they treat oxygen and pendulums (and perhaps also atoms and electrons) as the fundamental ingredients of their immediate experience. As a result of the paradigm-embodied experience of the race, the culture, and, finally, the profession, the world of the scientist has come to be populated with planets and pendulums, condensers and compound ores, and other such bodies besides. Compared with these objects of perception, both meter stick readings and retinal imprints are elaborate constructs to which experience has direct access only when the scientist, for the special purposes of his research, arranges that one or the other should do so. This is not to sug-

gest that pendulums, for example, are the only things a scientist could possibly see when looking at a swinging stone. (We have already noted that members of another scientific community could see constrained fall.) But it is to suggest that the scientist who looks at a swinging stone can have no experience that is in principle more elementary than seeing a pendulum. The alternative is not some hypothetical 'fixed' vision, but vision through another paradigm, one which makes the swinging stone something else.

All of this may seem more reasonable if we again remember that neither scientists nor laymen learn to see the world piecemeal or item by item. Except when all the conceptual and manipulative categories are prepared in advance – e.g., for the discovery of an additional transuranic element or for catching sight of a new house – both scientists and laymen sort out whole areas together from the flux of experience. The child who transfers the word 'mama' from all humans to all females and then to his mother is not just learning what 'mama' means or who his mother is. Simultaneously he is learning some of the differences between males and females as well as something about the ways in which all but one female will behave toward him. His reactions, expectations, and beliefs – indeed, much of his perceived world – change accordingly. By the same token, the Copernicans who denied its traditional title 'planet' to the sun were not only learning what 'planet' meant or what the sun was. Instead, they were changing the meaning of 'planet' so that it could continue to make useful distinctions in a world where all celestial bodies, not just the sun, were seen differently from the way they had been seen before. The same point could be made about any of our earlier examples. To see oxygen instead of dephlogisticated air, the condenser instead of the Leyden jar, or the pendulum instead of constrained fall, was only one part of an integrated shift in the scientist's vision of a great many related chemical, electrical, or dynamical phenomena. Paradigms determine large areas of experience at the same time.

It is, however, only after experience has been thus determined that the search for an operational definition or a pure observation-language can begin. The scientist or philosopher who asks what measurements or retinal imprints make the pendulum what it is must already be able to recognize a pendulum when he sees one. If he saw constrained fall instead, his question could not even be asked. And if he saw a pendulum,

but saw it in the same way he saw a tuning fork or an oscillating balance, his question could not be answered. At least it could not be answered in the same way, because it would not be the same question. Therefore, though they are always legitimate and are occasionally extraordinarily fruitful, questions about retinal imprints or about the consequences of particular laboratory manipulations presuppose a world already perceptually and conceptually subdivided in a certain way. In a sense such questions are parts of normal science, for they depend upon the existence of a paradigm and they receive different answers as a result of paradigm change.

To conclude this section, let us henceforth neglect retinal impressions and again restrict attention to the laboratory operations that provide the scientist with concrete though fragmentary indices to what he has already seen. One way in which such laboratory operations change with paradigms has already been observed repeatedly. After a scientific revolution many old measurements and manipulations become irrelevant and are replaced by others instead. One does not apply all the same tests to oxygen as to dephlogisticated air. But changes of this sort are never total. Whatever he may then see, the scientist after a revolution is still looking at the same world. Furthermore, though he may previously have employed them differently, much of his language and most of his laboratory instruments are still the same as they were before. As a result, postrevolutionary science invariably includes many of the same manipulations, performed with the same instruments and described in the same terms, as its prerevolutionary predecessor. If these enduring manipulations have been changed at all, the change must lie either in their relation to the paradigm or in their concrete results. I now suggest, by the introduction of one last new example, that both these sorts of changes occur. Examining the work of Dalton and his contemporaries, we shall discover that one and the same operation, when it attaches to nature through a different paradigm, can become an index to a quite different aspect of nature's regularity. In addition, we shall see that occasionally the old manipulation in its new role will yield different concrete results.

Throughout much of the eighteenth century and into the nineteenth, European chemists almost universally believed that the elementary atoms of which all chemical species consisted were held together by forces of mutual affinity. Thus a lump of silver cohered because of the forces of

affinity between silver corpuscles (until after Lavoisier these corpuscles were themselves thought of as compounded from still more elementary particles). On the same theory silver dissolved in acid (or salt in water) because the particles of acid attracted those of silver (or the particles of water attracted those of salt) more strongly than particles of these solutes attracted each other. Or again, copper would dissolve in the silver solution and precipitate silver, because the copper-acid affinity was greater than the affinity of acid for silver. A great many other phenomena were explained in the same way. In the eighteenth century the theory of elective affinity was an admirable chemical paradigm, widely and sometimes fruitfully deployed in the design and analysis of chemical experimentation.[20]

Affinity theory, however, drew the line separating physical mixtures from chemical compounds in a way that has become unfamiliar since the assimilation of Dalton's work. Eighteenth-century chemists did recognize two sorts of processes. When mixing produced heat, light, effervescence or something else of the sort, chemical union was seen to have taken place. If, on the other hand, the particles in the mixture could be distinguished by eye or mechanically separated, there was only physical mixture. But in the very large number of intermediate cases – salt in water, alloys, glass, oxygen in the atmosphere, and so on – these crude criteria were of little use. Guided by their paradigm, most chemists viewed this entire intermediate range as chemical, because the processes of which it consisted were all governed by forces of the same sort. Salt in water or oxygen in nitrogen was just as much an example of chemical combination as was the combination produced by oxidizing copper. The arguments for viewing solutions as compounds were very strong. Affinity theory itself was well attested. Besides, the formation of a compound accounted for a solution's observed homogeneity. If, for example, oxygen and nitrogen were only mixed and not combined in the atmosphere, then the heavier gas, oxygen, should settle to the bottom. Dalton, who took the atmosphere to be a mixture, was never satisfactorily able to explain oxygen's failure to do so. The assimilation of his atomic theory ultimately created an anomaly where there had been none before.[21]

One is tempted to say that the chemists who viewed solutions as compounds differed from their successors only over a matter of definition. In one sense that may have been the case. But that sense is not the one

that makes definitions mere conventional conveniences. In the eighteenth century mixtures were not fully distinguished from compounds by operational tests, and perhaps they could not have been. Even if chemists had looked for such tests, they would have sought criteria that made the solution a compound. The mixture-compound distinction was part of their paradigm – part of the way they viewed their whole field of research – and as such it was prior to any particular laboratory test, though not to the accumulated experience of chemistry as a whole.

But while chemistry was viewed in this way, chemical phenomena exemplified laws different from those that emerged with the assimilation of Dalton's new paradigm. In particular, while solutions remained compounds, no amount of chemical experimentation could by itself have produced the law of fixed proportions. At the end of the eighteenth century it was widely known that *some* compounds ordinarily contained fixed proportions by weight of their constituents. For some categories of reactions the German chemist Richter had even noted the further regularities now embraced by the law of chemical equivalents.[22] But no chemist made use of these regularities except in recipes, and no one until almost the end of the century thought of generalizing them. Given the obvious counterinstances, like glass or like salt in water, no generalization was possible without an abandonment of affinity theory and a reconceptualization of the boundaries of the chemist's domain. That consequence became explicit at the very end of the century in a famous debate between the French chemists Proust and Berthollet. The first claimed that all chemical reactions occurred in fixed proportion, the latter that they did not. Each collected impressive experimental evidence for his view. Nevertheless, the two men necessarily talked through each other, and their debate was entirely inconclusive. Where Berthollet saw a compound that could vary in proportion, Proust saw only a physical mixture.[23] To that issue neither experiment nor a change of definitional convention could be relevant. The two men were as fundamentally at cross-purposes as Galileo and Aristotle had been.

This was the situation during the years when John Dalton undertook the investigations that led finally to his famous chemical atomic theory. But until the very last stages of those investigations, Dalton was neither a chemist nor interested in chemistry. Instead, he was a meteorologist investigating the, for him, physical problems of the absorption of gases

by water and of water by the atmosphere. Partly because his training was in a different specialty and partly because of his own work in that specialty, he approached these problems with a paradigm different from that of contemporary chemists. In particular, he viewed the mixture of gases or the absorption of a gas in water as a physical process, one in which forces of affinity played no part. To him, therefore, the observed homogeneity of solutions was a problem, but one which he thought he could solve if he could determine the relative sizes and weights of the various atomic particles in his experimental mixtures. It was to determine these sizes and weights that Dalton finally turned to chemistry, supposing from the start that, in the restricted range of reactions that he took to be chemical, atoms could only combine one-to-one or in some other simple whole-number ratio.[24] That natural assumption did enable him to determine the sizes and weights of elementary particles, but it also made the law of constant proportion a tautology. For Dalton, any reaction in which the ingredients did not enter in fixed proportion was *ipso facto* not a purely chemical process. A law that experiment could not have established before Dalton's work, became, once that work was accepted, a constitutive principle that no single set of chemical measurements could have upset. As a result of what is perhaps our fullest example of a scientific revolution, the same chemical manipulations assumed a relationship to chemical generalization very different from the one they had had before.

Needless to say, Dalton's conclusions were widely attacked when first announced. Berthollet, in particular, was never convinced. Considering the nature of the issue, he need not have been. But to most chemists Dalton's new paradigm proved convincing where Proust's had not been, for it had implications far wider and more important than a new criterion for distinguishing a mixture from a compound. If, for example, atoms could combine chemically only in simple whole-number ratios, then a re-examination of existing chemical data should disclose examples of multiple as well as of fixed proportions. Chemists stopped writing that the two oxides of, say, carbon contained 56 per cent and 72 per cent of oxygen by weight; instead they wrote that one weight of carbon would combine either with 1.3 or with 2.6 weights of oxygen. When the results of old manipulations were recorded in this way, a 2:1 ratio leaped to the eye; and this occurred in the analysis of many well-known reactions and

of new ones besides. In addition, Dalton's paradigm made it possible to assimilate Richter's work and to see its full generality. Also, it suggested new experiments, particularly those of Gay-Lussac on combining volumes, and these yielded still other regularities, ones that chemists had not previously dreamed of. What chemists took from Dalton was not new experimental laws but a new way of practicing chemistry (he himself called it the "new system of chemical philosophy"), and this proved so rapidly fruitful that only a few of the older chemists in France and Britain were able to resist it.[25] As a result, chemists came to live in a world where reactions behaved quite differently from the way they had before.

As all this went on, one other typical and very important change occurred. Here and there the very numerical data of chemistry began to shift. When Dalton first searched the chemical literature for data to support his physical theory, he found some records of reactions that fitted, but he can scarcely have avoided finding others that did not. Proust's own measurements on the two oxides of copper yielded, for example, an oxygen weight-ratio of 1.47:1 rather than the 2:1 demanded by the atomic theory; and Proust is just the man who might have been expected to achieve the Daltonian ratio.[26] He was, that is, a fine experimentalist, and his view of the relation between mixtures and compounds was very close to Dalton's. But it is hard to make nature fit a paradigm. That is why the puzzles of normal science are so challenging and also why measurements undertaken without a paradigm so seldom lead to any conclusions at all. Chemists could not, therefore, simply accept Dalton's theory on the evidence, for much of that was still negative. Instead, even after accepting the theory, they had still to beat nature into line, a process which, in the event, took almost another generation. When it was done, even the percentage composition of well-known compounds was different. The data themselves had changed. That is the last of the senses in which we may want to say that after a revolution scientists work in a different world.

NOTES

* Chapter X of *The Structure of Scientific Revolutions*, by Thomas S. Kuhn. Copyright © 1962, 1970 by The University of Chicago. Reprinted by permission.

[1] The original experiments were by George M. Stratton, 'Vision without Inversion of the Retinal Image', *Psychological Review* IV (1897), 341–60, 463–81. A more up-to-date review is provided by Harvey A. Carr, *An Introduction to Space Perception*, New York, 1935, pp. 18–57.

[2] For examples, see Albert H. Hastorf, 'The Influence of Suggestion on the Relationship between Stimulus Size and Perceived Distance', *Journal of Psychology* **XXIX** (1950), 195–217; and Jerome S. Bruner, Leo Postman, and John Rodrigues, 'Expectations and the Perception of Color', *American Journal of Psychology* **LXIV** (1951), 216–27.

[3] N. R. Hanson, *Patterns of Discovery*, Cambridge, 1958, Chapter i.

[4] Peter Doig, *A Concise History of Astronomy*, London, 1950, pp. 115–16.

[5] Rudolph Wolf, *Geschichte der Astronomie*, Munich, 1877, pp. 513–15, 683–93. Notice particularly how difficult Wolf's account makes it to explain these discoveries as a consequence of Bode's Law.

[6] Joseph Needham, *Science and Civilization in China*, III, Cambridge, 1959, 423–29, 434–36.

[7] T. S. Kuhn, *The Copernican Revolution*, Cambridge, Mass., 1957, pp. 206–9.

[8] Duane Roller and Duane H. D. Roller, *The Development of the Concept of Electric Charge*, Cambridge, Mass., 1954, pp. 21–29.

[9] See the discussion in Section VII [Chapter VII of *The Structure of Scientific Revolutions* – ed.] and the literature to which the reference there cited in Note 9 will lead.

[10] Galileo Galilei, *Dialogues concerning Two New Sciences*, trans. by H. Crew and A. de Salvio, Evanston, Ill., 1946, pp. 80–81, 162–66.

[11] *Ibid.*, pp. 91–94, 244.

[12] M. Clagett, *The Science of Mechanics in the Middle Ages*, Madison, Wis., 1959, pp. 537–38, 570.

[13] [Jacques] Hadamard, *Subconscient intuition, et logique dans la recherche scientifique (Conférence faite au Palais de la Découverte le 8 Décembre 1945* [Alençon, n.d.]), pp. 7–8. A much fuller account, though one exclusively restricted to mathematical innovations, is the same author's *The Psychology of Invention in the Mathematical Field*, Princeton, 1949.

[14] T. S. Kuhn, 'A Function for Thought Experiments', in *Mélanges Alexandre Koyré*, ed. R. Taton and I. B. Cohen, to be published by Hermann, Paris, in 1963.

[15] A. Koyré, *Etudes Galiléennes*, Paris, 1939, I, 46–51; and 'Galileo and Plato', *Journal of the History of Ideas* IV (1943), 400–428.

[16] Kuhn, 'A Function for Thought Experiments', in *Mélanges Alexandre Koyré* (see n. 14 for full citation).

[17] Koyré, *Etudes...*, II, 7–11.

[18] Clagett, *op. cit.*, Chaps. iv, vi, and ix.

[19] N. Goodman, *The Structure of Appearance*, Cambridge, Mass., 1951, pp. 4–5. The passage is worth quoting more extensively: "If all and only those residents of Wilmington in 1947 that weigh between 175 and 180 pounds have red hair, then 'red-haired 1947 resident of Wilmington' and '1947 resident of Wilmington weighing between 175 and 180 pounds' may be joined in a constructional definition... The question whether there 'might have been' someone to whom one but not the other of these predicates would apply has no bearing... once we have determined that there is no such person.... It is fortunate that nothing more is in question; for the notion of 'possible' cases, of cases that do not exist but might have existed, is far from clear".

[20] H. Metzger, *Newton, Stahl, Boerhaave et la doctrine chimique*, Paris, 1930, pp. 34–68.

[21] *Ibid.*, pp. 124–29, 139–48. For Dalton, see Leonard K. Nash, *The Atomic-Molecular Theory*, 'Harvard Case Histories in Experimental Science', Case 4; Cambridge, Mass., 1950, pp. 14–21.

[22] J. R. Partington, *A Short History of Chemistry*, 2d ed.; London, 1951, pp. 161–63.

[23] A. N. Meldrum, 'The Development of the Atomic Theory: (1) Berthollet's Doctrine of Variable Proportions', *Manchester Memoirs* **LIV** (1910), 1–16.
[24] L. K. Nash, 'The Origin of Dalton's Chemical Atomic Theory', *Isis* **XLVII** (1956), 101–16.
[25] A. N. Meldrum, 'The Development of the Atomic Theory: (6) The Reception Accorded to the Theory Advocated by Dalton', *Manchester Memoirs* **LV** (1911), 1–10.
[26] For Proust, see Meldrum, 'Berthollet's Doctrine of Variable Proportions', *Manchester Memoirs* **LIV** (1910), 8. The detailed history of the gradual changes in measurements of chemical composition and of atomic weights has yet to be written, but Partington, *op. cit.*, provides many useful leads to it.

LAURENS LAUDAN

GRÜNBAUM ON 'THE DUHEMIAN ARGUMENT'*

In several recent publications[1], Professor Adolf Grünbaum has inveighed against the conventionalism of writers like Einstein, Poincaré, Quine and especially Duhem. Specifically, Grünbaum has assailed the view that a single hypothesis can never be conclusively falsified. Grünbaum claims that the conventionalists' insistence on the immunity of hypotheses from falsification is neither logically valid nor scientifically sound. Directing the weight of his argument against Duhem, Grünbaum launches a two-pronged attack. He insists, first, that conclusive falsifying experiments are possible, suggesting that Duhem's denial of such experiments is a logical non-sequitur. He then proceeds to show that, more than being merely possible, crucial falsifying experiments have occurred in physics. I do not intend to make a logical point against Grünbaum's critique so much as an historical and exegetical one. Put briefly, I believe that he has misconstrued Duhem's views on falsifiability and that the logical blunder which he discussed should not be ascribed to Duhem, but rather to those who have made Duhem's conventionalism into the doctrine which Grünbaum attacks. Whether there are any writers who accept the view he imputes to Duhem, or whether he is exploiting 'straw-men' to give weight to an otherwise trivial argument is an open question. For now, I simply want to suggest that his salvos are wrongly directed against Duhem.

The *locus classicus* for Grünbaum's interpretation is Duhem's discussion of falsifiability in his *Aim and Structure of Physical Theory*.[2] I submit that a careful analysis of the historical and textual context of Duhem's account of crucial experiments will indicate how far Grünbaum's argument misses the mark. In the *Aim and Structure*, Duhem was preoccupied with, and disturbed by, the naive realism with which most scientists of the late nineteenth century discussed their theories. They were firmly convinced that science was a search for truth and though they realized that theories could not be proved by verification, they still believed that by suitably eliminating (i.e., falsifying) rival hypotheses,

they could ultimately discover the residual true one. Experiments of this type, which simultaneously refuted one hypothesis and thereby verified another (supposedly its only logical alternative), were called crucial experiments (from Bacon's '*instantiae crucis*'). Though the phrase 'crucial experiment' dates from the seventeenth century, it was really the nineteenth century which adopted it as its own. In nineteenth century chemistry, geology, optics and thermodynamics, much use was made of what were thought to be crucial experiments. The notion of a crucial experiment, and the rigorous empiricism which accompanied it, were particularly common among those scientists who saw themselves in the experimental tradition of Newton, Bacon and Lavoisier. But the phrase 'crucial experiment' enjoyed even wider currency in nineteenth century philosophies of science than in science itself; perhaps because it was so compatible with that century's view of scientific progress as the evolution from 'false' theories to 'true' ones, crucial experiments being the instrument whereby the wrong were separated from the right. It was a dogma of research that some theories were really true and the crucial experiment offered itself as the unerring probe for finding the truth. The doctrine had an appealing, not to say appalling, simplicity.

In his incisive critique of this doctrine, Duhem pointed out that simultaneous falsification and verification required that two conditions be satisfied: (1) that an unambiguous falsification procedure exist and (2) that *reductio ad absurdum* methods be applicable to scientific inference. Duhem argued that neither of these conditions could be fulfilled. Disposing first of conclusive falsification, he pointed out that the alleged refutation of an hypothesis, H, by an observation, $\sim O$, presupposes that scientific reasoning follows the simple schema, $H \rightarrow O$. Falsification could then be represented by a modus tollens argument of the form, $[(H \rightarrow O) \cdot \sim O] \rightarrow \sim H$. Yet this is rarely, if ever, the structure of argument in the sciences. Every prediction the scientist makes is based, not on a single hypothesis, but on several – often tacit – assumptions and rules of inference. Since a falsified prediction, O, is a consequence of the conjunction of several hypotheses (i.e., $(H_1 + H_2 + \cdots + H_n) \rightarrow O$), we have no right to single out one of these hypotheses as the false one. All we are entitled to infer from a successful falsification is that the antecedent conjunction $(H_1 + H_2 + \cdots + H_n)$ is false. Beyond this, we can go no further; isolated hypotheses are immune from refutation. To maintain, as

J. F. Herschel did, that each of the hypotheses of classical mechanics can be independently tested is to misconstrue and oversimplify the relation of theory to experiment.

Duhem's corollary argument further strenghened his denial of crucial experiments. Even if one could, *per impossibile*, falsify a specific hypothesis (e.g., light is corpuscular), one has not thereby proved the truth of any alternative hypothesis (e.g., light is undular). There always remains the possibility of (1) future falsification of the alternative hypothesis and (2) the discovery of presently unknown explanations more satisfactory than the alternative. The only truth established by the falsification of H (assuming, of course, that one could falsify H) is $\sim H$, which is not an hypothesis, but a potentially infinite disjunction of hypotheses. Unlike the mathematician who can often exhaustively enumerate all conceivable cases, the physicist cannot list all the possible alternative hypotheses for explaining an event. If and only if he could would *reductio* methods be applicable. The effect of Duhem's dual attack on cruciality was to convince most philosophers of the impossibility of conclusive falsification. Reichenbach, Frank, Hanson, Quine and Toulmin have been among those who have sought to reinforce Duhem's position.

Recently, however, Professor Grünbaum has asserted that philosophers have been misled by Duhem's denial of crucial falsifying experiments[3] because his doctrine is both logically mistaken and a misrepresentation of the role of experiment in physics. Let us consider Grünbaum's two points in turn.

1. Grünbaum's first argument runs as follows: Duhem's denial of conclusive falsification rests on the assumption that for every hypothesis, H, there always exists a set of auxiliary assumptions, A', such that any observations whatever, O, are compatible with, and deducible from, the conjunction of H and A'. Schematically,

$$(H)\,(O)\,(\exists A')\,(H + A' \rightarrow O).$$

But, asks Grünbaum, what assurance have we that such an hypothesis-saver as A' will always exist? While there are occasions when a resourceful and imaginative scientist can save an apparently refuted hypothesis by modifying other assumptions, Duhem has given us no guarantee that such modifications will work for every hypothesis.[4] Unless we have a proof that an appropriate non-trivial A' exists for every H and O (a

proof that Duhem never offered), then we need not believe that every falsification is inconclusive. As Grünbaum summarizes his position:

> the failure of $\sim O$ to permit the deduction of $\sim H$ does *not* justify the assertion of Duhem's thesis that there always *exists* an A' such that the conjunction of H and A' entails $\sim O$![5]

There is much to be said about, though less to be said for, Grünbaum's argument. Because I think it stems from a misinterpretation of Duhem's text, it is not irrelevant to note that Grünbaum systematically avoids references to, or direct discussions of, the *Aim and Structure*. By conflating the views of Quine, Einstein, Weyl and Duhem, he sets out to refute what he variously calls "Duhem's conception as articulated by W. V. O. Quine" (*Ibid.*, 106), "Einstein's geometrical form of the Duhem thesis" (*Ibid.*, 135) and "Weyl's Duhemian thesis" (*Ibid.*). I question whether it is legitimate to assimilate Duhem's views to those of these other writers. In particular, I want to argue that Duhem did not make the logical blunder of asserting what Grünbaum calls the 'Duhem-thesis'; that is, Duhem does not maintain that

$$(H)(O)(\exists A')(H + A' \rightarrow O).$$

Duhem's position is a much milder and more modest one than Grünbaum's formulation of the D-thesis would lead us to believe. We must remember that Duhem wanted to show, not that falsification never occurred, but that such falsification was necessarily ambiguous. Given that H and A entail O and that $\sim O$, we can surely infer $\sim(H+A)$, though we have no reason to make an unequivocal denial of H rather than A. But to say that H is not refuted by $\sim O$ is certainly not to make the stronger claim that $(\exists A')(H+A' \rightarrow O)$. Duhem is not asserting that every hypothesis can be saved, but only that unless one had proved that it cannot be saved, then it is not falsified.

Duhem believed that the apparent asymmetry between verification and falsification (according to which the former was always inconclusive and the latter always conclusive) was chimerical. In fact, reasoned Duhem, we can no more falsify a single hypothesis than we can verify it. Every experiment calls into play a wide assortment of theoretical assumptions, which become tacit premises of the inference from hypothesis to event. The very meaning of the terms in an hypothesis depends upon

their use in other hypotheses and a theoretical statement cannot be understood except in the context of the conceptual web of which it forms a strand. When a theorist's predictions are falsified, all he knows is that at least one of the many assumptions he used for making the prediction was incorrect. But the experiment cannot single out the wrong premise(s) from the right one(s). There is a semantical linking between hypotheses which vitiates any attempt at testing them in isolation.[6] To continue to maintain H in the face of $\sim O$ is not necessarily to assert that a suitable A' *exists*, but simply to allow for the *possibility* that H may still be compatible with $\sim O$, given some suitable A'. The *onus probandi* is not, as Grünbaum supposes, on the scientist who refuses to call a refuted hypothesis false to show that his hypothesis can be saved by some suitable A'. Rather, the burden of proof is on those who deny H to show that there does not exist an A' which would make H compatible with $\sim O$. Schematically, the scientist who claims to have falsified an hypothesis, H, must prove that

$$\sim (\exists A')\,(H + A' \rightarrow \sim O).$$

Unless such a proof is forthcoming, a scientist is logically justified in seeking some sort of rapproachement between his hypothesis and the uncooperative data. It appears, then, that there are two versions of the D-thesis: a stronger one (which Grünbaum attacks) and a weaker (which I believe is Duhem's actual position). They can be formulated as follows:

Stronger D-thesis: For every hypothesis and every observation statement, there exists a set of non-trivial auxiliary assumptions, A', such that H and A' entail O.

Weaker D-thesis: In the absence of a proof that no appropriate hypothesis-saver exists (i.e., unless we prove that $\sim(\exists A')\,(H + A' \rightarrow \sim O)$), then $\sim O$ is not a conclusive refutation of H, even if $H + A' \rightarrow O$.

Insofar as Grünbaum imputes the stronger claim to Duhem, I think he is mistaken. I have already noted that he cites no textual evidence for his reading of Duhem, nor does he give us any reason to believe that Duhem would have accepted the stronger version. It is also worth observing that Grünbaum himself often formulates the D-thesis as if it were simply the claim that hypotheses cannot be conclusively refuted.[7] Surely, the weaker version is a more faithful translation of this claim than the stronger one.[8] By a subtle assimilation of the weaker D-thesis to the stronger and by then

refuting the stronger one, Grünbaum talks as if he has thereby refuted the weaker D-thesis as well. But the weaker (and, if I am right, authentic) version of the D-thesis is untouched by Grünbaum's critique.

2. We turn now to consider briefly Grünbaum's second attack on the Duhemian–Quinean–Einsteinean position. Here he seeks to show, by example, that conclusive falsifying experiments do occur. He suggests that we consider the schema

$$[(H + A \rightarrow O)(\bar{O} \cdot A) \rightarrow \bar{H}],$$

where H is a system of geometry, A is a proposition about the perturbation-free characteristics of solid rods, and O is the empirical statement that light rays coincide with rigid-body geodesics. We need not probe into the physics of the problem to understand the substance of Grünbaum's counterexample. He argues, and I think we can safely grant, that we have observed $\sim O$ and, in virtue of independent evidence, can assert with high probability that A is true. On these grounds, Grünbaum claims to have shown that the experiment has falsified the hypothesis H. On the face of it, this is a clear exception to the D-thesis. But there are two factors here which make this argument powerless against Duhem's position. In the first place, a system of geometry (the 'H' in this example) is not the sort of thing which counts for Duhem as an 'isolated hypothesis'. Duhem, we recall, insisted only that isolated hypotheses – not systems of hypotheses such as geometry or classical mechanics – were non-falsifiable. To say that a *set* of hypotheses has been falsified is no refutation of the D-thesis. The second, and more serious flaw in Grünbaum's counterexample, if I understand it correctly, is that A, though probable, is not known to be true: Despite A's high likelihood, a scientist is not forced to relinquish H unless A is known to be true. Since A is subject to some doubt, we cannot necessarily blame the failure of the prediction, O, on H rather than A. To give up H might be more prudent, but the demands of prudence do not carry logical weight. It is perhaps correct to remark that Grünbaum's experiment would cause a rational person to cease to expound H, but the experiment certainly does not provide an unambiguous falsification of H.

In light of the failure of both Grünbaum's logical analysis and his counterexample to make out a case for conclusive falsification, it seems we are still left with Duhem's conventionalism intact.

NOTES

* From *Philosophy of Science*, Vol. 32 (1965). Reprinted by permission.

[1] Cf. Grünbaum's 'The Duhemian Argument', *Philosophy of Science* 27 (1960), 75–87; 'Laws and Conventions in Physical Theory', in *Current Issues in the Philosophy of Science* (ed. Feigl & Maxwell), pp. 140–155 and 161–168; *Philosophical Problems of Space and Time*, pp. 106–152; and 'The Falsifiability of Theories: Total or Partial? A Contemporary Analysis of the Duhem–Quine Thesis', in *Boston Studies in the Philosophy of Science* (ed. Wartofsky).

[2] Cf. Duhem's *Aim and Structure of Physical Theory* (trans. by Wiener), Part II, Chapter vi and *passim*. The translation is based on the 1914 edition of *La Théorie Physique: Son Objet – Sa Structure*.

[3] Grünbaum agrees that crucial experiments cannot verify hypotheses. Like Duhem, he is opposed to the Baconian theory of crucial experiments. But, like Popper and against Duhem, he wants to assert that crucial falsifying experiments occur.

[4] Grünbaum readily grants that by suitable logical gyrations, or by a re-definition of terms, one could always find *ad hoc* some A' such that for a given O, $H+A' \rightarrow O$. (The simplest way, of course, would be to take O itself as A'.) But he rightly points out that Duhem's point becomes trivial on such an interpretation. Thus, Grünbaum argues that the Duhemian position requires that there exists a *non-trivial* A' for every H and every O.

[5] *Philosophical Problems of Space and Time*, p. 114.

[6] For example, Duhem writes that "The physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting the group is unacceptable and ought to be modified; but the experiment does not indicate which one ought to be changed". *Aim and Structure*, p. 187.

Elsewhere, he notes that "To seek to separate each of the hypotheses of theoretical physics from the other assumptions on which it rests in order to subject it in isolation to observational tests is to pursue a chimera...", *Ibid.*, p. 200.

[7] Thus, Grünbaum writes: "Duhem's thesis is that the falsifiability of the hypothesis H as an explanans of the actual empirical facts O' is always unavoidably inconclusive". 'Laws and Conventions in Physical Theory', in *Current Issues in the Philosophy of Science*, p. 145.

[8] As further evidence for the 'weaker' reading of Duhem, we might consider his discussion in *Aim and Structure*, Part II, Chapter iv, §10. There he explains that when a theory, T, is falsified, a scientist has two options: he can either modify one of the hypotheses in T to make it compatible with the phenomena, or he may discard the whole structure T and opt for an altogether different theory T'. If Duhem accepted the stronger version of the D-thesis, he presumably would have argued that the former alternative is always open. But he does not take this approach. He does say that we have no "right to condemn in advance the boldness" (p. 217) of the physicist who seeks to preserve his theory in the face of embarrassing evidence. But, on the other hand, Duhem admits that *there is no guarantee that he will be successful in finding an* A'. The scientist who takes this ploy is justified only if he succeeds in "satisfying the requirements of experiment" (p. 217), i.e., only if he finds some A' that saves the theory T. Duhem is not asserting that A's will always be found, or even that they exist, if they are not found. He readily admits that if, after a thorough analysis, no A's are forthcoming, then we should give up the entire theoretical system, T. But in giving it up, we must be careful not to say that the individual hypotheses within the theory have been falsified. The theory is no longer fruitful, but that does not mean that all its component hypotheses are false.

CARLO GIANNONI

QUINE, GRÜNBAUM, AND THE DUHEMIAN THESIS*

Quine in his paper 'Two Dogmas of Empiricism'[1] has propounded a radical conventionalist thesis, arguing that only science as a whole, including the laws of logic, is empirically testable. Grünbaum, on the other hand, has in various places including *Philosophical Problems of Space and Time*[2] been critical of an even moderate Duhemian conventionalism, and in particular attempts to show that the geometry of space is testable independently of other physical theory. Between these two extremes lies the Duhemian thesis. We will attempt to describe in this paper the exact nature and extent of Duhemian conventionalism in physics, indicating the manner in which it is a semantical conventionalism while not being trivially so.

Let us begin by looking at Quine's thesis of conventionalism in physics and logic. The point usually derived from the above mentioned paper is that the analytic-synthetic distinction is untenable. It is important to note that Quine in the first section of the paper admits that logically true statements are analytic, and directs his attack towards a second class of analytic statements which are generally called analytic by virtue of the meanings of the nonlogical words, such as 'No bachelor is married'. If synonyms were substituted for synonyms, these statements would be transformed into logical truths. Quine's main attack in the first four sections is on the lack of an adequate notion of synonymy. One might grant this in regard to natural language and say so much the worse for natural languages.

More interesting, more debatable, but less well defended is Quine's discussion in the last two sections of his paper. If the point which Quine makes here is valid for natural language, then it is equally valid for constructed languages. Essentially Quine is proposing an extension of Duhem's thesis of conventionalism in physics to include the truths of logic as well as *all* of the laws of science.

Let us analyze the structure of his argument. Quine suggests several arguments in defense of the analytic-synthetic distinction in the first four

sections of the paper, and criticizes each severely. He then goes on to suggest a method of defense based on the verification theory of meaning. Can we not say a statement is analytic if it is confirmed by anything whatever that happens in the world? If we can talk about the verification of particular statements, surely we can also talk about those statements which are verified "come what may". Quine's response to this ploy is to say that no individual statement can be verified. He says, "The unit of empirical significance is the whole of science".[3] Only science as a whole can be verified or falsified. *Ipso facto*, no individual statement can be vacuously confirmed by any experience, and the analytic-synthetic distinction becomes of minor interest in comparison to this far-reaching suggestion of Quine's that *no* individual statement is open to verification. If Quine can make this point, he has not only shown that the analytic-synthetic distinction is untenable, but also and even more important, that the true-false distinction is untenable, except as applied to science as a whole, and that the distinction between physics and logic is completely untenable. We anticipate from Quine at this point a defense as strong and as hard hitting as his critique of the analytic-synthetic distinction. But our expectations are not fulfilled. Quine merely gives examples of cases in which we must conventionally choose to reject one statement or another, for the statements jointly are falsifiable, but not individually. He does not, however, give us a philosophical underpinning to justify his general thesis that only science as a whole can be verified or falsified. Quine gives us this much to work with:

> Reevaluation of some statements entails reevaluation of others, because of their logical interconnections – the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reevaluated one statement we must reevaluate some others, which may be statements logically connected with the first or may be statements of logical connections themselves.[4]

One wishes that he had spelled out these logical interconnections in some detail. One also wants to know why he is now rejecting the criteria of logical truth of statements which he held at the beginning of the paper.

It is interesting to note here that a conclusion similar to Quine's was arrived at by Hempel and at about the same time (1951).[5] Because of the problem of theoretical terms Hempel arrived at the conclusion that the unit of empirical significance was not the terms or statements of science, but the theories of science. In contradistinction to Quine, how-

ever, Hempel gave detailed arguments for this position, possibly because he was reluctant to admit it. The theoretical terms of science cannot be explicitly or operationally defined using only observation terms, but rather must be introduced into science by means of the theories themselves. The theory consists of several statements at least two of which must contain the theoretical term. In effect, the theoretical term is implicitly defined by the theory. Moreover, statements can be deduced from the theory which do not contain theoretical terms, and which, in particular, do not contain the theoretical term in question. Because statements are deducible which do not contain theoretical terms, the theory as a whole can be said to have empirical significance, and can be confirmed or falsified. However, since the theoretical statements themselves contain theoretical terms which are only partially interpreted via the theory, they cannot be individually tested. Therefore, theories, by virtue of containing theoretical terms, must face experience as a whole. The unit of empirical significance and the unit which is to be tested must be the theory as a whole.

We see, therefore, that the Duhem-Quine thesis as applied to *theories* can be defended by purely logical analysis of theories and their terms. To defend the general Quinean thesis that all scientific knowledge is conventional we must be able to apply the same logical analysis throughout science. We will not attempt here to justify Quine's whole thesis, but rather we will attempt to defend that segment which is the same as the original Duhemian thesis. Hempel has already in effect given a defense of the thesis for that segment of science which includes theories. We propose now to defend the thesis for one other segment of science, namely, for that part of science which depends on the use of operational definitions which make reference to measuring instruments, and in particular where the measurement is derivative rather than fundamental.

In defending his thesis, Duhem distinguishes between common sense physical laws and the laws of physics as a developed science, for only to the latter is his thesis applicable. The distinction between these laws arises over the fact that physics uses terms which are not directly observable, such as 'mass', 'temperature', and 'pressure'. In the jargon of philosophy of science these are considered to be operationally defined terms. Duhem does not, however, consider these terms as introduced by definition. While one must certainly use instruments to measure these

various quantities, the very use of these instruments implies the belief in certain physical theories, viz., the theories of the instruments themselves.[6] While Duhem would agree with the operationalist that the meaning of these terms depends on the method of measuring them, he would go a step further and say that the meaning of the terms depends also on the theory of the instruments used. In effect, Duhem is treating these terms in the same manner that theoretical terms are currently treated, that is, they are implicitly defined by the theories of the instruments.

Duhem is concerned primarily with quantities which are the subject of derivative measurement rather than fundamental measurement. Of quantities which are derivatively measured we can distinguish three types: (1) Quantities which are *nominally* defined as a mathematical function of other quantities. For example, the density of a substance is defined as the mass of the substance divided by its volume. Thus, in order to measure density one first measures the mass and volume of the substance and then one *calculates* the density. (2) Quantities which are measured by other means than the means originally used in introducing the concept are derivative relative to this method of measurement. For example, length can be fundamentally measured by using meter sticks, but it also can be measured by sending a light beam along the length and back and measuring the time which it takes to complete the round trip. We can then calculate the length by finding the product of the velocity of light (c) and the time. Such a method of measurement is dependent not only on the measurement of time as is the first type of derivative quantity, but also on the law of nature that the velocity of light is a constant equal to c. (3) The third type of derivative quantity, and the type to which the Duhemian thesis is particularly applicable, is exemplified by the ordinary mercury thermometer measurement of temperature. Changes in temperature are correlated with changes in some other quantity, as in the first case of derivative measurement; however, here no significance is given to the absolute length of the strand of mercury. It is rather changes in the length of the strand which are relevant to changes in temperature. Changes in temperature are a function of change in length of the mercury, but temperature itself is not a function of the length. The significant aspect of this type of derivative measurement for the Duhemian thesis is the fact that this type of measurement depends on the assumption of certain laws of nature. While in a certain sense change in temperature is correlated

with change in the length of the strand of mercury *by stipulation*, it is also to a certain extent *a law-like connection*, for there are conditions under which one will either disregard or correct a temperature reading as given by a mercury thermometer. For example, scientists believe that the readings of a mercury thermometer placed in a strong magnetic field are incorrect. Therefore, the reading of a mercury thermometer is not taken at face value.

One might object that it is *part of the stipulation* that such perturbing influences as magnetic fields are to be avoided as a condition for *valid* use of the thermometer. Conditions could be added to the stipulation in either of two ways: one might make a blanket statement that the use of a mercury thermometer for temperature measurement is valid only if *all* perturbing influences are absent. This approach is obviously circular, for a perturbing influence is a condition which makes a measurement incorrect. Therefore we know that the measurement is incorrect only if we know that perturbing influences are present, and we know that perturbing influences are present only if we know that the measurement is incorrect. An alternative procedure which avoids the circularity would be to state specifically which conditions are to be considered perturbing. The difficulty with this latter approach is that at the time at which a particular method of derivative measurement is introduced one does not know all of the conditions which one would want to consider perturbing.

We think, rather, that the use of instruments, such as thermometers, involves both stipulative and causal elements. In using an instrument we assume that there is a law-like connection between the property measured and the property by which it is measured. If you will, change in temperature causes change in length of the mercury strand of the thermometer. As with all causal laws, it is assumed that the connection depends on certain conditions, but it is not usually known at first what all of these conditions are. It is stipulative in that the property which is being measured is not directly observable, for there is a difference between the crude concept of temperature of ordinary sensation and the exact concept of temperature of physics. Referring to the concepts of mass, temperature, and pressure Duhem has the following to say: "These ideas are not only abstract; they are, in addition, symbolic, and the symbols assume meaning only by grace of the physical theories"[7]. By an abstract idea Duhem means any idea which is general as opposed to proper names

which are particular. Our crude concept of temperature is abstract, but not symbolic. By symbolic, Duhem means theoretical, i.e., unobservable. Since the exact concept of temperature is unobservable, one is free to stipulate what relationship one wants to hold between the length of the mercury strand and the temperature. But in this case, the stipulated connection is nonetheless also a causal connection, for it is a connection which varies with external conditions, at least some of which are unknown. This is the heart of Duhemian conventionalism, that is, that there are *laws of nature* which by the nature of the case must be *stipulated to be true.*

Duhem further substantiates his view that the stipulated connection is causal by noting that when several methods are available for measuring a certain property, no one method is taken as the absolute criterion relative to which the other methods are derivative in the second sense of derivative measurement noted above. Each method is used as a check against the others. Duhem discusses an experiment which consists essentially of an electrical battery with a voltmeter, an incandescent lamp and a coil connected in parallel across it.[8] If the voltmeter indicated a certain reading one would say that the battery was operating. However, even if the voltmeter indicated zero, one might still say that the battery was operating based on the fact that the incandescent lamp was glowing and that the coil was warm. Conversely, if the lamp did not glow, one might still say that the battery was operating because gas bubbles were emerging from the battery or the voltmeter was giving a certain reading or the coil was warm. The operationalist likes to think that one method of measurement is the defining criterion for a certain quantity and that other measurements of the quantity are derivative relative to it, but this does not do justice to the way that scientists in fact proceed. The relationship between method of measurement and quantity measured is not one-one but many-one, and the relationship is many-one because to the physicist the measured quantity is causally related to more than one observable phenomenon. The very construction of measuring instruments oftentimes depends on the prior construction of a theory on the basis of which one can predict a number of observable phenomena, each of which is dependent on a single underlying cause and each of which can operate as a measure of the underlying property.

It is a truism that a scientific law can be tested only if all of the extralogical terms are interpreted, at least partially. One method of inter-

preting the symbolic terms of physics, such as temperature, is via measuring instruments, such as thermometers. What is tested then is a physical law together with a particular set of rules of interpretation for its symbolic concepts. If the rules of interpretation are analytic, then any falsification of an interpreted law encountered must be interpreted as a falsification of the law, for an analytic statement cannot be falsified. To insist that any law can be retained by changing the analytic rules of interpretation is to fall into a trivial semantic conventionalism. If Duhem is right, however, in insisting that the rules of interpretation for (some of) the symbolic concepts of physics are law-like, rather than analytic, then a falsification of the interpreted law cannot be taken to be a falsification of the law itself for it may be the interpretation that is false. We then arrive at Duhem's dictum that there are no crucial experiments in physics. This we believe is the extent of Duhem's conventionalism.

Recently Grünbaum has objected to the Duhemian thesis on the grounds that it is an empirical thesis, rather than a logical thesis, and therefore open to falsification, and furthermore that it is in fact false for the case of geometry. *We will attempt to show here that the Duhemian thesis as interpreted by Grünbaum is a logically defensible thesis relative to derivative measurement of the third type*, while agreeing with Grünbaum that it is an *empirical* thesis relative to fundamental measurement. We will furthermore attempt to clarify what we take to be a misunderstanding on the part of Grünbaum of certain aspects of the Duhemian thesis itself.

Let 'H' stand for the hypothesis being tested; 'A' for the interpreting *laws*, (i.e., auxiliary hypotheses); and 'O' for an observational statement. Let us further assume that '$((H \,\&\, A) \rightarrow O) \,\&\, \sim O$' is true. That is, the interpreted hypothesis has been falsified. Grünbaum interprets Duhem as asserting both of the following statements, where 'A'_{nt}' stands for a non-trivial alternative set of interpreting *laws* (i.e., auxiliary hypotheses), and 'H'' for an alternative hypothesis:

$$(\exists A'_{nt})\,(H \,\&\, A'_{nt}) \rightarrow \sim O),$$
$$(\exists H')\,((H' \,\&\, A) \rightarrow \sim O).$$

Duhem would certainly be agreeable to one or the other statement in any particular case *but not both*. His point is not that there is always both an 'A'_{nt}' and an 'H'', but that we cannot determine which it is that exists *independently of the rest of the system*. Within the context of a system,

however, he seems to believe that we can determine which it is. Duhem relates physical science to an organism, and he says that "if something goes wrong ... the physicist will have to ferret out through its effect on the entire system which organ needs to be remedied or modified without the possibility of isolating this organ and examining it apart"[9]. Furthermore, Duhem believes that despite the symbolic nature of the theories of physics, we can achieve a 'natural classification' of the world. When we achieve a system which orders and organizes a vast body of previously heterogeneous data of experience, then "it is impossible for us to believe that this order and this organization are not the reflected image of a real order and organization"[10].

Despite the fact that Duhem would not agree to Grünbaum's interpretation of his thesis, we believe that even the stronger thesis, namely, that *both* an 'H'' and an 'A'_{nt}' exist, is defensible. Let us consider the following situation: Charles' Law which holds that temperature times volume is a constant has been relatively well verified in the past. Nevertheless, we run a further test on it and it turns out false. We might say that the law is generally true but that it has exceptions, and that these exceptions can be excluded if we can determine the peculiar conditions operating in this experiment. We might then proceed to determine why the law does not hold in this particular case, that is, to ascertain the perturbing conditions. On the other hand, according to our previous analysis of temperature measurement we know that the result of a thermometer reading is to be accepted only under certain conditions, some of these conditions being unspecified and unknown. Since we might attribute the falsity of Charles' Law in this particular case to a particular perturbing condition, we might also argue that the perturbing condition has not affected the law but rather the measurement of temperature. Why should we say that perturbing conditions affect the law rather than the method of measurement *or* vice-versa? The parallel is complete between Charles' Law and the law of the thermometer. Both are generally reliable, and both are expected to be unreliable under certain conditions, some of which are unknown. When the pair do in fact turn out false we are free to attribute the falsity to either of the laws and to relate the perturbing condition to that law.

While we think that the strong Duhemian thesis is logically defensible with respect to derivative measurement, with respect to fundamental

measurement it is at best empirically true. While fundamental measurement includes not only length, but also time, mass, etc., we will concern ourselves only with the former. The fundamental measurement of length depends in the first instance on the existence of a measuring rod. Not any measuring rod will do, however, for most, if not all, measuring rods yield incorrect results because of the influence of perturbing conditions, such as heat. Measurement must therefore be performed with a rigid body where, according to Reichenbach, "Rigid bodies are solid bodies which are not affected by differential forces, or concerning which the influence of differential forces has been eliminated by corrections"[11]. The Duhemian thesis is defensible with respect to geometry if there are as a matter of fact no solid bodies uninfluenced by differential forces.

Grünbaum attempts to prove the falsity of the Duhemian thesis in two ways in regard to geometry. One of the proofs depends on the empirical fact that the material invar has a coefficient of linear expansion which is essentially zero. Invar is, therefore, a rigid material relative to the differential force heat. If one can find for each differential force a material which is rigid relative to it, then one has a standard which can be used to determine the extent to which other materials are influenced by that particular differential force. Whether or not such materials exist is an empirical question which will decide for or against the Duhemian thesis in regard to geometry.

Grünbaum's second proof, which we believe to be invalid, depends on the method of successive approximation. Let us consider the following situation: the length of a rod depends on temperature. In order to correct for the influence of temperature on the rod we use the law of linear expansion. But to determine the law of linear expansion we must use a thermometer. Let us assume that the thermometer is an ordinary mercury thermometer. Since the length of the thermometer itself depends on temperature, the reading of the thermometer itself must be corrected. The very influence that we are measuring affects the measurement of the influence. Have we been caught in a vicious circle? No. The law which is discovered with the mercury thermometer can be corrected by correcting the temperature reading of the thermometer. As we alternatively correct the law of linear expansion and the readings of the mercury thermometer, we wlll approach a limit, much in the same way that instantaneous velocity and acceleration are determined.

This is the method of successive approximation, but, as Lenzen has noted,[12] this method depends on the fact that the law of linear expansion as discovered by the mercury thermometer is correct to the *first* approximation and incorrect only to the second approximation. The revision of the law which depends on using the corrected thermometer readings will then be correct to the *second* approximation and incorrect only to the *third*, etc. It is for this reason that a limit is reached. Temperature will affect the length of the mercury thermometer only to a small extent relative to its total length, thereby affecting the temperature reading only to a small extent. Thus, the method of successive approximation will give us a law of linear expansion which will differ from the initial law only in the higher orders of approximation.

When we apply these same considerations to the case of geometry, we find that we are concerned on the one hand with a geometry and on the other hand with a set of laws for correcting the measured lengths. If we assume as our congruence standard solid bodies as corrected for differential forces, we know that the geometry of our immediate space has to be in a relatively narrow range centering about Euclidean geometry. Therefore all of the possible geometries will be in a certain sense the same in the first approximation. The corrections which we make for differential forces will then be crucial inasmuch as they will affect our determination of the geometry of space in precisely that order of approximation which makes all the difference. Correspondingly, the laws which we use for correction will differ only in higher orders of approximation. Let us say then that we have two candidates for the geometry of space, G and G′, and two candidates for the law of linear expansion, E and E′, where one member of each pair differs from the other only in the second order of approximation. Furthermore let us assume that E′, but not E, is a function of the curvature of space in the second order of approximation, and that E has been arrived at by the method of successive approximation such that E together with G successfully predicts the observable data. Let us further assume that G′ together with E′ also successfully predicts the observational data but does so on the basis of a value of E′ *which is known to be incorrect* as it assumes a curvature of space other than that of G′.[13] By the method of successive approximation we can alternatively revise E′ and G′ until we reach in the limit an E″ and a G″ which successfully predicts the observational data and such that E″ assumes the curva-

ture of G″. Is there any guarantee that the G″ and E″ so arrived at will be the same as G and E, respectively? Initially G differed from G′ in the second order of approximation. Since E″ differs from E′ in the second order of approximation, G″ will differ from G′ in the third order of approximation, but will be the same in the second order and therefore *different from G*. It seems then that one would *not* expect G and G′ to converge in the limit, that is, one would expect G and G″ to be different. Therefore, the method of successive approximation which will achieve a unique limit when we are concerned with correcting for the influence of the measured property on the measuring instrument, will not achieve a unique limit where the measured property – in this case the curvature of space – is taken to have an effect on a correctional law which in turn has an effect on the measuring instrument, that is, where a law which is necessary for correcting for differential forces is a function not only of the differential force itself but also of the property which is ultimately being measured. Consequently Duhemian conventionalism can be defended even in the case of geometry which depends on the fundamental measurement of length if it can be empirically shown that there are no bodies rigid relative to each differential force.

In the remainder of the paper we will consider the broad implications of the Duhemian thesis on our conception of scientific knowledge. Specifically we will be concerned with whether the Duhemian thesis is an epistemological thesis regarding our knowledge of the world or a semantical thesis regarding the meaning of scientific words, and scientific language. Let us assume that a hypothesis H together with its interpreting laws A has been falsified, and we determine that the interpreting laws need revision and that the hypothesis is to be retained. Since we have in fact changed the meaning of the symbols of the hypothesis, have we effected nothing more than a trivial semantic change, that is, is not the H which is left after the change in A in fact a new hypothesis with new factual content merely couched in the old symbols? "The symbols assume meaning only by grace of the physical theories"[14]. Certainly in changing the interpreting laws, we have changed the meaning of the symbols, for the symbols are only implicitly defined in the context of theory, but this is equally true if we change the hypothesis, for the symbols depend for their meaning not only on the laws of the measuring instruments but also on the hypotheses in which they occur. A and H together implicitly define

the symbols. Duhemian conventionalism is certainly semantical for we are free to change either H or A *because* we are free to change the meaning of the symbols ot the theory. The freedom we have in physics, however, is not solely the trivial sort by which we can always change the meaning of a word, for it is also a freedom which is necessitated by the very context of scientific discovery. We could rationally reconstruct physics so as to eliminate this latter freedom by fixing the meaning of the quantities introduced by derivative measurement once and for all by stipulation (definition), but this would be contrary not only to the way physics actually develops, but to the way it should develop. A 'definition' which shifts with the wind, in the light of empirical findings, which is continually modified to protect other laws, is not a definition except in name, and to treat it as a definition is to be intellectually dishonest. If one is going to insist that every statement is either analytic or synthetic, the laws of the instruments should be treated as synthetic; however, we would argue that it is foolish to insist on this distinction in the cases we have discussed above. Let us be quite clear that we are not interested in demolishing the analytic-synthetic distinction entirely. We merely point out that there are cases where it is pointless, and in fact damaging, to insist on it, even in a rational reconstruction of science.

One final problem is yet to be resolved. We know that Duhem holds that the individual statements of physics are not *testable* in isolation from physical theory, but a crucial question is whether or not they are *true or false* in isolation. A group of physical laws together can have factual content, but do the individual laws taken in isolation have factual content? As we interpret Duhem he does believe that the individual statements of physics have factual content and are true or false independently of the rest of physical theory. This belief is based on an 'act of faith' that physical theory does capture the real order in reality, "but while the physicist is powerless to justify this conviction, he is nonetheless powerless to rid his reason of it"[15].

This faith must logically depend on a certain semantical commitment, namely, on the belief that symbols such as temperature have an extra-linguistic meaning which is logically independent of the intra-linguistic meaning which they have by virture of implicit definition, but which is epistemologically dependent on it. The presupposition then is that there is in reality such a quantity as temperature and that there is a correct

measure of it. When a law of the instrument is modified, it may be interpreted as an attempt to bring the *actual* measure of temperature in line with the *correct* measure. When an hypothesis is modified, it may be interpreted as a belief that the actual measure of temperature is also the correct one, and that the difficulty lies with the hypothesis.

If there is such a real quantity in nature as temperature, the question arises as to why we should be at all concerned with it. As far as we can tell this concern in turn depends on the further assumption that nature has in itself a certain simplicity which will be reflected in the laws we discover only if we succeed in using the correct measures. The underlying assumption is a metaphysical belief regarding the nature of the universe. If this assumption is not made, then it does not matter whether or not there is a *real* quantity temperature, and whether or not if there is, we succeed in correctly measuring it.

The question of the nature of conventionalism ultimately resolves itself into whether we are to take a realistic (Platonistic) or a nominalistic approach to the symbols of physics. If we take a realistic position, the individual statements of physics have independent factual content and are true or false; they express propositions. In this case if one were to hold, in contradistinction to Duhem, that one cannot *determine* which statements of science are true, but only whether or not a system of statements adequately predicts the facts, then one would be holding to a merely epistemological thesis regarding the limitations of *our knowledge* of the world.

On the other hand, if one follows a nominalistic approach in regard to these symbols, one is saying that these symbols have extra-linguistic meaning only to the extent that they occur in theories which are ultimately connected with experience. When statements involving these symbols are abstracted from their theoretical context, the meaning of the symbols and *a fortiori* of the statement is stripped from them. Outside of their theoretical context they are meaningless and therefore neither true nor false, while within the theoretical context they are meaningful law-like statements which are either true or false. We would say in this case that a statement 'expresses a proposition' only within a theoretical context.

These considerations illustrate that the distinction between 'descriptive simplicity' and 'inductive simplicity' is not a logical distinction but an ontological one, depending as it does on one's approach to the symbols of

physics and the nature of the universe. If one is a realist, one searches for inductive simplicity, for one believes that nature in itself is simple and, therefore, the correct description of nature will be the simplest. If one is a nominalist, one searches for descriptive simplicity, for one believes that the world can be described in simple terms, although not necessarily that it is any truer than a more complex description. The former believes that quantities are to be discovered for they will lead to the correct, that is, the simplest description. The latter believes that quantities are to be created with the view in mind of achieving the simplest description.

NOTES

[1] In W. V. O. Quine, *From a Logical Point of View*, Harper and Row, New York, 1963, pp. 20–46.

[2] Alfred A. Knopf, New York, 1963, Chap. 4.

[3] Quine, 'Two Dogmas of Empiricism', p. 42.

[4] *Ibid.*

[5] See his 'The Concept of Cognitive Significance: A Reconsideration', *Proc. Am. Acad. Arts and Sciences* LXXX (1951), 61–77, and also his *Aspects of Scientific Explanation*, Free Press, New York, 1965, pp. 101–119.

[6] See Pierre Duhem, *The Aim and Structure of Physical Theory*, trans. by Philip P. Wiener, Princeton University Press, Princeton, 1954, pp. 165–79, for a discussion of the particular characteristics of the laws of physics.

[7] *Ibid.*, p. 166.

[8] *Ibid.*, p. 150.

[9] *Ibid.*, p. 188.

[10] *Ibid.*, p. 26.

[11] *The Philosophy of Space and Time*, Dover, New York, 1958, p. 22.

[12] 'Procedures of Empirical Science', *International Encyclopedia of Unified Science*, University of Chicago Press, Chicago, 1955, Vol. I, pp. 289–95.

[13] The curvature of space assumed by E′ must, of course, be the same as that of G′ if the combination E′ & G′ is to make a claim to truth. We assume in this discussion *as does Grünbaum* that the curvature of space is constant. With this assumption there is no difficulty in manufacturing a second order coefficient of E′ which is a function of curvature and such that the resultant E′ together with *some* G′ yields the same observational consequences as G & E. We are not begging the question for we are not assuming that we can *manufacture* an E′ which can assume the *same* value for the curvature of space as G′ and yield the same observational consequences as E & G.

[14] Duhem, *Aim and Structure of Physical Theory*, p. 166.

[15] *Ibid.*, p. 27. Compare with the following quotation from Albert Einstein, 'Physics and Reality', *Journal of the Franklin Institute* CCXXL (1936), reprinted in Edward H. Madden, *The Structure of Scientific Thought*, Houghton Mifflin, Boston, 1960, p. 84: "It is an outcome of faith that nature – as she is perceptible to our five senses – takes the character of such a well formulated puzzle [i.e., has only *one* correct solution]".

GARY WEDEKING

DUHEM, QUINE AND GRÜNBAUM ON FALSIFICATION*

I

A number of empirically-minded philosophers hold that there is a class of 'basic' (or 'protocol') sentences each of which is, in the words of Karl Popper, "a statement of singular fact" ([3], p. 43). Many of these philosophers also maintain, with Popper, that there are empirical theories which are 'falsifiable' in the sense that each such theory

> divides the class of all possible basic statements unambiguously into the following two nonempty subclasses. First, the class of all those basic statements with which it is inconsistent (or which it rules out, or prohibits): ... secondly, the class of those basic statements which it does not contradict (or which it 'permits') ([3], p. 86).

Duhem has objected to such theories of falsifiability on the grounds that they inadequately represent the complex experimental procedures of theoretical physics. It is not the case that a single physical theory is by itself inconsistent with any experimental findings. For the construction of the physicist's experimental apparatus is based on physical hypotheses other than the one being tested. Thus when the physicist predicts an experimental result O from a hypothesis H, he assumes the truth of all the physical laws that go into the construction of his apparatus and into the calculation of O from H. Thus the ordinary model of the falsification of a hypothesis

$$((H \rightarrow O) \cdot \sim O) \rightarrow \sim H$$

of inductive logic is overly simple. Such a model would represent the testing of a physical hypothesis only if we knew that the other assumptions upon which the prediction was based were true. But inductive logic itself assures us that this is something we cannot know, for it is an accepted inductive maxim that a law (or universally quantified statement) can never be completely verified.[1] An adequate logical model for physical experi-

ments must, then, include in the antecedent of the conditional, of which the consequent is O, not only the hypothesis, but also these other assumptions A. If the result of an experiment is negative, we cannot, therefore, logically conclude that H is false. "The only thing (such an) experiment teaches us is that among the propositions used to predict the phenomenon ... there is at least one error; but where this error lies is just what it does not tell us" ([4], p. 185). Thus for the above simple model of falsification we must substitute

(i) $$(((H \cdot A) \rightarrow O) \cdot \sim O) \rightarrow (\sim H \vee (\exists p)(p \in A \cdot \sim p)).$$

Grünbaum claims to refute the Duhemian thesis by arguing that that situation is logically characterized, not by (i), but by

(ii) $$(((H \cdot A) \rightarrow O) \cdot \sim O \cdot A) \rightarrow \sim H.$$

As we noted above, however, A can never be inductively certain. Let p be the probability of A (symbolized $P(A)=p$). Let us also suppose that the antecedent of (i) is true, that $\sim O$ is an experimental result and that $H \cdot A \rightarrow O$. Since according to (i), A implies $\sim H$, $\sim H$ must be at least as probable as is A, that is,

$$P(\sim H) \geqslant p.$$

By the Negation Theorem of the probability calculus

$$P(\sim H) = 1 - P(H),$$

which implies that

$$P(H) \leqslant 1 - p.$$

It follows from this that

$$\lim_{p \rightarrow 1} P(H) = 0.$$

Thus we see that as the value of p approaches 1 (certainty) the value of $P(H)$ approaches 0 (impossibility). Since A can be, as Grünbaum himself points out, "only more or less highly confirmed" ([2], p. 135), p is always

less than 1. Thus the value of $P(H)$ is never 0,[2] which implies that (ii) does *not* characterize the logical situation.

The result of the above paragraph is, in fact, so obvious from an inspection of (ii), that one might wonder how it is that Grünbaum is led to commit such a blunder. This error can be traced to Grünbaum's failure to understand what Duhem's thesis asserts.[3] Apologizing for the lack of inductive certainty for A, he states that "the inductive risk ... inherent in affirming A does not arise from the alleged inseparability of H and A, and that risk can be made exceedingly small without any involvement of H" ([2], p. 137). It is noteworthy, however, that Duhem nowhere says that the inductive risk of A arises from the inseparability of A and H. Duhem's theory asserts, in fact, the exact converse of this; the inductive inseparability of H and A arises from the inductive risk of A. It is because A is inductively uncertain that (i), rather than (ii), is the correct logical model. Only if A were certain would H be experimentally separable from A. Since it is not certain, the negative result of an experiment (i.e. $\sim O$) will not determine which of the disjuncts of the consequent of (i) is true. Thus any test of H must be a test of both H and A; that is, H and A are experimentally inseparable.

The thesis that Grünbaum seems to have mistaken for Duhem's is that involved in the 'vicious circle' problem propounded by D. M. Y. Sommerville (cf. [2], p. 126). According to Sommerville, there are cases in which an experimental procedure is particularly complicated due to the involvement of H in the determination of A. The paradigmatic case of this situation is when Euclidian measurements are used in experiments designed to determine the geometry of physical space. Such situations would complicate our evaluation of the results of an experiment in an obvious way. The Duhemian theory does not, of course, exclude the possibility of such situations. But neither does its truth depend upon the existence of such situations.

Grünbaum seems also to wish to hold Duhem responsible for a considerably stronger thesis of Quine's. According to this view, "any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system" ([4], p. 43, cf. [2], p. 108). Applying this theory to the philosophy of physics, Grünbaum reads it as asserting that no matter what the result O' of an experiment designed to test an hypothesis H is, there is always some new set of theoretical assumptions

A' which, together with H, implies O'. Duhem, however, does not hold that an A' having the above property always exists; his is the far weaker thesis that the physicist can never be *sure* that *no* such A' exists.

> ... experimental contradiction does not have the power to transform a physical hypothesis into an indisputable truth; in order to confer this power on it, it would be necessary to enumerate completely the various hypotheses which may cover a determinate group of phenomena; but the physicist is never sure he has exhausted all the imaginable assumptions ([1], p. 190).

In the interest of historical accuracy, therefore, we will not follow Grünbaum ([2], pp. 110f., etc.) in referring to Quine's theory as 'Duhem's thesis' in the succeeding analysis.

II

According to Grünbaum, Quine's theory asserts that whenever the experimental result O entailed by assumptions A and hypothesis H fails to materialize as expected, there is always some alternative set of assumptions A' such that A' and H entail the actual experimental result O'. That is,

$$\text{If} \quad (H \cdot A \rightarrow O) \cdot \sim O, \quad \text{then} \quad (\exists A')(H \cdot A' \rightarrow O').$$

Here Grünbaum is substituting H for "any statement" in Quine's assertion that "Any statement can be held true come what may" ([4], p. 43). But that we "can" hold H in the face of O' does not imply, as Grünbaum seems to hold, that H is "needed to deduce O'" ([2], p. 111). At most this only implies that there is some language S' in which H and O' are consistent one with another.[4] There is, as a matter of fact, no reason to assume that Quine must affirm O' to be deducible from any of the true sentences in S' (other than itself). In other words, there is no necessity for Quine to assert of *any* given sentence in S' that it is not independent of the other truths of S'. It might, in such a situation, be objected that no system of physics should contain "observation statements" which cannot be deduced from the "theoretical statements" of that system. But this kind of objection would not be provincial to S', any more than to S in which A is still retained as true. For there is no guarantee that there is *any* consistent language (which is in some sense adequate to the whole of physics) in which any arbitrary statement about an observation can be (nontrivially) deduced. It is true, of course, that we would *prefer* our A' to be such that within the language S' of which H and A' are parts, we

could "explain" the result O' of our experiment. But that this is a requirement that we could always impose upon our choice of A' is far from obvious. It is certainly not implied in Quine's thesis.

There is still another way in which Grünbaum's formulation is an inadequate statement of the Quinean thesis. Grünbaum assumes that O', presumably because of its special status as an "observation statement", must count as a true sentence of S'. But when Quine speaks of the possibility of "pleading hallucination" in order to maintain an adequate system "in the face of recalcitrant experience" ([4], p. 43), he is denying that even sentences as relevant to experience as is O' are immune to revision. Quine argues that "it is nonsense ... to speak of a linguistic component and a factual component in the truth of any individual statement" ([4], p. 42). Such nonsense, he writes, has led philosophers to imagine an "extreme case where the linguistic component is all that matters" ([4], p. 41), an analytic statement. Though Quine does not address himself to this issue in 'Two Dogmas', there would seem to be another side to the same coin. The same confusion has led philosophers to believe that there are statements in which all that matters is the factual component. Such "sense-data" statements are, like their polar opposites, analytic statements, thought to be indubitable – immune to revision in any situation whatsoever. Quine's point is that there are *no* statements which are indubitable in this sense. Thus it appears that Grünbaum is mistaken in imposing even the requirement that S' must be *consistent* with O' upon Quine's theory.

But surely there must be *some* restriction or other upon the construction of S'. Grünbaum contends that in order for Quine's theory to be nontrivial, that it is necessary to impose on it the restriction that "*the theoretical language be semantically stable* in the relevant respects" ([2], p. 20). This means, I suppose, that S' must be such that it includes no terms, in particular no terms occurring in H, which have different meanings than they have in S (in which A counts as true). In formulating this requirement, Grünbaum seems to have forgotten that the Quinean thesis under consideration occurs in the final section of an essay in which the analytic-synthetic distinction is denied. In denying this distinction, Quine rejects the opinion "that the truth of a statement is somehow analyzable into a linguistic component and a factual component" ([4], p. 41). Since Grünbaum's requirement of semantical stability (RSS)

presupposes that we can always separate these components sufficiently to hold one constant while the other is altered, this requirement makes no sense as applied to the *Quinean* thesis. As Grünbaum offers no argument against the thesis developed in the earlier sections of 'Two Dogmas', we must reject RSS as a legitimate restriction. What, then, is a restriction that we *can* impose upon the construction of S' in order to test the validity of Quine's theory?

A requirement having certain intuitive appeal is that *all true statements be true in* S'. But this requirement is clearly inadequate, for it does not specify the language in which these sentences are true. That is, it does not tell us what language L to compare S' with in order to determine that all the sentences true in L are true in S'. It certainly cannot be construed as requiring that all true statements of any language whatever be true in S'. For it is not in general the case that all true sentences of any given language are even sentences (much less true sentences) of S'.

Perhaps what is needed is some requirement of translatability of truths. A conceivable requirement is that *for every language L there is a truth-preserving translation in S′ for each true sentence in L.* But we need *not* require translatability between S' and L for any L whatever. We do not, for example, have to worry about inconsistent languages, nor need we concern ourselves with any language which contains a true sentence such as 'There are centaurs', when there are in fact no centaurs. As a matter of fact, there is no obvious way to specify the class of languages with which we *need* concern ourselves. Certainly we would like each L in this class to be in some sense true to experience. In order to maintain any reasonable translatability requirement, then, we must restrict our choice of L to those languages which are 'adequate' to the facts of experience. But without an independent method of determining this class of languages, this reduces to the requirement that *S′ must be 'adequate' to the facts of experience.* As Quine puts it, "The edge of the system must be kept squared with experience" ([4], p. 45).

The Quinean thesis now becomes *for any (noncontradictory) hypothesis H there is a system S′ which 'squares' with experience, and in which H is true.* This thesis, it should be noted, is, even without Grünbaum's RSS, far from obvious. For it is by no means clear that there is *any* language (with or without H as a true sentence) which is adequate to experience in the sense required. And this is because it is not clear what sense of "squar-

ing with experience" is required in this connection. Quine contends that our choice *between* the various languages having the required property of squaring with experience is determined by *pragmatic* considerations. This would seem to require a notion of "squaring with experience" which can be rigorously delimited from these pragmatic considerations. But it is not evident that such a rigid delimitation is possible, especially in light of Quine's rejection of the notion of purely factual statements.

The above considerations lead me to doubt the truth of the Quinean thesis in any but its most trivial form, viz. that *a language S' is always possible in which any given sentence H counts as true.* Whenever we attempt to impose additional restrictions on our choice of S' in order to prove Quine's theory non-trivial, its meaning begins to elude us. But this is not to say that the theory is worthless, for it results in an observation which is significant indeed: the difference between those situations in which Quine's thesis holds only trivially and those in which we are confronted with a real and significant alternative as to which sentences we should reject as false and which we should retain as true is only a matter of degree.

APPENDIX

Since the above paper was written in the spring of 1965, it does not take into consideration the more recent literature on the controversy. I would like to point out, in particular, Laurens Laudan, 'Grünbaum on "The Duhemian Argument"', *Philosophy of Science* **32** (1965), 295–99, and Carlo Giannoni, 'Quine, Grünbaum, and the Duhemian Thesis', *Nous* **1** (1967), 283–98, both of which give Duhem an interpretation similar to my own. In Grünbaum's more recent writings ('The falsifiability of a Component of a Theoretical System', in Feyerabend and Maxwell (eds.), *Mind, Matter, and Method*, University of Minnesota Press, Minneapolis, 1966, pp. 273–305, and 'Can We Ascertain the Falsity of a Scientific Hypothesis', Chapter 17 of *Philosophical Problems of Space and Time*, Second, enlarged edition, D. Reidel, Dordrecht-Holland, 1973 [excerpted below]), he appears to concede that his was a misinterpretation of Duhem, although he continues to refer to a thesis as 'Duhem's' which is apparently of his own invention, at any rate, certainly not of Duhem's. The issues raised in the second part of my paper could be raised basically unchanged against Grünbaum's newer writings.

NOTES

* From *Philosophy of Science* (1969). Reprinted by permission.

[1] This is not to say that *no* universally quantified statement can be known to be true. For those that are logically true can be known to be true. But in these cases the concept of verifications is hardly applicable.

[2] We are assuming, of course, that H cannot be known to be false on noninductive grounds, i.e. that H is not self-contradictory.

[3] The extent of Grünbaum's misunderstanding of Duhem can be seen from the fact that he finds no inconsistency in attacking Duhem's denial of falsifiability while holding as "eminently sound" the very basis of this denial, viz. that "the logic of every disconfirmation, no less than of every confirmation of an isolated scientific hypothesis H is such as to *involve at some stage or other* an entire network of interwoven hypotheses in which H is ingredient rather than in every stage merely the separate hypothesis H". ([2], p. 111.)

[4] It may be objected that in any system in which both H and O' are true, 'H implies O'' is also true. But this is true only of the material implication. And Grünbaum clearly intends the relation between H and O' to consist in the *necessity* of using H in the deduction of A.

BIBLIOGRAPHY

[1] Duhem, P.: *The Aim and Structure of Physical Theory* (trans. by Philip P. Wiener), Atheneum, New York, 1962.

[2] Grünbaum, A.: *Philosophical Problems of Space and Time*, Alfred Knopf, New York, 1963.

[3] Popper, Karl R.: *The Logic of Scientific Discovery*, Science Editions, New York, 1961.

[4] Quine, W. V. O.: 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Harper & Row, New York, 1963, pp. 20–46.

MARY HESSE

DUHEM, QUINE AND A NEW EMPIRICISM*

1. THE DUHEM–QUINE THESIS

As in the case of great books in all branches of philosophy, Pierre Duhem's *La Théorie Physique*, first published in 1906, can be looked to as the progenitor of many different and even conflicting currents in subsequent philosophy of science. On a superficial reading, it seems to be an expression of what later came to be called deductivist and instrumentalist analyses of scientific theory. Duhem's very definition of physical theory, put forward early in the book, is the quintessence of instrumentalism:

> A physical theory is not an explanation. It is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as simply, as completely, and as exactly as possible a set of experimental laws [p. 19].

The instrumentalist overtones of this become clear from the implications of the denial that theories are explanations. For Duhem an explanation is a metaphysical entity, and science should be independent of metaphysics. But this dictum is not intended, as with the positivists, to dispose of metaphysics as irrational or meaningless; it is rather an assertion of the autonomy and dignity of metaphysics as alone capable of expressing the truth of how things are in the world. Metaphysics according to Duhem is not independent of experience, but its methods are not those of science, and its conclusions stand independently of changing fashions in science. Thus it is for Duhem a grave error to interpret scientific theory as itself providing a metaphysics – a global theory drawn from science such as mechanism is not only false, because science outgrows it by its own methods, but also it is not the kind of theory that could ever be true, because it illegitimately uses the methods of mathematical representation of experimental facts to construct an ontology and to give answers to substantial questions about the nature of the world and of man. But only metaphysics, and in particular a religious metaphysics, can

do that. The aim of science must be more modest. A non-interference pact must be established between the domains of science and metaphysics.

Duhem was not the first nor the last philosopher of religion to see the answer to teasing conflicts between science and religion in terms of a complete separation of their spheres of influence, but this is not the aspect of Duhem's thought that I want to discuss here. Indeed, if this were all there were to say about Duhem's philosophy of science it would deserve no more than a minor place in the history of positivism. But his extra-scientific preoccupations did not after all mislead him into so crude an analysis of science itself as his definition of scientific theory would entail. He is saved by a discussion of the observational basis of science that is far subtler than that presupposed by later deductivists and instrumentalists, and paradoxically it is a discussion which can be made to undermine the very foundations of the dichotomy of mathematical theory and explanation, science and metaphysics, that his theory of explanation presupposes.

Most empiricist accounts of science have been based, usually tacitly, on the notion of a comparatively unproblematic observation language. It matters little how this is construed – whether in terms of hard sense data, operational definitions, ordinary language, or what not – the essential point is that there are statements of some kind whose meaning as descriptions of states of affairs is supposed to be transparent, and whose truth-value is supposed to be directly and individually decidable by setting up the appropriate observation situations. It is a long time since anyone seriously claimed that the truth of such statements can be known *incorrigibly*, but most eyes have been averted from the consequences of the significant admission of fallibility of even observation statements, and attention has been concentrated on the way in which meaning and truth-value is conveyed to theories, regarded as in these respects parasitic upon observation statements and clearly distinguishable from them. The consequences for deductivism have been proliferation of a number of insoluble and unnecessary problems regarding the meaning of theoretical statements and the possibility of confirming them, and the result has been a slide into instrumentalism in which, in the end, only observation statements and not theories have empirical interpretation. What that interpretation and its significance is still remains unanalysed.

Duhem introduces two important modifications into this type of classical empiricism. They may be expressed as a new theory of *correspondence* and a new theory of *coherence*.

(i) In his theory of *correspondence*, attention is shifted away from the empirical basis of traditional empiricism to the theoretical *interpretation* of that basis. Duhem sees that what is primarily significant for science is not the precise nature of what we directly observe, which in the end is a *causal* process, itself susceptible of scientific analysis. What is significant is the interpretive expression we give to what is observed, what he calls the *theoretical facts*, as opposed to the 'raw data' represented by *practical facts*. This distinction may best be explained by means of his own example. Consider the theoretical fact 'The temperature is distributed in a certain manner over a certain body' (p. 133). This, says Duhem, is susceptible of precise mathematical formulation with regard to the geometry of the body and the numerical specification of the temperature distribution. Contrast the practical fact. Here geometrical description is at best an idealisation of a more or less rigid body with a more or less indefinite surface. The temperature at a given point cannot be exactly fixed, but is only given as an average value over vaguely defined small volumes. The theoretical fact is an imperfect translation, or interpretation, of the practical fact. Moreover, the relation between them is not one-one, but rather many-many, for an infinity of idealisations may be made to more or less fit the practical fact, and an infinity of practical facts may be expressed by means of one theoretical fact.

Duhem is not careful in his exposition to distinguish *facts* from *linguistic expressions of facts*. Sometimes both practical and theoretical facts seem to be intended as linguistic statements (for instance, where the metaphor of 'translation' is said to be appropriate). But even if this is his intention, it is clear that he does not wish to follow traditional empiricism into a search for forms of expression of practical facts which will constitute the basis of science. Practical facts are not the appropriate place to look for such a basis – they are imprecise, ambiguous, corrigible, and on their own ultimately meaningless. Moreover, there is a sense in which they are literally inexpressible. The absence of distinction between fact and linguistic expression here is not accidental. As soon as we begin to try to capture a practical fact in language, we are committed to some theoretical interpretation. Even to say of the solid

body that 'its points are more or less worn down and blunt' is to commit ourselves to the categories of an ideal geometry.

What, then, is the 'basis' of scientific knowledge for Duhem? If we are to use this conception at all, we must say that the basis of science is the set of theoretical facts in terms of which experience is interpreted. But we have just seen that theoretical facts have only a more or less loose and ambiguous relation with experience. How can we be sure that they provide a firm empirical foundation? The answer must be that we cannot be sure. There is no such foundation. It must be admitted that Duhem himself is not consistent on this point, for he sometimes speaks of the persistence of the network of theoretical facts as if this, once established, takes on the privileged character ascribed to observation statements in classical positivism. But this is not the view that emerges from his more careful discussion of examples. For he is quite clear, as in the case of the correction of the 'observational' laws of Kepler by Newton's theory (p. 193), that more comprehensive mathematical representations may show particular theoretical facts to be false.

However, we certainly seem to have a problem here, because if it is admitted that subsets of the theoretical facts may be removed from the corpus of science, and if we yet want to retain empiricism, the decision to remove them can be made only by reference to *other* theoretical facts, whose status is in principle equally insecure. The correspondence with experience, though loose and corrigible, must still be retained, and still remains unanalysed.

(ii) Duhem's theory of *coherence* is indispensable to a satisfactory resolution of this problem. The theory has been much discussed, but unfortunately not always in the context in which Duhem set it, with the result that it has often been misunderstood and even trivialised.

Theoretical facts do not stand on their own, but are bound together, in a network of laws which constitutes the total mathematical representation of experience. The putative theoretical fact that was Kepler's third law of planetary motion, for example, does not fit the network of laws established by Newton's theory. It is therefore modified, and this modification is possible without violating experience because of the many-one relation between the theoretical fact and that practical fact understood as the ultimately inexpressible situation which obtains in regard to the orbits of planets. It follows that neither the truth nor the falsity of a

theoretical fact or a lawlike relation connecting such facts can be determined in isolation from the rest of the network. Systems of hypotheses have to come to the test of experience as wholes. Individual hypotheses are not individually falsifiable any more than they are individually verifiable.

Quine, as is well known, has taken up both aspects of Duhem's new empiricism. A bare remnant of empirical correspondence is implied by his dictum that 'our statements about the external world face the tribunal of sense experience not individually but only as a corporate body' – for Quine they do face it; how they face it has come in his recent writings to be a question for a stimulus-response psychology (Quine, 1960, 1968). The coherence of our knowledge is also implied, in the very strong sense (which is never explicitly claimed by Duhem) not only that generally speaking hypotheses cannot individually be shown to be false by experience, but that *no* statement can be; any statement can be maintained true in the face of any evidence: 'Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system' (Quine, 1953, p. 43). Because it is doubtful whether we ever want to 'hold a hypothesis *true*' rather than highly confirmed or highly probable, and because I do not here want to beg or examine that question, I shall discuss Quine's claim in a slightly more weakened form than is implied by this quotation. The weaker form, which I shall call the Q-thesis, is that

No descriptive statement can be individually falsified by evidence, whatever the evidence may be, since adjustments in the rest of the system can always be devised to prevent its falsification.

It has seemed to many commentators that to replace the observational basis of science with this shifting network is to open the floodgates to conventionalism, and to a vicious circularity of truth-value and meaning which is in effect an abandonment of empiricism. Popper (1959, p. 78), for example, classes Duhem with Poincaré as a conventionalist. But if by conventionalism is meant, as Poincaré apparently intended in regard to the geometry of physical space, that any given total theoretical system can be imposed upon any logically possible experience, then surely to class Duhem as a conventionalist is a mistake. For neither Duhem nor Quine say anything to imply that a total system is not refutable by experience; indeed that it is so refutable is entailed by their contrast between refutability of individual hypotheses and refutability of the linked system

of hypotheses. Once parts of the system have been fixed, perhaps conventionally, there are some extensions of it that are empirically excluded.

But elsewhere Popper (1963, p. 238ff.) demands something more than this:

> We can be reasonably successful in attributing our refutations to definite portions of the theoretical maze. (For we *are* reasonably successful in this – a fact which must remain inexplicable for one who adopts Duhem's or Quine's views on the matter) [p. 243].

The 'holistic argument goes much too far' if it denies that it is ever possible to find out which is the guilty hypothesis. There are, he suggests, three ways in which it may in fact be identified:

(i) We may provisionally take for granted the background knowledge common to two theories for which we design a crucial experiment, and regard the experiment as refuting one or other of the theories rather than the background knowledge. But neither Duhem nor Quine would ever deny this possibility, and it is of course not sufficient to refute Q, since it does not require acceptance of the background knowledge to be anything but *provisional.*

(ii) We may be able to axiomatise the whole theoretical system in such a way as to isolate the effect of a single axiom, which may then be refuted in isolation. But even if we disregard the extreme impracticability of such axiomatisation in the case of most interesting scientific theories, its ideal possibility still does not refute Q, because no axiomatisation can fully account for the empirical applicability of the system, and the correctness of the conditions of application (the so-called 'correspondence rules') might always be called into question to avoid abandonment of any of the axioms.

(iii) Theories need to make successful predictions (to be 'corroborated' in Popper's terminology) as well as being refuted if false. When successful predictions have occurred, Popper seems to suggest, we are more reluctant to abandon those parts of the theory responsible for them, and more willing to locate the responsibility for subsequent refutations in other, less well corroborated parts of the network. Popper's notion of corroboration here as elsewhere is far from clear, but it is difficult to interpret this suggestion in any sense other than in terms of relative inductive *confirmation* of some parts of the system in comparison with others. Some theory of confirmation of the system by experience does indeed seem to be a requirement of the network analysis, and I shall re-

turn to this requirement below; but as far as Popper's suggestion goes, he only regards this method of picking out a guilty hypothesis as indicative and not conclusive, and so the method in any case would not refute Q.

The Q-thesis has also recently come under attack from Adolf Grünbaum (1963, Ch. 4; 1966). In a series of articles Grünbaum has sought to show that Q is true only in trivial cases in which 'drastic adjustments elsewhere in the system' are construed as allowing *ad hoc* changes in the rules of English usage. Clearly if a hypothesis predicts that roses are red, and they turn out to be black, Q is not satisfied except trivially by interchanging the uses of 'red' and 'black' in observation reports in English. 'Hence', Grünbaum continues, 'a *necessary* condition for the non-triviality of Duhem's thesis is that *the theoretical language be semantically stable* in the relevant respects' (1966, p. 278). He does not, however, claim to give general sufficient conditions for the non-triviality of auxiliary hypotheses or rules which would preserve the truth of a hypothesis H in the face of apparently contrary evidence, nor does he attempt to spell out in detail what it would be for the theoretical language to be 'semantically stable' or for H to remain the 'same hypothesis', arguing only that Quine's suggestion of resort to a non-standard logic must at least be regarded as trivial, as must *ad hoc* changes in the meanings of descriptive terms.[1]

This criticism suggests, therefore, a second requirement for the Q-thesis to be viable, namely some theory of change and retention of *meaning* within the network, in addition to the first requirement of some theory of confirmation.

To summarise these important and pervasive kinds of doubt about the viability of the Q-thesis, it is convenient to quote further from Grünbaum. Discussing Einstein's assertion that any metric geometry can be preserved in the face of any empirical evidence, he says:

> Indeed, if the Duhemian is to maintain, as he does, that a *total theoretical system* is falsifiable by observations, then surely he must assume that the relevant falsifying observations present us with sufficient *relatively* stubborn fact to be falsifying.... And if there were no relatively stubborn fact... how could the Duhemian avoid the following conclusion: "Observational findings are always so unrestrictedly ambiguous as not to permit even the refutation of any given total theoretical system"? But such a result would be tantamount to the absurdity that any total theoretical system can be espoused *a priori* (1966, p. 288).

One could earn a quick point against this passage by remarking a *non sequitur* between 'if there were no *relatively* stubborn fact' and 'ob-

servational findings are always so *unrestrictedly* ambiguous ...'. Might there not, one wants to ask, be relatively *un*stubborn facts which were nevertheless not *so* unrestrictedly ambiguous as to warrant the conclusion that any theory might be espoused as true *a priori*? Neither does it obviously follow that if there is no conclusive refutation, any theory goes, for there may be available a theory of relative confirmation. More fundamentally, it should be noted that Grünbaum has almost unwittingly fallen into just the habit of distinguishing the theoretical system from the 'relatively stubborn facts' that are called in question by the Q-thesis. That the facts are only *relatively* stubborn does not save him, for the whole thrust of Q is against the practice of looking in the system for those statements which can, even relatively, form its basis, and upon which the rest of the system is propped up. 'Relatively' in this context is always taken in classical empiricist accounts to imply 'relative to some *more* basic statements which we could uncover if we had time or necessity'. But the only relativity of stubbornness that can be allowed in a Q-system is relativity with respect to the *other theoretical* statements. The structure is mutually supporting. Where the points of external support are applied is a subsidiary matter which cannot be decided independently of the character of the network itself. Grünbaum might well reply that this leaves the theory completely up in the air, and removes it from empirical control in just the way he fears. So we are left with the two requirements of a network theory as constituting problems for explication:

(i) That some relative empirical confirmation should be provided, and that without being able to identify any statement of the system which expresses the evidence incorrigibly.

(ii) That some means of analysing stability and change of meaning in the network should be provided.

2. Criteria of confirmation

When faced with a philosophical tangle which seems to involve logical circularities or contradictions, it is often illuminating to try to conceive of a mechanism which simulates the conditions of the problem, and to see whether a self-consistent model of it is possible. What we need in this case is a machine capable of representing and theorising about its environment according to the conditions just described. We can distinguish in the usual

way between the hardware and the software of the machine: the machine has a certain physical constitution, the hardware, which we will assume remains fixed (the machine is not at present regarded as a structurally evolving organism), and its software includes a certain system of coding according to which some of its physical states can be represented in its own 'language', and the representation perhaps printed on an output tape, so that the machine is capable of 'reporting' on its environment. Suppose the machine goes through the following stages of operation:

(i) Physical input from the environment causally modifies part of the machine (its 'receptor').

(ii) The information thus conveyed to the receptor is represented in the machine language according to a code present in the machine. We may assume at this stage that the code is not, at least in practice, infinitely and exactly competent, so that (a) if the input is potentially infinitely various, some information present in the receptor is lost at the coding stage, and (b) the mechanism may make mistakes in a small proportion of translation of the input into code. The product of this stage will be called the *coded input* (C.I.), and corresponds to the set of observation statements produced by a human investigator as the representation in language resulting from experienced sensory input. Notice in particular that C.I. is not necessarily a complete or accurate representation of the input.

(iii) C.I. is examined for repetitions which yield inductive generalisations, and for more complex patterns which yield theories. If the machine is inductivist it may run through all possible systems of generalisations seeking that which is in some specified sense most probable or most simple. If it is deductivist it may have a small stock of patterns to try out on C.I., rejecting those whose fit is too bad, and retaining those whose fit is 'best' or good enough. In either case it is not necessary that the theory arrived at by the machine should be consistent with *every* piece of C.I., only that *most* of it should be consistent. Moreover, *which* parts of C.I. are going to be consistent with the best theory cannot be determined in advance, but only by examining the theory in the light of the complete C.I., and adjusting it to make the best fit. In other words, no single statement of C.I. is incorrigible relative to a good theory, only most of it must be. There are no epistemologically privileged protocol statements, but the element of correspondence which implies an empiricist check on the

theory is still present in the whole set of observation statements. Thus the first requirement of the network model of science regarding the possibility of confirmation can be met by providing (a) the empirical check or correspondence element present in the whole of C.I., and (b) some principles of probability or simplicity of theories which are used as the coherence element to choose the best theory and modify and perhaps discard some small part of C.I.

Does the Q-thesis hold for such a machine? No C.I. statement (say C_1) is logically immune from correction by the best theory or theories, and therefore if some given descriptive statement S is contradicted by C_1, S can in principle always be regarded as unfalsified by taking C_1 to be itself false. But what does 'false' mean here? It cannot mean that C_1 is false as a direct representation of the input, because all that can be known about this is that a certain small proportion of such C.I. statements are false, not which ones. It must mean false relative to a 'best' theory constructed in the light of the whole C.I. and the internal coherence criteria for construction of theories. If these criteria are sufficiently modified no doubt C_1 could be made consistent with some theory which satisfied them, but it does not follow that this could be done according to any criteria which would be accepted as reasonable for a good theory. However, this limitation on the applicability of Q only highlights the importance for judging 'truth' and 'falsity' of what are taken to be reasonable criteria, and indicates that these are not immediately determined by the input, but are in a sense *a priori* relative to that input. They may, of course, possibly be regarded as modifiable in a second order sense in the light of the type of theories which seem to be successful for large amounts of input, but then the principles of modification would presuppose yet higher-level criteria, and so on. I shall not pursue here the problem of specifying and justifying criteria for 'best' theories, for in the light of current discussion of various kinds of confirmation theory the only thing that is clear is that the problem is turning out to be unexpectedly deep and difficult, and as yet hardly rewarding. But it does seem important to emphasise that some statement of confirmation criteria for theories seems to be a necessary condition of rebutting the charge that the Q-thesis effectively abandons empiricism.

It may seem that this requirement contradicts the claim that is also integral to the Quinean approach, namely that there is no ultimate

distinction between the *a priori* and the *a posteriori*, the analytic and the synthetic. Quine himself (1968) has recently willingly accepted that the possession of some 'innate ideas' is a corollary of his network model of language. This is in reply to objections by Chomsky (1968), who curiously reads Quine, not as abandoning empiricism, but as sticking too closely to it in his analysis of sensory conditioning as the foundation of language learning. There is no empirical evidence, Chomsky claims, for the kind of language learning Quine seems to require, rather all the evidence we have (which incidentally is from syntax rather than semantics, and therefore not clearly relevant to the conditions of applicability of descriptive predicates – but let us take Chomsky for the moment at his own valuation) points to the presence of innate, interlingual dispositions to certain standard linguistic principles. Quine's acceptance of this point seems motivated rather by desire to conform to the present state of empirical linguistics and psychology of perception than by general arguments such as have been put forward here. Innate principles which are understood merely as conditions causally operating on sensory data perhaps need not count for Quine as *a priori* principles which refute his conflation of the prior and the posterior. However, we can hardly be content with this understanding of the principles. To remain so would be like accepting a physico-physiological account of the processes which go on when we do sums, and regarding this as excluding rational discussion of the logically systematic principles involved in doing sums correctly. To adapt the favourite metaphor which Quine takes from Neurath, modifying parts of the network while relying on other parts may be like rebuilding a boat plank by plank while it is afloat, but there are right and wrong ways of doing the rebuilding. To provide a normative inductive logic in which the innate principles are systematically explicated does not *preclude* empirical investigations of the scientific and social facts about inductive reasoning, but it tells us more about them, by showing why and under what conditions they can be regarded as rational. There is a close parallel here with the programme of rational decision theory, which may be assisted, but is not determined, by empirical investigations of practical decision making. In the sense of a rational inductive logic, then, the innate principles would be *a priori* relative to the data which they process, but this sense need not be objectionable to Quine, since no claim is made about the eternal immutability of the principles – different

external conditions may cause adaptive organisms to modify these principles too.

3. Meaning

The alternative possibility of saving S from falsification, which has been dismissed by Grünbaum as trivial, is to so 'change the meaning of S' that it no longer contradicts the evidence. How can we understand in terms of the machine model the demand that meanings shall be stable in order to exclude trivial satisfaction of Q? We cannot directly and immediately apply the usual empiricist interpretation of 'the meaning of S' as the empirical conditions necessary and sufficient for S to have the truth-value true, because the only criteria of truth we have are relative to the coherence of the system as well as to its empirical constraints. Indeed, the truth- or probability-value of S *relative to the current best theory* may change as additional evidence replaces best theory by another, and that without direct observation of the empirical conditions of satisfaction of S. So in this sense the meaning of S, like its truth-value, is not invariant to accumulating evidence. Is such instability of meaning an objection to the network model?

It can certainly be interpreted in such a way as to constitute a *reductio ad absurdum* of any model of science which attempts to retain an element of empiricism, including the network model. It has been so interpreted by several recent writers in the guise of what has come to be called the 'meaning variance thesis' (Hanson, 1958; Feyerabend, 1962; Kuhn, 1962), and since I want to distinguish the network model from this thesis in important respects, I shall start by stating and examining the thesis itself.

The original context of the meaning variance thesis was an attack upon the deductive model of theories with its accompanying assumption that there is a comparatively stable and transparent observation language, upon which theoretical language is parasitic. It is pointed out, first, that reliance on deducibility in the deductive account of explanation of observation by theory, and reduction of one theory to another, is vain, because there is always a measure of approximation in such inferences, and hence it is always possible for the same data to be 'explained' in mutually contradictory forms by mutually contradictory theories. For example, Galileo's law is not a logical consequence of Newton's theory; in fact it is contradicted by that theory, because the law asserts that the

acceleration of bodies falling along the earth's radii is constant. It was possible to hold Galileo's law to be true only because this discrepancy was concealed by experimental error. And yet Newton's theory is held to *explain* the facts about falling bodies in spite of contradicting the experimental law which had been accepted up to then as a description of those facts. Again, Newtonian mechanics cannot simply be reduced by deduction to the more comprehensive relativity mechanics, because relativity mechanics entails, among other things, that space and time are mutually dependent and inseparable dimensions, and that the mass of a body is not an invariant property, but a function of the body's speed relative to whatever happens to be taken as the rest frame. Such consequences of relativity are strictly *inconsistent* with Newtonian mechanics. Similar objections may be made to the alleged deductive reduction of phenomenological thermodynamics to statistical mechanics, and of quantum to classical electrodynamics. Many of these examples involve something even more radical than mere numerical approximations. It is meaningless, for example, to speak of Newtonian mechanics 'approximating' to relativistic mechanics 'when the constant velocity of light c is taken as infinite', or of quantum theory 'approximating' to classical physics 'when the quantum of action h is taken as zero', because it is of the essence of relativity and quantum theory that the respective constant c and h are *finite constants*, having experimentally specifiable values. Moreover, in passing from one theory to another there are *conceptual* as well as numerical changes in the predicates involved: mass as invariant property becomes variable relation, temperature as property becomes a relational function of velocity, atom as indestructible homogeneous stuff becomes divisible and internally structured.

Such examples as these lead to the second, and more radical, part of the meaning variance thesis, namely that deducibility is impossible not only because numerical fit between theory and observation is at best approximate, but also because the concepts of different theories are governed by rules of syntax and use implicit in the respective theories, and since different theories in a given experimental domain in general conflict, these rules of usage are in general inconsistent. Hence explanation of observation by theory, or reduction of one theory to another, cannot take place by identification of the concepts of one theory with those of observation or of another theory, nor by empirically established

relations between them. We cannot even know that different theories are 'about' the same observational subject matter, for if the meaning of the predicates of observation statements are determined by the theoretical beliefs held by their reporters, and if these meanings differ in different theories, then we seem to have an incommensurability between theories which allows no logical comparison between them, and in particular allows no relations of consistency, incompatibility or relative confirmation.

The thrust of the meaning variance thesis is therefore primarily against the notion of a neutral observation language which has meaning invariant to changes of theory. But the thesis becomes impaled on a dilemma. Either there is such an independent observation language, in which case according to the thesis its predicates cannot be related deductively or in any other logical fashion with any theoretical language, or there is no such observation language, in which case every theory provides its own 'theory-laden' observation predicates, and no theory can be logically compared with any other. The consequences of meaning variance can be put in paradoxical form as follows:

(1) The meaning of a term in one theory is not the same as its meaning in another prima facie conflicting theory.

(2) Therefore no statement, and in particular no observation statement, containing the predicate in one theory can contradict a statement containing the predicate in the other.

(3) Therefore no observation statement which belongs to one theory can be used as a test for another theory. There are no crucial experiments between theories.

A similar paradox can be derived from (1) with regard to both explanation and confirmation.

(1a) The meaning of a predicate in the pre-theoretical observation language is different from its meaning in a theory which is said to explain that domain of observation and to be confirmed by it.

(2a) Therefore if the theory entails some observation statement, that statement cannot be the same as any pre-theoretical observation statement, even if it is typographically similar to it.

(3a) Therefore no theory can explain or be confirmed by the statements of the pre-theoretical observation language.

That such paradoxes seem to follow from the meaning variance thesis

has been taken to be a strong objection to the thesis, and hence strong support for the view, presupposed in the deductive account, that observation statements have meaning independent of theories. On the other hand there is certainly a prima facie case for item (1) of the meaning variance thesis, and the network model itself is committed to a similar abandonment of the theory-neutral observation language. Must the notion of a theory-laden observation language lead to paradox?

First, it may be wondered whether so radical a departure from deductivism as indicated by (2) is really warranted by the argument for (1). Suppose we grant for the moment (1) in some sense of 'the meaning of a predicate' which could be incorporated into deductivism, for example that the predicates of a theory are 'implicitly defined' by the postulates of that theory, which entails (1). Even so, for the paradoxes to go through, a further step is required. It must be shown either (i) that the sense of 'meaning' required to make (2) true is the same as that required for the truth of (1), or (ii) that another concept of 'meaning' is implicit in (2), that for this concept meaning is also theory-variant, hence that (2) is still true, and the paradoxes follow. (i) can be disposed of very quickly. In order to establish (i) it would be necessary to show that the difference of meaning of 'P' in different theories which is asserted in (1) is such as to preclude substitutivity of 'P' in one theory T for 'P' in the other theory T′, so that no relations of consistency, entailment or contradiction could be set up between statements of T and T′. If this were true, however, it would also be impossible to speak of the difference of meaning *of* 'P' in T and T′, for this formulation already presupposes some meaning-identity of 'P' which is not theory-variant. Hence (1) would be not just false, but inexpressible. What, then, is the relevant identity of 'P' presupposed by the possibility of asserting (1) which will also make (2) false and hence dissolve the paradoxes? Here typographic similarity will clearly not do. We must appeal somehow to the external empirical reference of T and T′ to give the meaning-identity of 'P' that will allow substitutivity of 'P' between the theories.

The suggestion that naturally springs to mind within the deductive framework is to take the class of objects that satisfy P, that is, the extension of 'P', and identify the relevant meaning of 'P' with this extension. In pursuit of this suggestion Israel Scheffler (1967, Ch. 3) proposes to construe 'meaning' in the classic Fregean manner as having

two separable components: 'sense' and 'reference'. (1) may be regarded as the assertion that the sense, or definition, or synonymy relations of predicates differ in different theories, but in considering the logical relations of deducibility, consistency, contradiction, and so on, it is sameness of reference or extension that is solely involved. Difference of sense does not imply difference of reference, hence (2) and (3) do not follow from (1). Thus Scheffler claims to reconcile variance of meaning between theories, and between theory and observation, with invariance of reference and hence of logical relations.

Unfortunately this reconciliation does not work even within the deductive framework. Waiving difficulties about construing sense in terms of definitional synonymy relations, the most serious objection is that 'same reference' is neither necessary nor sufficient for the logical comparability that is required of different theories. It is not *sufficient* because the properties ascribed to objects in science are not extensional properties. Suppose two theories T_1 and T_2 are 'about' two quite distinct aspects of a domain of objects: say their colour relations, and their shapes. It may happen that T_1 and T_2 are such that there is an exact one-to-one correspondence between the sets of predicates of T_1 and T_2 respectively, and that as far as is known T_1 is true of any set of objects if and only if T_2 is also true of it. Then the corresponding predicates of the two theories have the same referential meaning. But this does not imply that the theories are the same. So long as no predicates are added to their respective predicate-sets, no development of T_1 can be either consistent with or contradictory to any development of T_2. In other words, because science is about *intensional* properties, sameness of extension does not suffice for logical comparability. Furthermore, sameness of reference is not *necessary* for logical comparability. Two different theories may make use of different categorisations or classifications of objects: thus Dalton's atoms have different extensions from Cannizzaro's atoms, yet we want to be able to say of some of Cannizzaro's statements that they entail or contradict some of Dalton's.

The network model gives promise of resolving the paradoxes by, first, giving a more subtle analysis of the observation language than that presupposed by deductivism, in terms of which what I have called 'intension' of predicates as well as their extension has a place, and second by allowing a distinction to be made between meanings which are internal

to a theory, and meanings which are empirically related. Return for a moment to the observing machine described earlier. We have already noticed that the meaning of descriptive statements is internally related to the best theory and its criteria in something like the way the meaning variance theorists describe. It is also the case that no simple account of the meaning of descriptive predicates in terms of their extension is possible in this model, because all we can know about extension is also relative to the state of the evidence. It may be true or highly probable that P applies to a given object according to one best theory, but false or highly improbable according to another theory adopted on different evidence. There is, however, a relation between machine hardware and input that does remain constant during the process of data collection and theory building that has been described. This is the set of physical conditions under which input becomes coded input. These conditions do not demand infinite exactness nor complete freedom from error, but in what has been said so far they have been assumed sufficiently stable to permit the assertion that a high proportion of statements in the C.I. are true, though we don't know which. This stability is sufficient to ensure that trivial changes of meaning are not resorted to to save theories come what may. Translated into terms of human language-users, this stability does not require that they be aware of some transparent empirical relation between observed properties and linguistic predicates, nor even that they always entertain the same theories; it requires only that by learning to apply predicates in an intersubjectively acceptable manner, they have acquired physical dispositions which are invariant to change of evidence.

To express the matter thus is to invite the comment: does not this kind of stability entail undue inflexibility in the use of descriptive predicates? Do not the meanings of our predicates sometimes change even in this respect under pressure of evidence? In other words, does not evidence also educate our dispositions? It seems fairly clear from the history of science that it does. Consider the predicates 'heavy' and 'light' after Newton's theory had been accepted. It then became incorrect to use the word 'light' of air, and correct to use the word 'heavy', because in Newton's theory all material substances are heavy by definition, even if they can be made to cause a balloon to rise. In such cases there is indeed no substitutibility with retention of truth-value of 'heavy' before and after the change, and so the meaning paradoxes seem to arise. But con-

sider the reason why such a change might occur. In machine terms, we might find that certain applications in observational situations of a given predicate to objects of a certain kind were always contradicted by the best theory for a wide variety of evidence. This would not of course *force* on us a change of disposition to apply that predicate to those objects under the appropriate input, because we expect a small proportion of such applications to be in error relative to the best theory. But if these errors seemed to be concentrated in an unexpected way around certain predicates, we might well decide to change the use of these predicates to fit better the best theories as determined by the large proportion of other observation statements which are assumed true. It might even be possible to state explicit rules for such changes of use and disposition, depending for example on the small probability values of the observation statements involved relative to the rest of the evidence. But all this of course depends on any particular occasion on the presence of many predicates which are not so subject to change of use. The solution of the meaning variance paradoxes requires that there are always many stable predicates when one theory gives way to another.

The possibility of some change of use according to empirically controlled rules shows, however, that Grünbaum's requirement of 'stability of meaning' to save the Q-thesis from triviality is too stringent. Allowing the sort of flexibility of meaning which has obviously often occurred in the development of science need not open the floodgates to apriorism.

4. Summary

In summary let me try to state explicitly the main principles of the new Duhem–Quine empiricism in distinction from the old.

(i) There is no need to make a fundamental epistemological distinction between the theoretical and observational aspects of science, either in regard to decidability of truth-value, or transparency of empirical meaning. The network of relatively observational statements can be imagined to be continuous with a network of theoretical relationships. Indeed much of the recent argument in the literature which has been designed to show that there is no sharp line between theory and observation has depended upon examples of quasi-direct recognition in some circumstances of the empirical applicability of what are normally called

theoretical predicates (such as 'particle-pair annihilation', 'glaciation'). The corresponding theoretical properties cannot, of course, be directly observed independently of the surrounding network of theory and empirical laws, but neither can the so-called observable properties. The difference between them is pragmatic and dependent on causal conditions of sense-perception rather than epistemological.

(ii) The corollary is that empirical applications of observation predicates are not incorrigible, and the empirical laws accepted as holding between them are not infallible. A whole theoretical network may force corrections upon empirical laws in any part of it, but not all, or even most, of it can be corrected at once. Moreover, there is no way of telling *a priori* by separating the theoretical from the observational, *which* part may need correction in the light of subsequent evidence and theory.

(iii) Corrections may strongly suggest changes in the conditions of correct intersubjective application of some of the descriptive predicates, and these changes may be made explicitly according to rules which presuppose that other predicates are not subject to change on the same occasion. To save the notion of 'same theory' which is required to avoid the meaning variance paradoxes, there must be some such stability, indeed the majority of descriptive predicates must be stable in this sense, but just as we do not know *a priori* which observation statements will be retained as true in the next theory, neither do we know which observation predicates will retain stability of meaning. Had Aristotle been a Carnapian, 'heavy' would undoubtedly have appeared in his list of primary observation predicates, and he would have held it to be observable that air is not heavy.

(iv) To avoid total arbitrariness in adoption of the 'best' theory on given sensory input, some prior principles of selection of well-confirmed theories, and criteria for shifts of applicability of some observation predicates, must be assumed. This does not seem, however, to be an objectionable apriorism in the context of the new empiricism, since it is always possible that these principles themselves might change under pressure of the evidence in second or higher order network adjustments.

(v) Lurking within many of these elements of the new empiricism is a systematic conflation of certain aspects of the epistemological problem with causal mechanisms. This occurs at the point of what has been called 'coding' of the input into the coded input, and the identifica-

tion of this process in human observers with the causal process by means of which descriptive language has been learned. Doubtless to the old empiricism this is a fatal cricularity in the network model, because the question will immediately be asked: How do we know anything about the causal coding and the input it processes except in terms of the usual scientific method of observation and theorising? And if this in turn is subject to the conditions of the network model is not the regress irreducibly vicious? Similar objection, it will be recalled, was made to Russell's causal account of the reception of sense data. But there is a crucial difference between the aims of the new empiricism and those of Russell. Russell, in common with most old empiricists, was looking for 'hard data'; new empiricists accept that these are not to be had. This, incidentally, suggests that the approach suggested here to the relatively prior principles of data processing, via a search for a rational inductive logic, is a better reflection of new empiricism than is the purely scientific search for invariants of language which Chomsky favours, or for psychological and machine models of human learning with which some investigators replace the study of inductive logic. Such empirical approaches are always open to the regressive argument, and leave unanswered the question of what prior principles they themselves depend on. The approach via a rational inductive theory, on the other hand, has the merit of exploring possible rational strategies in possible worlds, independently at least of the details of actual learning processes. But it provides no assurance like that sought by old empiricists, that our knowledge of *this* world is firmly based, only that *if* we were given certain interconnected prior conditions, of whose actuality we can never in practice be certain (for example, that the world is not infinitely various), then we could give reasons for our conscious methods of developing science in a world where these conditions obtain. Duhem might hasten to applaud this conclusion as confirming his view that after all scientific knowledge is superficial and transient compared to the revealed truths of a theological metaphysics. We, who do not have this assurance either, must make do with what we have, a poor thing perhaps, but enough.

NOTES

* From The Royal Institute of Philosophy Lectures, *Knowledge and Necessity*, Macmillan and Co., London, Ltd., Vol. III.

[1] *Note added in proof:* Professor Grünbaum has now developed these arguments further in 'Can We Ascertain the Falsity of a Scientific Hypothesis?', Chapter 17 of *Philosophical Problems of Space and Time*, Second, enlarged edition, D. Reidel, Dordrecht-Holland, 1973 [excerpted below – ed.].

BIBLIOGRAPHY

Chomsky, N.: 'Quine's Empirical Assumptions', *Synthese* **19** (1968), 53.

Duhem, P.: *The Aim and Structure of Physical Theory*, Princeton, N.J., 1906; trans. Wiener, Oxford, 1954.

Feyerabend, P. K.: 'Explanation, Reduction and Empiricism', in *Minnesota Studies*, III, ed. by H. Feigl and G. Maxwell, Minneapolis, 1962, p. 28.

Grünbaum, A.: *Philosophical Problems of Space and Time*, New York, 1963.

Grünbaum, A.: 'The Falsifiability of a Component of a Theoretical System', in *Mind, Matter, and Method*, ed. by P. K. Feyerabend and G. Maxwell, Minneapolis, 1966, p. 273.

Hanson, N. R.: *Patterns of Discovery*, Cambridge, 1958.

Kuhn, T. S.: *The Structure of Scientific Revolutions*, Chicago, 1962.

Popper, K. R.: *The Logic of Scientific Discovery*, London, 1959.

Popper, K. R.: *Conjectures and Refutations*, London, 1963.

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

Quine, W. V. O.: *Word and Object*, New York, 1960.

Quine, W. V. O.: 'Replies', *Synthese* **19** (1968), 264.

Scheffler, I.: *Science and Subjectivity*, Indianapolis, 1967.

IMRE LAKATOS

FALSIFICATION AND THE METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES*

1. SCIENCE: REASON OR RELIGION?

For centuries knowledge meant proven knowledge – proven either by the power of the intellect or by the evidence of the senses. Wisdom and intellectual integrity demanded that one must desist from unproven utterances and minimize, even in thought, the gap between speculation and established knowledge. The proving power of the intellect or the senses was questioned by the sceptics more than two thousand years ago; but they were browbeaten into confusion by the glory of Newtonian physics. Einstein's results again turned the tables and now very few philosophers or scientists still think that scientific knowledge is, or can be, proven knowledge. But few realize that with this the whole classical structure of intellectual values falls in ruins and has to be replaced: one cannot simply water down the ideal of proven truth – as some logical empiricists do – to the ideal of 'probable truth'[1] or – as some sociologists of knowledge do – to 'truth by [changing] consensus'.[2]

Popper's distinction lies primarily in his having grasped the full implications of the collapse of the best-corroborated scientific theory of all times: Newtonian mechanics and the Newtonian theory of gravitation. In his view virtue lies not in caution in avoiding errors, but in ruthlessness in eliminating them. Boldness in conjectures on the one hand and austerity in refutations on the other: this is Popper's recipe. Intellectual honesty does not consist in trying to entrench, or establish one's position by proving (or 'probabilifying') it – intellectual honesty consists rather in specifying precisely the conditions under which one is willing to give up one's position. Committed Marxists and Freudians refuse to specify such conditions: this is the hallmark of their intellectual dishonesty. *Belief* may be a regrettably unavoidable biological weakness to be kept under the control of criticism: but *commitment* is for Popper an outright crime.

Kuhn thinks otherwise. He too rejects the idea that science grows by accumulation of eternal truths.[3] He too takes his main inspiration from

Einstein's overthrow of Newtonian physics. His main problem too is *scientific revolution.* But while according to Popper science is 'revolution in permanence', and criticism the heart of the scientific enterprise, according to Kuhn revolution is exceptional and, indeed, extra-scientific, and criticism is, in 'normal' times, anathema. Indeed for Kuhn the transition from criticism to commitment marks the point where progress – and 'normal' science – begins. For him the idea that on 'refutation' one can demand the rejection, the eliminaton of a theory, is 'naive' falsificationism. Criticism of the dominant theory and proposals of new theories are only allowed in the rare moments of 'crisis'. This last Kuhnian thesis has been widely criticized[4] and I shall not discuss it. My concern is rather that Kuhn, having recognized the failure both of justificationism and falsificationism in providing rational accounts of scientific growth, seems now to fall back on irrationalism.

For Popper scientific change is rational or at least rationally reconstructible and falls in the realm of the *logic of discovery.* For Kuhn scientific change – from one 'paradigm' to another – is a mystical conversion which is not and cannot be governed by rules of reason and which falls totally within the realm of the *(social) psychology of discovery.* Scientific change is a kind of religious change.

The clash between Popper and Kuhn is not about a mere technical point in epistemology. It concerns our central intellectual values, and has implications not only for theoretical physics but also for the underdeveloped social sciences and even for moral and political philosophy. If even in science there is no other way of judging a theory but by assessing the number, faith and vocal energy of its supporters, then this must be even more so in the social sciences: truth lies in power. Thus Kuhn's position would vindicate, no doubt, unintentionally, the basic political *credo* of contemporary religious maniacs ('student revolutionaries').

In this paper I shall first show that in Popper's logic of scientific discovery two different positions are conflated. Kuhn understands only one of these, 'naive falsificationism' (I prefer the term 'naive methodological falsificationism'); I think that his criticism of it is correct, and I shall even strengthen it. But Kuhn does not understand a more sophisticated position the rationality of which is not based on 'naive' falsificationism. I shall try to explain – and further strengthen – this stronger Popperian position which, I think, may escape Kuhn's strictures and present scien-

tific revolutions as constituting rational progress rather than as religious conversions.

2. Fallibilism versus falsificationism

(a) *Dogmatic (or Naturalistic) Falsificationism. The Empirical Basis*

To see the conflicting theses more clearly, we have to reconstruct the problem situation as it was in philosophy of science after the breakdown of 'justificationism'.

According to the 'justificationists' scientific knowledge consisted of proven propositions. Having recognized that strictly logical deductions enable us only to infer (transmit truth) but not to prove (establish truth), they disagreed about the nature of those propositions (axioms) whose truth can be proved by extra-logical means. *Classical intellectualists* (or 'rationalists' in the narrow sense of the term) admitted very varied – and powerful – sorts of extralogical 'proofs' by revelation, intellectual intuition, experience. These, with the help of logic, enabled them to prove every sort of scientific proposition. *Classical empiricists* accepted as axioms only a relatively small set of 'factual propositions' which expressed the 'hard facts'. Their truth-value was established by experience and they constituted the *empirical basis* of science. In order to prove scientific *theories* from nothing else but the narrow empirical basis, they needed a logic much more powerful than the deductive logic of the classical intellectualists: '*inductive logic*'. All justificationists, whether intellectualists or empiricists, agreed that a singular statement expressing a 'hard fact' may *disprove* a universal theory[5]; but few of them thought that a finite conjunction of factual propositions might be sufficient to *prove* 'inductively' a universal theory.[6]

Justificationism, that is, the identification of knowledge with proven knowledge, was the dominant tradition in rational thought throughout the ages. Scepticism did not deny justificationism: it only claimed that there was (and could be) no proven knowledge and *therefore* no knowledge whatsoever. For the sceptics 'knowledge' was nothing but animal belief. Thus justificationist scepticism ridiculed objective thought and opened the door to irrationalism, mysticism, superstition.

This situation explains the enormous effort invested by classical rationalists in trying to save the synthetical *a priori* principles of intellectualism

and by classical empiricists in trying to save the certainty of an empirical basis and the validity of inductive inference. For all of them *scientific honesty demanded that one assert nothing that is unproven.* However, both were defeated: Kantians by non-Euclidean geometry and by non-Newtonian physics, and empiricists by the logical impossibility of establishing an empirical basis (as Kantians pointed out, facts cannot prove propositions) and of establishing an inductive logic (no logic can infallibly increase content). It turned out that *all theories are equally unprovable.*

Philosophers were slow to recognize this, for obvious reasons: classical justificationists feared that once they conceded that theoretical science is unprovable, they would have also to concede that it is sophistry and illusion, a dishonest fraud. The philosophical importance of *probabilism* (or '*neojustificationism*') lies in the denial that such a concession is necessary.

Probabilism was elaborated by a group of Cambridge philosophers who thought that although scientific theories are equally unprovable, they have different degrees of probability (in the sense of the calculus of probability) relative to the available empirical evidence.[7] *Scientific honesty then requires less than had been thought: it consists in uttering only highly probable theories: or even in merely specifying, for each scientific theory, the evidence, and the probability of the theory in the light of this evidence.*

Of course, replacing proof by probability was a major retreat for justificationist thought. But even this retreat turned out to be insufficient. It was soon shown, mainly by Popper's persistent efforts, that under very general conditions all theories have zero probability, whatever the evidence; *all theories are not only equally unprovable but also equally improbable.*[8]

Many philosophers still argue that the failure to obtain at least a probabilistic solution of the problem of induction means that we "throw over almost everything that is regarded as knowledge by science and common sense".[9] It is against this background that one must appreciate the dramatic change brought about by falsificationism in evaluating theories, and in general, in the standards of intellectual honesty. Falsificationism was, in a sense, a new and considerable retreat for rational thought. But since it was a retreat from utopian standards, it cleared away much hypocrisy and muddled thought, and thus, in fact, it represented an advance.

First I shall discuss a most important brand of falsificationism: dogmatic (or 'naturalistic') falsificationism. Dogmatic falsificationism admits

the fallibility of *all* scientific theories without qualification, but it retains a sort of infallible empirical basis. It is strictly empiricist without being inductivist: it denies that the certainty of the empirical basis can be transmitted to theories. *Thus dogmatic falsificationism is the weakest brand of justificationism.*

It is extremely important to stress that admitting [fortified] empirical counterevidence as a final arbiter against a theory does not make one a dogmatic falsificationist. Any Kantian or inductivist will agree to such arbitration. But both the Kantian and the inductivist, while bowing to a negative crucial experiment, will also specify conditions of how to establish, entrench one unrefuted theory more than another. Kantians held that Euclidean geometry and Newtonian mechanics were established with certainty; inductivists held they had probability 1. For the dogmatic falsificationist, however, empirical *counter*evidence is the *one and only* arbiter which may judge a theory.

The hallmark of dogmatic falsificationism is then the recognition that all theories are equally conjectural. Science cannot *prove* any theory. But although science cannot *prove*, it can *disprove*: it "can perform with complete logical certainty [the act of] repudiation of what is false",[10] that is there is an absolutely firm empirical basis of facts which can be used to disprove theories. Falsificationists provide new – very modest – standards of scientific honesty: they are willing to regard a proposition as 'scientific' not only if it is a proven factual proposition, but even if it is nothing more than a falsifiable one, that is, if there are factual propositions available at the time with which it may clash, or, in other words, if it has potential falsifiers.[11]

Scientific honesty then consists of specifying, in advance, an experiment such that if the result contradicts the theory, the theory has to be given up.[12] The falsificationist demands that once a proposition is disproved, there must be no prevarication: the proposition must be unconditionally rejected. To (non-tautologous) unfalsifiable propositions the dogmatic falsificationist gives short shrift: he brands them 'metaphysical' and denies them scientific standing.

Dogmatic falsificationists draw a sharp demarcation between the theoretician and the experimenter: the theoretician proposes, the experimenter – in the name of Nature – disposes. As Weyl put it: "I wish to record my unbounded admiration for the work of the experimenter in his

struggle to wrest interpretable facts from an unyielding Nature who knows so well how to meet our theories with a decisive *No* – or with an inaudible *Yes*".[13]Braithwaite gives a particularly lucid exposition of dogmatic falsificationism. He raises the problem of the objectivity of science: "To what extent, then, should an established scientific deductive system be regarded as a free creation of the human mind, and to what extent should it be regarded as giving an objective account of the facts of nature?" His answer is:

> The form of a statement of a scientific hypothesis and its use to express a general proposition, is a human device; what is due to Nature are the observable facts which refute or fail to refute the scientific hypothesis... [In science] we hand over to Nature the task of deciding whether any of the contingent lowest-level conclusions are false. This objective test of falsity it is which makes the deductive system, in whose construction we have very great freedom, a deductive system of scientific hypotheses. Man proposes a system of hypotheses: Nature disposes of its truth or falsity. Man invents a scientific system, and then discovers whether or not it accords with observed fact.[14]

According to the logic of dogmatic falsificationism, science grows by repeated overthrow of theories with the help of hard facts. For instance, according to this view, Descartes's vortex theory of gravity was refuted – and eliminated – by the *fact* that planets moved in ellipses rather than in Carteisan circles; Newton's theory, however, explained successfully the then available facts, both those which had been explained by Descartes's theory and those which refuted it. Therefore Newton's theory replaced Descartes's theory. Analogously, as seen by falsificationists, Newton's theory was, in turn, refuted – proved false – by the anomalous perihelion of Mercury, while Einstein's explained that too. Thus science proceeds by bold speculations, which are never proved or even made probable, but some of which are later eliminated by hard, conclusive refutations and then replaced by still bolder, new and, at least at the start, unrefuted speculations.

Dogmatic falsificationism, however, is untenable. It rests on two false assumptions and on a too narrow criterion of demarcation between scientific and non-scientific.

The *first assumption* is that there is a natural, *psychological* borderline between theoretical or speculative propositions on the one hand and factual or observational (or basic) propositions on the other. (I shall call this – following Popper – the *naturalistic doctrine of observation.*)

The *second assumption* is that if a proposition satisfies the psychological

criterion of being factual or observational (or basic) then it is true; one may say that it was *proved* from facts. (I shall call this the *doctrine of observational (or experimental) proof.*)[15]

These two assumptions secure for the dogmatic falsificationist's deadly disproofs an empirical basis from which proven falsehood can be carried by deductive logic to the theory under test.

These assumptions are complemented by a *demarcation criterion*: only those theories are 'scientific' which forbid certain observable states of affairs and therefore are factually disprovable. *Or, a theory is 'scientific' if it has an empirical basis.*[16]

But both assumptions are false. Psychology testifies against the first, logic against the second, and, finally, methodological judgment testifies against the demarcation criterion. I shall discuss them in turn.

(1) A first glance at a few characteristic examples already undermines the *first assumption.* Galileo claimed that he could 'observe' mountains on the moon and spots on the sun and that these 'observations' refuted the time-honoured theory that celestial bodies are faultless crystal balls. But his 'observations' were not 'observational' in the sense or being observed by the – unaided – senses: their reliability depended on the reliability of his telescope – and of the optical theory of the telescope – which was violently questioned by his contemporaries. It was not Galileo's – pure, untheoretical – *observations* that confronted Aristotelian *theory* but rather Galileo's 'observations' in the light of his optical theory that confronted the Aristotelians' 'observations' in the light of their theory of the heavens.[17] This leaves us with two inconsistent theories, *prima facie* on a par. Some empiricists may concede this point and agree that Galileo's 'observations' were not genuine observations; but they still hold that there is a 'natural demarcation' between statements impressed on an empty and passive mind directly by the senses – only these constitute genuine 'immediate knowledge' – and between statements which are suggested by impure, theory-impregnated sensations. Indeed, *all* brands of justificationist theories of knowledge which acknowledge the senses as a source (whether as *one* source or as *the* source) of knowledge are bound to contain a *psychology of observation.* Such psychologies specify the 'right', 'normal', 'healthy', 'unbiased', 'careful' or 'scientific' state of the senses – or rather the state of mind as a whole – in which they observe truth as it is. For instance, Aristotle – and the Stoics – thought that the

right mind was the medically healthy mind. Modern thinkers recognized that there is more to the right mind than simple 'health'. Descartes's right mind is one steeled in the fire of sceptical doubt which leaves nothing but the final loneliness of the *cogito* in which the *ego* can then be re-established and God's guiding hand found to recognize truth. All schools of modern justificationism can be characterized by the particular *psychotherapy* by which they propose to prepare the mind to receive the grace of proven truth in the course of a mystical communion. In particular, for classical empiricists the right mind is a *tabula rasa*, emptied of all original content, freed from all prejudice of theory. But it transpires from the work of Kant and Popper – and from the work of psychologists influenced by them – that such empiricist psychotherapy can never succeed. For there are and can be no sensations unimpregnated by expectations and therefore *there is no natural (i.e. psychological) demarcation between observational and theoretical propositions.*[18]

(2) But even if there was such a natural demarcation, logic would still destroy the *second assumption* of dogmatic falsificationism. For the truth-value of the 'observational' propositions cannot be indubitably decided: *no factual proposition can ever be proved from an experiment.* Propositions can only be derived from other propositions, they cannot be derived from facts: one cannot prove statements from experiences – "no more than by thumping the table".[19] This is one of the basic points of elementary logic, but one which is understood by relatively few people even today.[20]

If factual propositions are unprovable then they are fallible. If they are fallible then clashes between theories and factual propositions are not 'falsifications' but merely inconsistencies. Our imagination may play a greater role in the formulation of 'theories' than in the formulation of 'factual propositions',[21] but they are both fallible. Thus *we cannot prove theories and we cannot disprove them either.*[22] The demarcation between the soft, unproven 'theories' and the hard, proven 'empirical basis' is non-existent: *all* propositions of science are theoretical and, incurably, fallible.[23]

(3) Finally, even if there were a natural demarcation between observation statements and theories, and even if the truth-value of observation statements could be indubitably established, dogmatic falsificationism would still be useless for eliminating the most important class of what are commonly regarded as scientific theories. For even if experiments *could*

prove experimental reports, their disproving power would still be miserably restricted: *exactly the most admired scientific theories simply fail to forbid any observable state of affairs.*

To support this last contention, I shall first tell a characteristic story and then propose a general argument.

The story is about an imaginary case of planetary misbehaviour. A physicist of the pre-Einsteinian era takes Newton's mechanics and his law of gravitation (N), the accepted initial conditions, I, and calculates, with their help, the path of a newly discovered small planet, p. But the planet deviates from the calculated path. Does our Newtonian physicist consider that the deviation was forbidden by Newton's theory and therefore that, once established, it refutes the theory N? No. He suggests that there must be a hitherto unknown planet p' which perturbs the path of p. He calculates the mass, orbit, etc., of this hypothetical planet and then asks an experimental astronomer to test his hypothesis. The planet p' is so small that even the biggest available telescopes cannot possibly observe it: the experimental astronomer applies for a research grant to build yet a bigger one.[24] In three years' time the new telescope is ready. Were the unknown planet p' to be discovered, it would be hailed as a new victory of Newtonian science. But it is not. Does our scientist abandon Newton's theory and his idea of the perturbing planet? No. He suggests that a cloud of cosmic dust hides the planet from us. He calculates the location and properties of this cloud and asks for a research grant to send up a satellite to test his calculations. Were the satellite's instruments (possibly new ones, based on a little-tested theory) to record the existence of the conjectural cloud, the result would be hailed as an outstanding victory for Newtonian science. But the cloud is not found. Does our scientist abandon Newton's theory, together with the idea of the perturbing planet and the idea of the cloud which hides it? No. He suggests that there is some magnetic field in that region of the universe which disturbed the instruments of the satellite. A new satellite is sent up. Were the magnetic field to be found, Newtonians would celebrate a sensational victory. But it is not. Is this regarded as a refutation of Newtonian science? No. Either yet another ingenious auxiliary hypothesis is proposed or ... the whole story is buried in the dusty volumes of periodicals and the story never mentioned again.[25]

This story strongly suggests that even a most respected scientific theory, like Newton's dynamics and theory of gravitation, may fail to forbid any

observable state of affairs.[26] Indeed, *some scientific theories forbid an event occurring in some specified finite spatio-temporal region (or briefly, a 'singular event') only on the condition that no other factor* (possibly hidden in some distant and unspecified spatio-temporal corner of the universe) *has any influence on it.* But then *such theories never alone contradict a 'basic' statement:* they contradict at most a conjunction of a basic statement describing a spatio-temporally singular event and of a universal non-existence statement saying that no other relevant cause is at work anywhere in the universe. And the dogmatic falsificationist cannot possibly claim that such universal non-existence statements belong to the empirical basis: that they can be observed and proved by experience.

Another way of putting this is to say that some scientific theories are normally interpreted as containing a *ceteris paribus* clause:[27] in such cases it is always a specific theory *together* with this clause which may be refuted. But such a refutation is inconsequential for the *specific* theory under test because by replacing the *ceteris paribus* clause by a different one the *specific* theory can always be retained whatever the tests say.

If so, the 'inexorable' disproof procedure of dogmatic falsificationism breaks down in these cases *even if* there were a firmly established empirical basis to serve as a launching pad for the arrow of the *modus tollens*: the prime target remains hopelessly elusive.[28] And as it happens, it is exactly the most important, 'mature' theories in the history of science which are *prima facie* undisprovable in this way. Moreover, by the standards of dogmatic falsificationism all probabilistic theories also come under this head: for no finite sample can ever *disprove* a universal probabilistic theory;[29] probabilistic theories, like theories with a *ceteris paribus* clause, have no empirical basis. But then the dogmatic falsificationist relegates the most important scientific theories *on his own admission* to metaphysics where rational discussion – consisting, by his standards, of proofs and disproofs – has no place, since a metaphysical theory is neither provable nor disprovable. The demarcation criterion of dogmatic falsificationism is thus still strongly antitheoretical.

(Moreover, *one can easily argue that ceteris paribus clauses are not exceptions, but the rule in science.* Science, after all, must be demarcated from a curiosity shop where funny local – or cosmic – oddities are collected and displayed. The assertion that 'all Britons died from lung cancer between 1950 and 1960 is logically possible, and might even have been true. But

if it has been only an occurrence of an event with minute probability, it would have only curiosity value for the crankish fact-collector, it would have a macabre entertainment value, but no scientific value. A proposition might be said to be scientific only if it aims at expressing a causal connection: such connection between being a Briton and dying of lung cancer may not even be intended. Similarly, 'all swans are white', if true, would be a mere curiosity unless it asserted that swanness *causes* whiteness. But then a black swan would not refute this proposition, since it may only indicate *other causes* operating simultaneously. Thus 'all swans are white' is either an oddity and easily disprovable or a scientific proposition with a *ceteris paribus* clause and therefore undisprovable. *Tenacity of a theory against empirical evidence would then be an argument for rather than against regarding it as 'scientific'. 'Irrefutability' would become a hallmark of science.*[30])

To sum up: classical justificationists only admitted proven theories; neoclassical justificationists probable ones; dogmatic falsificationists realized that in either case no theories are admissible. They decided to admit theories if they are disprovable – disprovable by a finite number of observations. But even if there were such disprovable theories – those which can be contradicted by a finite number of observable facts – they are still logically too near to the empirical basis. For instance, on the terms of the dogmatic falsificationist, a theory like 'All planets move in ellipses' may be disproved by five observations; therefore the dogmatic falsificationist will regard it as scientific. A theory like 'All planets move in circles' may be disproved by four observations; therefore the dogmatic falsificationist will regard it as still more scientific. The acme of scientificness will be a theory like 'All swans are white' which is disprovable by one single observation. On the other hand, he will reject all probabilistic theories together with Newton's, Maxwell's, Einstein's theories, as unscientific, for no finite number of observations can ever disprove them.

If we accept the demarcation criterion of dogmatic falsificationism, *and* also the idea that facts can prove 'factual' propositions, we have to declare that the most important, if not all, theories ever proposed in the history of science are metaphysical, that most, if not all, of the accepted progress is pseudo-progress, that most, if not all, of the work done is irrational. If, however, still accepting the demarcation criterion of dogmatic falsificationism, we deny that facts can prove propositions, then we cer-

tainly end up in complete scepticism: then all science is undoubtedly irrational metaphysics and should be rejected. *Scientific theories are not only equally unprovable, and equally improbable, but they are also equally undisprovable.* But the recognition that not only the theoretical but *all* the propositions in science are fallible, means the total collapse of *all* forms of dogmatic justificationism as theories of scientific rationality.

(b) *Methodological Falsificationism. The 'Empirical Basis'*

The collapse of dogmatic falsificationism because of fallibilistic arguments seems to bring us back to square one. If *all* scientific statements are fallible theories, one can criticize them only for inconsistency. But then, in what sense, if any, is science empirical? If scientific theories are neither provable, nor probabilifiable, nor disprovable, then the sceptics seem to be finally right: science is no more than vain speculation and there is no such thing as progress in scientific knowledge. Can we still oppose scepticism? *Can we save scientific criticism from fallibilism*? Is it possible to have a fallibilistic theory of scientific progress? In particular, if scientific criticism is fallible, on what ground can we ever eliminate a theory?

A most intriguing answer is provided by *methodological falsificationism.* Methodological falsificationism is a brand of conventionalism; therefore in order to understand it, we must first discuss conventionalism in general.

There is an important demarcation between *'passivist' and 'activist' theories of knowledge.* 'Passivists' hold that true knowledge is Nature's imprint on a perfectly inert mind: mental *activity* can only result in bias and distortion. The most influential passivist school is classical empiricism. 'Activists' hold that we cannot read the book of Nature without mental activity, without interpreting them in the light of our expectations or theories.[31] Now *conservative 'activists'* hold that we are born with our basic expectations; with them we turn the world into 'our world' but must then live for ever in the prison of our world. The idea that we live and die in the prison of our 'conceptual frameworks' was developed primarily by Kant; pessimistic Kantians thought that the real world is for ever unknowable because of this prison, while optimistic Kantians thought that God created our conceptual framework to fit the world.[32] But *revolutionary activists* believe that conceptual frameworks can be developed and also replaced by new, *better* ones; it is *we* who create our 'prisons' and we can also, critically, demolish them.[33]

New steps from conservative to revolutionary activism were made by Whewell and then by Poincaré, Milhaud and Le Roy. Whewell held that theories are developed by trial and error – in the 'preludes to the inductive epochs'. The best ones among them are then 'proved' – during the 'inductive epochs' – by a long primarily *a priori* consideration which he called 'progressive intuition'. The 'inductive epochs' are followed by 'sequels to the inductive epochs'; cumulative developments of auxiliary theories.[34] Poincaré, Milhaud and Le Roy were averse to the idea of *proof* by progressive intuition and preferred to explain the continuing historical success of Newtonian mechanics by a *methodological decision* taken by scientists: after a considerable period of initial empirical success scientists may *decide* not to allow the theory to be refuted. Once they have taken this decision, they solve (or dissolve) the apparent anomalies by auxiliary hypotheses or other 'conventionalist stratagems'.[35] This *conservative conventionalism* has, however, the disadvantage of making us unable to get out of our self-imposed prisons, once the first period of trial-and-error is over and the great decision taken. It cannot solve the problem of the elimination of those theories which have been triumphant for a long period. According to conservative conventionalism, experiments may have sufficient power to refute young theories, but not to refute old, established theories: *as science grows, the power of empirical evidence diminishes.*[36]

Poincaré's critics refused to accept his idea, that, although the scientists build their conceptual frameworks, there comes a time when these frameworks turn into prisons which cannot be demolished. This criticism gave rise to two rival schools of *revolutionary conventionalism*: Duhem's simplicism and Popper's methodological falsificationism.[37]

Duhem accepts the conventionalists' position that no physical theory ever crumbles merely under the weight of 'refutations', but claims that it still may crumble under the weight of "continual repairs, and many tangled-up stays" when "the worm-eaten columns" cannot support "the tottering building" any longer[38]; then the theory loses its original simplicity and has to be replaced. But falsification is then left to subjective taste or, at best, to scientific fashion, and leaves too much leeway for dogmatic adherence to a favourite theory.

Popper set out to find a criterion which is both more objective and more hard-hitting. He could not accept the emasculation of empiricism, in-

herent even in Duhem's approach, and proposed a methodology which allows experiments to be powerful even in 'mature' science. Popper's methodological falsificationism is both conventionalist and falsificationist, but he "differs from the [conservative] conventionalists in holding that the statements decided by agreement are *not* [spatio-temporally] universal but [spatio-temporally] singular"[39]; and he differs from the dogmatic falsificationist in holding that the truth-value of such statements cannot be proved by facts but, in some cases, may be decided by agreement.[40]

The *conservative conventionalist* (or methodological justificationist, if you wish) makes unfalsifiable by *fiat* some (spatio-temporally) universal theories, which are distinguished by their explanatory power, simplicity or beauty. Our *revolutionary conventionalist* (or 'methodological falsificationist') makes unfalsifiable by *fiat* some (spatio-temporally) singular statements which are distinguishable by the fact that there exists at the time a 'relevant technique' such that 'anyone who has learned it' will be able to *decide* that the statement is 'acceptable'.[41] Such a statement may be called an 'observational' or 'basic' statement, but only in inverted commas.[42] Indeed, the very selection of all such statements is a matter of a decision, which is not based on exclusively psychological considerations. This decision is then followed by a second kind of decision concerning the separation of the set of *accepted* basic statements from the rest.

These *two decisions* correspond to the *two assumptions* of dogmatic falsificationism. But there are important differences. First, the methodological falsificationist is not a justificationist, he has no illusions about 'experimental proofs' and is fully aware of the fallibility of his decisions and the risks he is taking.

The methodological falsificationist realizes that in the 'experimental techniques' of the scientist fallible theories are involved,[43] 'in the light of which' he interprets the facts. In spite of this he 'applies' these theories, he regards them in the given context not as theories under test but as *unproblematic background knowledge* "which we accept (tentatively) as unproblematic while we are testing the theory".[44] He may call these theories – and the statements whose truth-value he decides in their light – 'observational': but this is only a manner of speech which he inherited from naturalistic falsificationism.[45] The methodological falsificationist *uses our most successful theories as extensions of our senses* and widens the

range of theories which can be applied in testing far beyond the dogmatic falsificationist's range of strictly observational theories. For instance, let us imagine that a big radio-star is discovered with a system of radio-star satellites orbiting it. We should like to test some gravitational theory on this planetary system – a matter of considerable interest. Now let us imagine that Jodrell Bank succeeds in providing a set of space-time co-ordinates of the planets which is inconsistent with the theory. We shall take these statements as potential falsifiers. Of course, these basic statements are not 'observational' in the usual sense but only "'observational'". They describe planets that neither the human eye nor optical instruments can reach. Their truth-value is arrived at by an 'experimental technique'. This 'experimental technique' is based on the 'application' of a well-corroborated theory of radio-optics. Calling these statements 'observational' is no more than a manner of saying that, in the context of his problem, that is, in testing our gravitational theory, the methodological falsificationist uses radio-optics uncritically, as 'background knowledge'. *The need for decisions to demarcate the theory under test from unproblematic background knowledge is a characteristic feature of this brand of methodological falsificationism.*[46] (This situation does not really differ from Galileo's 'observation' of Jupiter's satellites: moreover, as some of Galileo's contemporaries rightly pointed out, he relied on a virtually non-existent optical theory – which then was less corroborated, and even less articulated, than present-day radio-optics. On the other hand, calling the reports of our human eye 'observational' only indicates that we 'rely' on some vague physiological theory of human vision.[47])

This consideration shows the conventional element in granting – in a given context – the (methodologically) 'observational' status to a theory.[48] Similarly, there is a considerable conventional element in the decision concerning the actual truth-value of a basic statement which we take after we have decided which 'observational theory' to apply. One single observation may be the stray result of some trivial error: in order to reduce such risks, methodological falsificationists prescribe some safety control. The simplest such control is to repeat the experiment (it is a matter of convention how many times); another is to 'fortify' the potential falsifier by a 'well-corroborated falsifying hypothesis'.[49]

The methodological falsificationist also points out that, as a matter of fact, these conventions are institutionalized and endorsed by the scientific

community; the list of 'accepted' falsifiers is provided by the verdict of the experimental scientists.[50]

This is how the methodological falsificationist establishes his 'empirical basis'. (He uses inverted commas in order 'to give ironical emphasis' to the term.[51]) This 'basis' can be hardly called a 'basis' by justificationist standards: there is nothing proven about it – it denotes 'piles driven into a swamp'.[52] Indeed, if this 'empirical basis' clashes with a theory, the theory may be *called* 'falsified', but it is not falsified in the sense that it is disproved. Methodological 'falsification' is very different from dogmatic falsification. If a theory is falsified, it is proven false; if it is 'falsified', it may still be true. If we follow up this sort of 'falsification' by the actual 'elimination' of a theory, we may well end up by eliminating a true, and accepting a false, theory (a possibility which is thoroughly abhorrent to the old-fashioned justificationist).

Yet the methodological falsificationist advises that exactly this is to be done. The methodological falsificationist realizes that if we want to reconcile fallibilism with (non-justificationist) rationality, we *must* find a way to eliminate *some* theories. If we do not succeed, the growth of science will be nothing but growing chaos.

Therefore the methodological falsificationist maintains that "[if we want] to make the method of selection by elimination work, and to ensure that only the fittest theories survive, their struggle for life must be made severe".[53] Once a theory has been falsified, in spite of the risk involved, it must be eliminated: "[with theories we work only] as long as they stand up to tests".[54] The elimination must be methodologically conclusive: "In general we regard an inter-subjectively testable falsification as final... A corroborative appraisal made at a later date... can replace a positive degree of corroboration by a negative one, but not *vice versa*".[55] This is the methodological falsificationist's explanation of how we get out of a rut: "It is always the experiment which saves us from following a track that leads nowhere".[56]

The methodological falsificationist separates rejection and disproof, which the dogmatic falsificationist had conflated.[57] He is a fallibilist but his fallibilism does not weaken his critical stance: he turns fallible propositions into a 'basis' for a hard-line policy. On these grounds he proposes a *new demarcation criterion*: only those theories – that is, non-'observational' propositions – which forbid certain 'observable 'states of affairs,

and therefore may be 'falsified' and rejected, are 'scientific': or, briefly, *a theory is 'scientific' (or 'acceptable') if it has an 'empirical basis'*. This criterion brings out sharply the difference between dogmatic and methodological falsificationism.[58]

This methodological demarcation criterion is much more liberal than the dogmatic one. Methodological falsificationism opens up new avenues of criticism: many more theories may qualify as 'scientific'. We have already seen that there are more 'observational' theories than observational theories, and therefore there are more 'basic' statements than basic statements. [59] Furthermore, probabilistic theories may qualify now as 'scientific': although they are not falsifiable they can be easily made 'falsifiable' by an *additional (third type) decision* which the scientist can make by specifying certain rejection rules which may make statistically interpreted evidence 'inconsistent' with the probabilistic theory.[60]

But even these three decisions are not sufficient to enable us to 'falsify' a theory which cannot explain anything 'observable' without a *ceteris paribus* clause. No finite number of 'observations' is enough to 'falsify' such a theory. However, if this is the case how can one reasonably defend a methodology which claims to "interpret natural laws or theories as... statements which are partially decidable, i.e. which are, for logical reasons, not verifiable but, in an asymmetrical way, falsifiable..."?[61] How can we interpret theories like Newton's theory of dynamics and gravitation as 'one-sidedly decidable'?[62] How can we make in such cases genuine "attempts to weed out false theories – to find the weak points of a theory in order to reject it if it is falsified by the test"?[63] How can we draw them into the realm of rational discussion? The methodological falsificationist solves the problem by making a further *(fourth type) decision*: when he tests a theory together with a *ceteris paribus* clause and finds that this conjunction has been refuted, he must decide whether to take the refutation also as a refutation of the specific theory. For instance, he may accept Mercury's 'anomalous' perihelion as a refutation of the treble conjunction N_3 of Newton's theory, the known initial conditions and the *ceteris paribus* clause. Then he tests the initial conditions 'severely'[64] and may decide to relegate them into the 'unproblematic background knowledge'. This decision implies the refutation of the double conjunction N_2 of Newton's theory and the *ceteris paribus* clause. Now he has to take the crucial decision: whether to relegate also the *ceteris paribus* clause into

the pool of 'unproblematic background knowledge'. He will do so if he finds the *ceteris paribus* clause well corroborated.

How can one test a *ceteris paribus* clause severely? By assuming that there *are* other influencing factors, by specifying such factors, and by testing these specific assumptions. If many of them are refuted, the *ceteris paribus* clause will be regarded as well-corroborated.

Yet the decision to 'accept' a *ceteris paribus* clause is a very risky one because of the grave consequences it implies. If it is decided to accept it as part of such background knowledge, the statements describing Mercury's perihelion from the empirical basis of N_2 are turned into the empirical basis of Newton's specific theory N_1 and what was previously a mere 'anomaly' in relation to N_1, becomes now crucial evidence against it, its falsification. (We may call an event described by a statement A an '*anomaly* in relation to a theory T' if A is a potential falsifier of the conjunction of T and a *ceteris paribus* clause but it becomes a potential falsifier of T itself after having decided to relegate the *ceteris paribus* clause into "unproblematic background knowledge".) Since, for our savage falsificationist, falsifications are methodologically conclusive, the fateful decision amounts to the methodological elimination of Newton's theory, making further work on it irrational. If the scientist shrinks back from such bold decisions he will "never benefit from experience", "believing, perhaps, that it is his business to defend a successful system against criticism as long as it is not *conclusively disproved*".[65] He will degenerate into an apologist who may always claim that "the discrepancies which are asserted to exist between the experimental results and the theory are only apparent and they will disappear with the advance of our understanding."[66] But for the falsificationist this is "the very reverse of the critical attitude which is the proper one for the scientist",[67] and is impermissible. To use one of the methodological falsificationist's favourite expressions: the theory "must be made to stick its neck out".

The methodological falsificationist is in a serious plight when it comes to deciding where to draw the demarcation, even if only in a well-defined context, between the problematic and unproblematic. The plight is most dramatic when he has to make a decision about *ceteris paribus* clauses, when he has to promote one of the hundreds of 'anomalous phenomena' into a 'crucial experiment', and decide that in such a case the experiment was 'controlled'.[68]

Thus, with the help of this fourth type of decision,[69] our methodological falsificationist has finally succeeded in interpreting even theories like Newton's theory as 'scientific'.[70]

Indeed, there is no reason why he should not go yet another step. Why not decide that a theory – which even these four decisions cannot turn into an empirically falsifiable one – is falsified if it clashes with another theory which is scientific on some of the previously specified grounds and is also well-corroborated?[71] After all, if we reject one theory because one of its potential falsifiers is seen to be true in the light of an observational theory, why not reject another theory because it clashes *directly* with one that may be relegated into unproblematic background knowledge? This would allow us, by a *fifth type decision*, to eliminate even 'syntactically metaphysical' theories, that is, theories, which, like 'all-some' statements or purely existential statements,[72] because of their *logical form* cannot have spatio-temporally singular potential falsifiers.

To sum up: the methodological falsificationist offers an interesting solution to the problem of combining hard-hitting criticism with fallibilism. Not only does he offer a philosophical basis for falsification after fallibilism had pulled the carpet from under the feet of the dogmatic falsificationist, but he also widens the range of such criticism very considerably. By putting falsification in a new setting, he saves the attractive code of honour of the dogmatic falsificationist: that scientific honesty consists in specifying, in advance, an experiment such, that if the result contradicts the theory, the theory has to be given up.

Methodological falsificationism represents a considerable advance beyond both dogmatic falsificationism and conservative conventionalism. It recommends risky decisions. But the risks are daring to the point of recklessness and one wonders whether there is no way of lessening them.

Let us first have a closer look at the risks involved.

Decisions play a crucial role in this methodology – as in any brand of conventionalism. Decisions however may lead us disastrously astray. The methodological falsificationist is the first to admit this. But this, he argues, is the price which we have to pay for the possibility of progress.

One has to appreciate the dare-devil attitude of our methodological falsificationist. He feels himself to be a hero who, faced with two catastrophic alternatives, dared to reflect coolly on their relative merits and choose the lesser evil. One of the alternatives was sceptical fallibilism,

with its 'anything goes' attitude, the despairing abandonment of all intellectual standards, and hence of the idea of scientific progress. Nothing can be established, nothing can be rejected, nothing even communicated: the growth of science is a growth of chaos, a veritable Babel. For two thousand years, scientists and scientifically-minded philosophers chose justificationist illusions of some kind to escape this nightmare. Some of them argued that *one has to choose between inductivist justificationism and irrationalism*: "I do not see any way out of a dogmatic assertion that we know the inductive principle or some equivalent; the only alternative is to throw over almost everything that is regarded as knowledge by science and common sense".[73] Our methodological falsificationist proudly rejects such escapism: he dares to measure up to the full impact of fallibilism and yet escape scepticism by a daring and risky conventionalist policy, with no dogmas. He is fully aware of the risks but insists that *one has to choose between some sort of methodological falsificationism and irrationalism.* He offers a game in which one has little hope of winning, but claims that it is still better to play than to give up.[74]

Indeed, those critics of naive falsificationism who offer no alternative method of criticism are inevitably driven to irrationalism. For instance, Neurath's muddled argument, that the falsification and ensuing elimination of a hypothesis may turn out to have been "an obstacle in the progress of science",[75] carries no weight as long as the only alternative he seems to offer is chaos. Hempel is, no doubt, right in stressing that "science offers various examples [when] a conflict between a highly-confirmed theory and an occasional recalcitrant experiential sentence may well be resolved by revoking the latter rather than by sacrificing the former"[76]; nevertheless he admits that he can offer no other 'fundamental standard' than that of naive falsificationism.[77] Neurath – and, seemingly, Hempel – reject falsificationism as 'pseudo-rationalism'[78]; but where is 'real rationalism'? Popper warned already in 1934 that Neurath's permissive methodology (or rather lack of methodology) would make science unempirical and therefore irrational:

> We need a set of rules to limit the arbitrariness of 'deleting' (or else 'accepting') a protocol sentence. Neurath fails to give any such rules and thus unwittingly throws empiricism overboard... Every system becomes defensible if one is allowed (as everybody is, in Neurath's view) simply to 'delete' a protocol sentence if it is inconvenient.[79]

Popper agrees with Neurath that all propositions are fallible; but he

forcefully makes the crucial point that we cannot make progress unless we have a firm rational strategy or method to guide us when they clash.[80]

But is not the firm strategy of the brand of methodological falsificationism hitherto discussed *too firm*? Are not the decisions it advocates bound to be *too arbitrary*? Some may even claim that all that distinguishes methodological from dogmatic falsificationism is that *it pays lip-service to fallibilism*!

To criticize a theory of criticism is usually very difficult. Naturalistic falsificationism was relatively easy to refute, since it rested on an empirical psychology of perception: one could show that it was simply *false*. But how can methodological falsificationism be falsified? No disaster can ever disprove a non-justificationist theory of rationality. Moreover, how can we ever recognize an epistemological disaster? We have no means to judge whether the verisimilitude of our successive theories increases or decreases.[81] At this stage we have not yet developed a general theory of criticism even for scientific theories, let alone for theories of rationality[82]; therefore if we want to falsify our methodological falsificationism, we have to do it before having a theory of how to do it.

If we look at history of science, if we try to see how some of the most celebrated falsifications happened, we have to come to the conclusion that either some of them are plainly irrational, or that they rest on rationality principles radically different from the ones we just discussed. First of all, our falsificationist must deplore the fact that stubborn theoreticians frequently challenge experimental verdicts and have them reversed. In the falsificationist conception of scientific 'law and order' we have described there is no place for such successful appeals. Further difficulties arise from the falsification of theories to which a *ceteris paribus* clause is appended.[83] Their falsification as it occurs in actual history is *prima facie* irrational by the standards of our falsificationist. By his standards, scientists frequently seem to be irrationally slow: for instance, eighty-five years elapsed between the acceptance of the perihelion of Mercury as an anomaly and its acceptance as a falsification of Newton's theory, in spite of the fact that the *ceteris paribus* clause was reasonably well corroborated. On the other hand, scientists frequently seem to be irrationally rash: for instance, Galileo and his disciples accepted Copernican heliocentric celestial mechanics in spite of the abundant evidence against the rotation of the Earth; or Bohr and his disciples accepted a theory of light emission in

spite of the fact that it ran counter to Maxwell's well-corroborated theory.

Indeed, it is not difficult to see at least two crucial characteristics common to both dogmatic and our methodological falsificationism which are clearly dissonant with the actual history of science: that (1) *a test is – or must be made – a two-cornered fight between theory and experiment so that in the final confrontation only these two face each other; and (2) the only interesting outcome of such confrontation is (conclusive) falsification: "[the only genuine] discoveries are refutations of scientific hypotheses"*.[84] However, history of science suggests that (1′) tests are – at least – three-cornered fights between rival theories and experiment and (2′) some of the most interesting experiments result, *prima facie*, in confirmation rather than falsification.

But if – as seems to be the case – the history of science does not bear out our theory of scientific rationality, we have two alternatives. One alternative is to abandon efforts to give a rational explanation of the success of science. Scientific method (or 'logic of discovery'), conceived as the discipline of rational appraisal of scientific theories – and of criteria of *progress* – vanishes. We, may, of course, still try to explain *changes* in 'paradigms' in terms of social psychology.[85] This is Polanyi's and Kuhn's way.[86] The other alternative is to try at least to *reduce* the conventional element in falsificationism (we cannot possibly eliminate it) and replace the *naive* versions of methodological falsificationism – characterized by the theses (1) and (2) above – by a *sophisticated* version which would give a new *rationale* of falsification and thereby rescue methodology and the idea of scientific *progress*. This is Popper's way, and the one I intend to follow.

(c) *Sophisticated Versus Naive Methodological Falsificationism. Progressive and Degenerating Problemshifts.*

Sophisticated falsificationism differs from naive falsificationism both in its rules of *acceptance* (or 'demarcation criterion') and its rules of *falsification* or elimination. For the naive falsificationist any theory which can be interpreted as experimentally falsifiable, is 'acceptable' or 'scientific'. For the sophisticated falsificationist a theory is 'acceptable' or 'scientific' only if it has corroborated excess empirical content over its predecessor (or rival), that is, only if it leads to the discovery of novel facts. This condition can be analysed into two clauses: that the new theory has excess empirical content ('*acceptability*'$_1$) and that some of this excess content is

verified ('*acceptability*'$_2$). The first clause can be checked instantly by *a priori* logical analysis; the second can be checked only empirically and this may take an indefinite time.

Again, for the naive falsificationist a theory is *falsified* by a ('fortified') 'observational' statement which conflicts with it (or rather, which he decides to interpret as conflicting with it). The sophisticated falsificationist regards a scientific theory T as falsified if and only if another theory T' has been proposed with the following characteristics: (1) T' has excess empirical content over T: that is, it predicts *novel* facts, that is, facts improbable in the light of, or even forbidden, by T;[87] (2) T' explains the previous success of T, that is, all the unrefuted content of T is contained (within the limits of observational error) in the content of T'; and (3) some of the excess content of T' is corroborated.[88]

In order to be able to appraise these definitions we need to understand their problem background and their consequences. First, we have to remember the conventionalists' methodological discovery that no experimental result can ever kill a theory: any theory can be saved from counter-instances either by some auxiliary hypothesis or by a suitable reinterpretation of its terms. Naive falsificationists solved this problem by relegating – in crucial contexts – the auxiliary hypotheses to the realm of unproblematic background knowledge, eliminating them from the deductive model of the test-situation and thereby *forcing* the chosen theory into logical isolation, in which it becomes a sitting target for the attack of test-experiments. But since this procedure did not offer a suitable guide for a rational reconstruction of the history of science, we may just as well completely rethink our approach. Why aim at falsification at any price? Why not rather impose certain standards on the theoretical adjustments by which one is allowed to save a theory? Indeed, some such standards have been well-known for centuries, and we find them expressed in age-old wisecracks against *ad hoc* explanations, empty prevarications, face-saving, linguistic tricks.[89] We have already seen that Duhem adumbrated such standards in terms of 'simplicity' and 'good sense'. But *when* does lack of 'simplicity' in the protective belt of theoretical adjustments reach the point at which the theory *must* be abandoned?[90] In what sense was Copernican theory, for instance, 'simpler' than Ptolemaic?[91] The vague notion of Duhemian 'simplicity' leaves, as the naive falsificationist correctly argued, the decision very much to taste and fashion.

Can one improve on Duhem's approach? Popper did. His solution – a sophisticated version of methodological falsificationism – is more objective and more rigorous. Popper agrees with the conventionalists that theories and factual propositions can always be harmonized with the help of auxiliary hypotheses: he agrees that the problem is how to demarcate between scientific and pseudoscientific *adjustments*, between rational and irrational changes of theory. According to Popper, saving a theory with the help of auxiliary hypotheses which satisfy certain well-defined conditions represents scientific progress; but saving a theory with the help of auxiliary hypotheses which do not, represents degeneration. Popper calls such inadmissible auxiliary hypotheses *ad hoc* hypotheses, mere linguistic devices, 'conventionalist stratagems'.[92] But then any scientific theory has to be appraised rogether with its auxiliary hypotheses, initial conditions, etc., and, especially, together with its predecessors so that we may see by what sort of *change* it was brought about. Then, of course, what we appraise is a *series of theories* rather than isolated *theories*.

Now we can easily understand why we formulated the criteria of acceptance and rejection of sophisticated methodological falsificationism as we did. But it may be worth while to reformulate them slightly, couching them explicitly in terms of *series of theories*.

Let us take a series of theories, T_1, T_2, T_3,... where each subsequent theory results from adding auxiliary clauses to (or from semantical reinterpretations of) the previous theory in order to accommodate some anomaly, each theory having at least as much content as the unrefuted content of its predecessor. Let us say that such a series of theories is *theoretically progressive (or 'constitutes a theoretically progresssive problemshift')* if each new theory has some excess empirical content over its predecessor, that is, if it predicts some novel, hitherto unexpected fact. Let us say that a theoretically progressive series of theories is also *empirically progressive (or 'constitutes an empirically progressive problemshift')* if some of this excess empirical content is also corroborated, that is, if each new theory leads us to the actual discovery of some *new fact*.[93] Finally, let us call a problemshift *progressive* if it is both theoretically and empirically progressive, and *degenerating* if it is not.[94] We '*accept*' problemshifts as 'scientific 'only if they are at least theoretically progressive; if they are not, we '*reject*' them as 'pseudoscientific'. Progress is measured by the degree to which a problemshift is progressive, by the degree

to which the series of theories leads us to the discovery of novel facts. We regard a theory in the series 'falsified' when it is superseded by a theory with higher corroborated content.

This demarcation between progressive and degenerating problemshifts sheds new light on the appraisal of *scientific – or, rather, progressive – explanations*. If we put forward a theory to resolve a contradiction between a previous theory and a counterexample in such a way that the new theory, instead of offering a content-increasing (scientific) *explanation*, only offers a content-decreasing (linguistic) *reinterpretation*, the contradiction is resolved in a merely semantical, unscientific way. *A given fact is explained scientifically only if a new fact is also explained with it.*[95]

Sophisticated falsificationism thus shifts the problem of how to appraise *theories* to the problem of how to appraise *series of theories*. Not an isolated *theory*, but only a series of theories can be said to be scientific or unscientific: to apply the term 'scientific' to one *single* theory is a category mistake.[96]

The time-honoured empirical criterion for a satisfactory theory was agreement with the observed facts. Our empirical criterion for a series of theories is that it should produce new facts. *The idea of growth and the concept of empirical character are soldered into one.*

This revised form of methodological falsificationism has many new features. First, it denies that "in the case of a scientific theory, our decision depends upon the results of experiments. If these confirm the theory, we may accept it until we find a better one. If they contradict the theory, we reject it".[97] It denies that "what ultimately decides the fate of a theory is the result of a test, i.e. an agreement about basic statements".[98] Contrary to naive falsificationism, *no experiment, experimental report, observation statement or well-corroborated low-level falsifying hypothesis alone can lead to falsification. There is no falsification before the emergence of a better theory.*[99] But then the distinctively negative character of naive falsificationism vanishes; criticism becomes more difficult, and also positive, constructive. But, of course, if falsification depends on the emergence of better theories, on the invention of theories which anticipate new facts, then falsification is *not* simply a relation between a theory and the empirical basis, but a multiple relation between competing theories, the original 'empirical basis', and the empirical growth resulting from the competition. Falsification can thus be said to have a '*historical character*'.[100]

Moreover, some of the theories which bring about falsification are frequently proposed *after* the 'counterevidence'. This may sound paradoxical for people indoctrinated with naive falsificationism. Indeed, this epistemological theory of the relation between theory and experiment differs sharply from the epistemological theory of naive falsificationism. The very term 'counterevidence' has to be abandoned in the sense that no experimental result must be interpreted directly as 'counterevidence'. If we still want to retain this time-honoured term, we have to redefine it like this: 'counterevidence to T_1' is a corroborating instance to T_2 which is either inconsistent with or independent of T_1 (with the *proviso* that T_2 is a theory which satisfactorily explains the empirical success of T_1). This shows that '*crucial counterevidence*' – or '*crucial experiments*' – can be recognized as such among the scores of anomalies only *with hindsight*, in the light of some superseding theory.[101]

Thus the crucial element in falsification is whether the *new theory* offers any novel, excess information compared with its predecessor and whether some of this excess information is corroborated. Justificationists valued 'confirming' instances of a theory; naive falsificationists stressed 'refuting' instances; for the methodological falsificationists it is the – rather rare – corroborating instances of the *excess* information which are the crucial ones; these receive all the attention. We are no longer interested in the thousands of trivial verifying instances nor in the hundreds of readily available anomalies: the few crucial *excess-verifying instances* are decisive.[102] This consideration rehabilitates – and reinterprets – the old proverb: *Exemplum docet, exempla obscurant.*

'Falsification' in the sense of naive falsificationism (corroborated counterevidence) is not a *sufficient* condition for eliminating a specific theory: in spite of hundreds of known anomalies we do not regard it as falsified (that is, eliminated) until we have a better one.[103] Nor is 'falsification' in the naive sense *necessary* for falsification in the sophisticated sense: a progressive problemshift does not have to be interspersed with 'refutations'. Science can grow without any 'refutations' leading the way. Naive falsificationists suggest a linear growth of science, in the sense that theories are followed by powerful refutations which eliminate them; these refutations in turn are followed by new theories.[104] It is perfectly *possible* that theories be put forward 'progressively' in such a rapid succession that the 'refutation' of the *n*-th appears only as the corroboration of the

$n+1$th. The problem fever of science is raised by proliferation of rival theories rather than counterexamples or anomalies.

This shows that the slogan of *proliferation of theories* is much more important for sophisticated than for naive falsificationism. For the naive falsificationist science grows through repeated experimental overthrow of theories; new rival theories proposed before such 'overthrows' may speed up growth but are not absolutely necessary[105], constant proliferation of theories is optional but not mandatory. For the sophisticated falsificationist proliferation of theories cannot wait until the accepted theories are 'refuted' (or until their protagonists get into a Kuhnian crisis of confidence).[106] While naive falsificationism stresses "the urgency of replacing a *falsified* hypothesis by a better one",[107] sophisticated falsificationism stresses the urgency of replacing *any* hypothesis by a better one. Falsification cannot "compel the theorist to search for a better theory",[108] simply because falsification cannot precede the better theory.

The problem-shift from naive to sophisticated falsificationism involves a semantic difficulty. For the naive falsificationist a 'refutation' is an experimental result which, by force of his decisions, is made to conflict with the theory under test. But according to sophisticated falsificationism one must not take such decisions before the alleged 'refuting instance' has become the confirming instance of a new, better theory. Therefore whenever we see terms like 'refutation', 'falsification', 'counterexample', we have to check in each case whether these terms are being applied in virtue of decisions by the naive or by the sophisticated falsificationist.[109]

Sophisticated methodological falsificationism offers new standards for intellectual honesty. Justificationist honesty demanded the acceptance of only what was proven and the rejection of everything unproven. Neo-justificationist honesty demanded the specification of the probability of any hypothesis in the light of the available empirical evidence. The honesty of naive falsificationism demanded the testing of the falsifiable and the rejection of the unfalsifiable and the falsified. Finally, the honesty of sophisticated falsificationism demanded that one should try to look at things from different points of view, to put forward new theories which anticipate novel facts, and to reject theories which have been superseded by more powerful ones.

Sophisticated methodological falsificationism blends several different

traditions. From the empiricists it has inherited the determination to learn primarily from experience. From the Kantians it has taken the activist approach to the theory of knowledge. From the conventionalists it has learned the importance of decisions in methodology.

I should like to emphasize here a further distinctive feature of sophisticated methodological empiricism: the crucial role of excess corroboration. For the inductivist, learning about a new theory is learning how much confirming evidence supports it; about refuted theories one *learns* nothing (learning, after all, is to build up proven or probable *knowledge*). For the dogmatic falsificationist, learning about a theory is learning whether it is refuted or not; about confirmed theories one learns nothing (one cannot prove or probabilify anything), about refuted theories one learns that they are disproved.[110] For the sophisticated falsificationist, learning about a theory is primarily learning which new facts it anticipated: indeed, for the sort of Popperian empiricism I advocate, the only relevant evidence is the evidence anticipated by a theory, and *empiricalness (or scientific character) and theoretical progress are inseparably connected.*[111]

This idea is not entirely new. Leibnitz, for instance, in his famous letter to Conring in 1678, wrote: "It is the greatest commendation of an hypothesis (next to [proven] truth) if by its help predictions can be made even about phenomena or experiments not tried."[112] Leibnitz's view was widely accepted by scientists. But since appraisal of a scientific theory, before Popper, meant appraisal of its degree of justification, this position was regarded by some logicians as untenable. Mill, for instance, complains in 1843 in horror that "it seems to be thought that an hypothesis... is entitled to a more favourable reception, if besides accounting for all the facts previously known, it has led to the anticipation and prediction of others which experience afterwards verified".[113] Mill had a point: this appraisal was in conflict both with justificationism and with probabilism: why should an event *prove* more, if it was anticipated by the theory than if it was known already before? As long as *proof* was the only criterion of the scientific character of a theory, Leibnitz's criterion could only be regarded as irrelevant.[114] Also, the *probability* of a theory given evidence cannot possibly be influenced, as Keynes pointed out, by *when* the evidence was produced: the probability of a theory given evidence can depend only on the theory and the evidence,[115] and not upon whether the evidence was produced before or after the theory.

In spite of this convincing justificationist criticism, the criterion survived among some of the best scientists, since it formulated their strong dislike of merely *ad hoc* explanations, which "though [they] truly express the facts [they set out to explain, are] not born out by any other phenomena".[116]

But it was only Popper who recognized that the *prima facie* inconsistency between the few odd, casual remarks against *ad hoc* hypotheses on the one hand and the huge edifice of justificationist philosophy of knowledge must be solved by demolishing justificationism and by introducing new, nonjustificationist criteria for appraising scientific theories based on anti-adhocness.

Let us look at a few examples. Einstein's theory is not better than Newton's *because* Newton's theory was 'refuted' but Einstein's was not: there are many known 'anomalies' to Einsteinian theory. Einstein's theory is better than – that is, represents progress compared with – Newton's theory *anno 1916* (that is, Newton's laws of dynamics, law of gravitation, the known set of initial conditions; 'minus' the list of known anomalies such as Mercury's perihelion) *because* it explained everything that Newton's theory had successfully explained, and it explained also *to some extent* some known anomalies and, in addition, forbade events like transmission of light along straight lines near large masses about which Newton's theory had said nothing but which had been permitted by other well-corroborated scientific theories of the day; moreover, *at least some* of the unexpected excess Einsteinian content was in fact *corroborated* (for instance, by the eclipse experiments).

On the other hand, according to these sophisticated standards, Galileo's theory that the natural motion of terrestrial objects was circular, introduced no improvement since it did not forbid anything that had been not forbidden by the relevant theories he intended to improve upon (that is, by Aristotelian physics and by Copernican celestial kinematics). This theory was therefore *ad hoc* and therefore – from the heuristic point of view – valueless.[117]

A beautiful example of a theory which satisfied only the first part of Popper's criterion of progress (excess content) but not the second part (corroborated excess content) was given by Popper himself: the Bohr-Kramers–Slater theory of 1924. This theory was refuted in *all* its new predictions.[118]

Let us finally consider how much conventionalism remains in sophisticated falsificationism. Certainly *less* than in naive falsificationism. We need *fewer* methodological decisions. The '*fourth-type decision*' which was essential for the naive version has become completely redundant. To show this we only have to realize that if a scientific theory, consisting of some 'laws of nature', initial conditions, auxiliary theories (but without a *ceteris paribus* clause) conflicts with some factual propositions we do not have to decide which – explicit or 'hidden' – part to replace. We may try to replace *any* part and only when we have hit on an explanation of the anomaly with the help of some content-increasing change (or auxiliary hypothesis), and nature corroborates it, do we move on to eliminate the 'refuted' complex. Thus sophisticated falsification is a slower but possibly safer process than naive falsification.

Let us take an example. Let us assume that the course of a planet differs from the one predicted. Some conclude that this refutes the dynamics and gravitational theory applied: the initial conditions and the *ceteris paribus* clause have been ingeniously corroborated. Others conclude that this refutes the initial conditions used in the calculations: dynamics and gravitational theory have been superbly corroborated in the last two hundred years and all suggestions concerning further factors in play failed. Yet others conclude that this refutes the underlying assumption that there were no other factors in play except for those which were taken into account: these people may possibly be motivated by the metaphysical principle that any explanation is only approximative because of the infinite complexity of the factors involved in determining any single event. Should we praise the first type as '*critical*', scold the second type as '*hack*', and condemn the third as '*apologetic*'? No. We do not need to draw any conclusions about such 'refutation'. We never reject a specific theory simply by *fiat*. If we have an inconsistency like the one mentioned, we do not have to decide which ingredients of the theory we regard as problematic and which ones as unproblematic: we regard all ingredients as problematic in the light of the conflicting accepted basic statement and try to replace all of them. If we succeed in replacing some ingredient in a 'progressive' way (that is, the replacement has more corroborated empirical content than the original), we call it 'falsified'.

We do not need the *fifth type decision* of the naive falsificationist either. In order to show this let us have a new look at the problem of the appraisal

of (syntactically) metaphysical theories – and the problem of their retention and elimination. The 'sophisticated' solution is obvious. We retain a syntactically metaphysical theory as long as the problematic instances can be explained by content-increasing changes in the auxiliary hypotheses appended to it. [119] Let us take, for instance, Cartesian metaphysics C: "in *all* natural processes *there is* a clockwork mechanism regulated by (*a priori*) animating principles". This is syntactically irrefutable: it can clash with no – spatiotemporally singular – 'basic statement'. It may, of course, clash with a refutable theory like N: "gravitation is a force equal to fm_1m_2/r^2 which *acts at a distance*". But N will only clash with C if 'action at a distance' is interpreted literally and possibly, in addition, as representing an *ultimate* truth, irreducible to any still deeper cause. (Popper would call this an 'essentialist' interpretation.) Alternatively we can regard 'action at a distance' as a mediate cause. Then we interpret 'action at a distance' figuratively, and regard it as a shorthand for some hidden mechanism of action by contact. (We may call this a 'nominalist' interpretation.) In this case we can attempt to explain N by C – Newton himself and several French physicists of the eighteenth century tried to do so. If an auxiliary theory which performs this explanation (or, if you wish, 'reduction') produces novel facts (that is, it is 'independently testable'), Cartesian metaphysics should be regarded as good, scientific, empirical metaphysics, generating a progressive problemshift. A progressive (syntactically) metaphysical theory produces a sustained progressive shift in its protective belt of auxiliary theories. If the reduction of the theory to the 'metaphysical' framework does not produce new empirical content, let alone novel facts, then the reduction represents a degenerating problemshift, it is a mere linguistic exercise. The Cartesian efforts to bolster up their 'metaphysics' in order to explain Newtonian gravitation is an outstanding example of such a merely linguistic reduction.[120]

Thus we do not eliminate a (syntactically) metaphysical theory if it clashes with a well-corroborated scientific theory, as naive falsificationism suggests. We eliminate it if it produces a degenerating shift in the long run and there is a better, rival, metaphysics to replace it. The methodology of a research programme with a 'metaphysical' core does not differ from the methodology of one with a 'refutable' core except perhaps for the logical level of the inconsistencies which are the driving force of the programme.

(It has to be stressed, however, that the very choice of the logical form in which to articulate a theory depends to a large extent on our methodological decision. For instance, instead of formulating Cartesian metaphysics as an 'all-some' statement, we can formulate it as an 'all-statement': 'all natural processes are clockworks'. A 'basic statement' contradicting this would be: '*a* is a natural process and it is not clockwork'. The question is whether according to the 'experimental techniques', or rather, to the interpretative theories of the day, '*x* is not a clockwork' can be 'established' or not. Thus the rational choice of the logical form of a theory depends on the state of our knowledge; for instance, a metaphysical 'all-some' statement of today may become, with the change in the level of observational theories, a scientific 'all-statement' tomorrow. I have already argued that only series of theories and not theories should be classified as scientific or non-scientific; now I have indicated that even the logical form of a theory can only be rationally chosen on the basis of a critical appraisal of the state of the research programme in which it is embedded.)

The first, second, and third type decisions of naive falsificationism, however, cannot be avoided, but as we shall show, the conventional element in the second decision – and also in the third – can be slightly reduced. We cannot avoid the decision which sort of propositions should be the 'observational' ones and which the 'theoretical' ones. We cannot avoid either the decision about the truth-value of some 'observational propositions'. These decisions are vital for the decision whether a problemshift is empirically progressive or degenerating. But the sophisticated falsificationist may at least mitigate the arbitrariness of this second decision by allowing for an *appeal procedure.*

Naive falsificationists do not lay down any such appeal procedure. They accept a basic statement if it is backed up by a well-corroborated falsifying hypothesis,[121] and let it overrule the theory under test – even though they are well aware of the risk.[122] But there is no reason why we should not regard a falsifying hypothesis – and the basic statement it supports – as being just as problematic as a falsified hypothesis. Now how exactly can we expose the problematicality of a basic statement? On what grounds can the protagonists of the 'falsified' theory appeal and win?

Some people may say that we might go on testing the basic statement (or the falsifying hypothesis) 'by their deductive consequences' until agree-

ment is finally reached. In this testing we deduce – in the same deductive model – further consequences from the basic statement either with the help of the theory under test or some other theory which we regard as unproblematic. Although this procedure 'has no natural end', we always come to a point when there is no further disagreement.[123]

But when the theoretician appeals against the verdict of the experimentalist, the appeal court does not normally cross-question the basic statement directly but rather questions the *interpretative theory* in the light of which its truth-value had been established.

One typical example of a series of successful appeals is the Proutians' fight against unfavourable experimental evidence from 1815 to 1911. For decades Prout's theory T ("that all atoms are compounds of hydrogen atoms and thus 'atomic weights' of all chemical elements must be expressible as whole numbers") and falsifying 'observational' hypotheses, like Stas's 'refutation' R ("the atomic weight of chlorine is 35·5") confronted each other. As we know, in the end T prevailed over R.[124]

The first stage of any serious criticism of a scientific theory is to reconstruct, improve, its logical deductive articulation. Let us do this in the case of Prout's theory *vis à vis* Stas's refutation. First of all, we have to realize that in the formulation we just quoted, T and R were *not* inconsistent. (Physicists rarely articulate their theories sufficiently to be pinned down and caught by the critic.) In order to show them up as inconsistent we have to put them in the following form. T: "the atomic weight of all pure (homogeneous) chemical elements are multiples of the atomic weight of hydrogen", and R: "chlorine is a pure (homogeneous) chemical element and its atomic weight is 35·5". The last statement is in the form of a falsifying hypothesis which, if well corroborated, would allow us to use basic statements of the form B: "Chlorine X is a pure (homogeneous) chemical element and its atomic weight is 35·5" – where X is the proper name of a 'piece' of chlorine determined, say, by its space-time co-ordinates.

But how well-corroborated is R? Its first component depends on R_1: "Chlorine X is a pure chemical element". This was the verdict of the experimental chemist after a rigorous application of the 'experimental techniques' of the day.

Let us have a closer look at the fine-structure of R_1. In fact R_1 stands for a conjunction of two longer statements T_1 and T_2. The first statement, T_1, could be this: "If seventeen chemical purifying procedures $p_1, p_2 \cdots p_{17}$

are applied to a gas, what remains will be pure chlorine". T_2 is then: "X was subjected to the seventeen procedures $p_1, p_2 \ldots p_{17}$". The careful 'experimenter' carefully applied all seventeen procedures: T_2 is to be accepted. But the conclusion that therefore what remained *must* be pure chlorine is a 'hard fact' only in virtue of T_1. The experimentalist, while *testing T, applied* T_1. He *interpreted* what he saw in the light of T_1: the result was R^1. *Yet in the monotheoretical deductive model of the test situation this interpretative theory does not appear at all.*

But what if T_1, the interpretative theory, is false? Why not 'apply' T rather than T_1 and claim that atomic weights *must be* whole numbers? Then *this* will be a 'hard fact' in the light of T, and T_1 will be overthrown. Perhaps additional new purifying procedures must be invented and applied.

The problem is then *not* when we should stick to a '*theory*' in the face of '*known facts*' and when the other way round. The problem is *not* what to do when 'theories' clash with 'facts'. Such a 'clash' is only suggested by the '*monotheoretical deductive model*'. Whether a proposition is a '*fact*' or a '*theory*' in the context of a test-situation depends on our methodological decision. 'Empirical basis of a theory' is a mono-theoretical notion, it is *relative* to some mono-theoretical deductive structure. We may use it as first approximation; but in case of 'appeal' by the theoretician, we must use a *pluralistic model.* In the pluralistic model the clash is not 'between theories and facts' but between two high-level theories: between an *interpretative theory* to provide the facts and an *explanatory theory* to explain them; and the interpretative theory may be on quite as high a level as the explanatory theory. The clash is then not any more between a logically higher-level theory and a lower-level falsifying hypothesis. The problem should not be put in terms of whether a '*refutation*' is real or not. The problem is how to repair an *inconsistency* between the 'explanatory theory' under test and the – explicit or hidden – 'interpretative' theories; or, if you wish, *the problem is which theory to consider as the interpretative one which provides the 'hard' facts and which the explanatory one which 'tentatively' explains them.* In a mono-theoretical model we regard the higher-level theory as an *explanatory theory to be judged by the 'facts'* delivered from outside (by the authoritative experimentalist): in the case of a clash we reject the explanation.[125] In a pluralistic model we may decide, alternatively, to regard the higher level theory as an *interpretative theory to*

judge the 'facts' delivered from outside: in case of a clash we may reject the 'facts' as 'monsters'. In a pluralistic model of testing, several theories – more or less deductively organized – are soldered together.

This argument alone would be enough to show the correctness of the conclusion, which we drew from a different earlier argument, that experiments do not simply overthrow theories, that no theory forbids a state of affairs specifiable in advance. It is not that we propose a theory and Nature may shout NO; rather we propose a maze of theories, and Nature may shout INCONSISTENT.[126]

The problem is then *shifted* from the old problem of replacing a theory refuted by 'facts' to the new problem of how to resolve inconsistencies between closely associated theories. Which of the mutually inconsistent theories should be eliminated? The sophisticated falsificationist can answer that question easily: one had to try to replace first one, then the other, then possibly both, and opt for that new set-up which provides the biggest increase in corroborated content, which provides the most progressive problemshift.[127]

Thus we have established an appeal procedure in case the theoretician wishes to question the negative verdict of the experimentalist. The theoretician may demand that the experimentalist specify his 'interpretative theory',[128] and he may then replace it – to the experimentalist's annoyance – by a better one in the light of which his originally 'refuted' theory may receive positive appriasal.[129]

But even this appeal procedure cannot do more than *postpone* the conventional decision. For the verdict of the appeal court is not infallible either. When we decide whether it is the replacement of the 'interpretative' or of the 'explanatory' theory that produces novel facts, we again must take a decision about the acceptance or rejection of basic statements. But then we have only *postponed* – and possibly *improved* – the decision, not avoided it.[130] The difficulties concerning the empirical basis which confronted 'naive' falsificationism cannot be avoided by 'sophisticated' falsificationism either. Even if we regard a theory as 'factual', that is, if our slow-moving and limited imagination cannot offer an alternative to it (as Feyerabend used to put it), we have to make, at least occasionally and temporarily, decisions about its truth-value. *Even then, experience still remains, in an important sense, the 'impartial arbiter'*[131] *of scientific controversy.* We cannot get rid of the problem of the 'empirical basis', if we want to

learn from experience[132]: but we can make our learning less dogmatic – but also less fast and less dramatic. By regarding some observational theories as problematic we may make our methodology more flexible: but we cannot articulate and include *all* 'background knowledge' (or 'background ignorance'?) into our critical deductive model. This process is bound to be piecemeal and some conventional line must be drawn at any given time.

There is one objection even to the sophisticated version of methodological falsificationism which cannot be answered without some concession to Duhemian 'simplicism'. The objection is the so-called 'tacking paradox'. According to our definitions, adding to a theory completely disconnected low-level hypotheses may constitute a 'progressive shift'. It is difficult to eliminate such makeshift shifts without demanding that the additional assertions must be connected with the original assertion *more intimately* than by mere conjunction. This, of course, is a sort of simplicity requirement which would assure the continuity in the series of theories which can be said to constitute *one* problemshift.

This leads us to further problems. For one of the crucial features of sophisticated falsificationism is that it replaces the concept of *theory* as the basic concept of the logic of discovery by the concept of *series of theories. It is a succession of theories and not one given theory which is appraised as scientific or pseudo-scientific.* But the members of such series of theories are usually connected by a remarkable *continuity* which welds them into *research programmes.* This *continuity* – reminiscent of Kuhnian 'normal science' – plays a vital role in the history of science; the main problems of the logic of discovery cannot be satisfactorily discussed except in the framework of a *methodology of research programmes.*

3. A METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES

I have discussed the problem of objective appraisal of scientific growth in terms of progressive and degenerating problemshifts in series of scientific theories. The most important such series in the growth of science are characterized by a certain *continuity* which connects their members. This continuity evolves from a genuine research programme adumbrated at the start. The programme consists of methodological rules: some tell us

what paths of research to avoid (*negative heuristic*), and others what paths to pursue (*positive heuristic*).[133]

Even science as a whole can be regarded as a huge research programme with Popper's supreme heuristic rule: "devise conjectures which have more empirical content than their predecessors." Such methodological rules may be formulated, as Popper pointed out, as metaphysical principles.[134] For instance, the *universal* anti-conventionalist rule against exception-barring may be stated as the metaphysical principle: 'Nature does not allow exceptions'. This is why Watkins called such rules 'influential metaphysics'.[135]

But what I have primarily in mind is not science as a whole, but rather *particular* research programmes, such as the one known as 'Cartesian metaphysics'. Cartesian metaphysics, that is, the mechanistic theory of the universe – according to which the universe is a huge clockwork (and system of vortices) with push as the only cause of motion – functioned as a powerful heuristic principle. It discouraged work on scientific theories – like [the 'essentialist' version of] Newton's theory of action at a distance – which were inconsistent with it (*negative heuristic*). On the other hand, it encouraged work on auxiliary hypotheses which might have saved it from apparent counterevidence – like Keplerian ellipses (*positive heuristic*).[136]

(a) *Negative Heuristic: The 'Hard Core' of the Programme.*

All scientific research programmes may be characterized by their '*hard core*'. The negative heuristic of the programme forbids us to direct the *modus tollens* at this 'hard core'. Instead, we must use our ingenuity to articulate or even invent 'auxiliary hypotheses', which form a *protective belt* around this core, and we must redirect the *modus tollens* to *these*. It is this protective belt of auxiliary hypotheses which has to bear the brunt of tests and get adjusted and re-adjusted, or even completely replaced, to defend the thus-hardened core. A research programme is successful if all this leads to a progressive problemshift; unsuccessful if it leads to a degenerating problemshift.

The classical example of a successful research programme is Newton's gravitational theory; possibly the most successful research programme ever. When it was first produced, it was submerged in an ocean of 'anomalies' (or, if you wish, 'counterexamples'), and opposed by the observational theories supporting these anomalies. But Newtonians turned, with

brilliant tenacity and ingenuity, one counter-instance after another into corroborating instances, primarily by overthrowing the original observational theories in the light of which this 'contrary evidence' was established. In the process they themselves produced new counter-examples which they again resolved. They "turned each new difficulty into a new victory of their programme".[137]

In Newton's programme the negative heuristic bids us to divert the *modus tollens* from Newton's three laws of dynamics and his law of gravitation. This 'core' is 'irrefutable' by the methodological decision of its protagonists: anomalies must lead to changes only in the 'protective' belt of auxiliary, 'observational' hypotheses and initial conditions.[138]

I have given a contrived micro-example of a progressive Newtonian problemshift.[139] If we analyse it, it turns out that each successive link in this exercise predicts some new fact; each step represents an increase in empirical content: the example constitutes a *consistently progressive theoretical shift*. Also, each prediction is in the end verified; although on three subsequent occasions they may have seemed momentarily to be 'refuted'.[140] While 'theoretical progress' (in the sense here described) may be verified immediately, 'empirical progress' cannot, and in a research programme we may be frustrated by a long series of 'refutations' before ingenious and lucky content-increasing auxiliary hypotheses turn a chain of defeats – *with hindsight* – into a resounding success story, either by revising some false 'facts' or by adding novel auxiliary hypotheses. We may then say that we must require that each step of a research programme be consistently content-increasing: that each step constitute a *consistently progressive theoretical problemshift*. All we need in addition to this is that at least every now and then the increase in content should be seen to be retrospectively corroborated: the programme as a whole should also display an *intermittently progressive empirical shift*. We do not demand that each step produce *immediately* an *observed* new fact. Our term '*intermittently*' gives sufficient *rational* scope for dogmatic adherence to a programme in face of *prima facie* 'refutations'.

The idea of 'negative heuristic' of a scientific research programme rationalizes classical conventionalism to a considerable extent. We may rationally decide not to allow 'refutations' to transmit falsity to the hard core as long as the corroborated empirical content of the protecting belt of auxiliary hypotheses increases. But our approach differs from Poinca-

ré's justificationist conventionalism in the sense that, unlike Poincaré's, we maintain that if and when the programme ceases to anticipate novel facts, its hard core might have to be abandoned: that is, *our* hard core, unlike Poincaré's, may crumble under certain conditions. In this sense we side with Duhem who thought that such a possibility must be allowed for; but for Duhem the reason for such crumbling is purely *aesthetic*, while for us it is mainly *logical and empirical.*

(b) *Positive Heuristic: The Construction of the 'Protective Belt' and the Relative Autonomy of Theoretical Science.*

Research programmes, besides their negative heuristic, are also characterized by their positive heuristic.

Even the most rapidly and consistently progressive research programmes can digest their 'counter-evidence' only piecemeal: anomalies are never completely exhausted. But it should not be thought that yet unexplained anomalies – 'puzzles' as Kuhn might call them – are taken in random order, and the protective belt built up in an eclectic fashion, without any preconceived order. The order is usually decided in the theoretician's cabinet, independently of the *known* anomalies. Few theoretical scientists engaged in a research programme pay undue attention to 'refutations'. They have a long-term research policy which anticipates these refutations. This research policy, or order of research, is set out – in more or less detail – in the *positive heuristic* of the research programme. The negative heuristic specifies the 'hard core' of the programme which is 'irrefutable' by the methodological decision of its protagonists; the positive heuristic consists of a partially articulated set of suggestions or hints on how to change, develop the 'refutable variants' of the research-programme, how to modify, sophisticate, the 'refutable' protective belt.

The positive heuristic of the programme saves the scientist from becoming confused by the ocean of anomalies. The positive heuristic sets out a programme which lists a chain of ever more complicated *models* simulating reality: the scientist's attention is riveted on building his models following instructions which are laid down in the positive part of his programme. He ignores the *actual* counterexamples, the available '*data*'.[141] Newton first worked out his programme for a planetary system with a fixed point-like sun and one single point-like planet. It was in this

model that he derived his inverse square law for Kepler's ellipse. But this model was forbidden by Newton's own third law of dynamics, therefore the model had to be replaced by one in which both sun and planet revolved round their common centre of gravity. This change was not motivated by any observation (the data did not suggest an 'anomaly' here) but by a theoretical difficulty in developing the programme. Then he worked out the programme for more planets as if there were only heliocentric but no interplanetary forces. Then he worked out the case where the sun and planets were not mass-points but mass-*balls*. Again, for this change he did not *need* the observation of an anomaly; infinite density was forbidden by an (inarticulated) touchstone theory, therefore planets *had* to be extended. This change involved considerable mathematical difficulties, held up Newton's work – and delayed the publication of the *Principia* by more than a decade. Having solved this 'puzzle', he started work on *spinning balls* and their wobbles. Then he admitted interplanetary forces and started work on *perturbations*. At this point he started to look more anxiously at the facts. Many of them were beautifully explained (qualitatively) by this model, many were not. It was then that he started to work on *bulging* planets, rather than round planets, etc.

Newton despised people who, like Hooke, stumbled on a first naive model but did not have the tenacity and ability to develop it into a research programme, and who thought that a first version, a mere aside, constituted a 'discovery'. He held up publication until his programme had achieved a remarkable progressive shift.[142]

Most, if not all, Newtonian 'puzzles', leading to a series of new variants superseding each other, were foreseeable at the time of Newton's first naive model and no doubt Newton and his colleagues *did* forsee them: Newton must have been fully aware of the blatant falsity of his first variants.[143] Nothing shows the existence of a positive heuristic of a research programme clearer than this fact: this is why one speaks of 'models' in research programmes. A '*model*' is a set of initial conditions (possibly together with some of the observational theories) which one knows is *bound* to be replaced during the further development of the programme, and one even knows, more or less, how. This shows once more how irrelevant 'refutations' of any specific variant are in a research programme: their existence is fully expected, the positive heuristic is there as the strategy both for predicting (producing) and digesting them. Indeed, if

the positive heuristic is clearly spelt out, the difficulties of the programme are mathematical rather than empirical.[144]

One may formulate the 'positive heuristic' of a research programme as a 'metaphysical' principle. For instance one may formulate Newton's programme like this :"the planets are essentially gravitating spinning-tops of roughly spherical shape". This idea was never *rigidly* maintained: the planets are not *just* gravitational, they have also, for example, electromagnetic characteristics which may influence their motion. Positive heuristic is thus in general more flexible than negative heuristic. Moreover, it occasionally happens that when a research programme gets into a degenerating phase, a little revolution or a *creative shift* in its positive heuristic may push it forward again.[145] It is better therefore to separate the 'hard core' from the more flexible metaphysical principles expressing the positive heuristic.

Our considerations show that the positive heuristic forges ahead with almost complete disregard of 'refutations': it may seem that it is the '*verifications*'[146] rather than the refutations which provide the contact points with reality. Although one must point out that any 'verification' of the $n+1$th version of the programme is a refutation of the n'th version, we cannot deny that *some* defeats of the subsequent versions are always foreseen: it is the 'verifications' which keep the programme going, recalcitrant instances notwithstanding.

We may appraise research programmes, even after their 'elimination', for their *heuristic power*: how many new facts did they produce, how great was "their capacity to explain their refutations in the course of their growth"?[147]

(We may also appraise them for the stimulus they gave to mathematics. The real difficulties for the theoretical scientist arise rather from the *mathematical difficulties* of the programme than from anomalies. The greatness of the Newtonian programme comes partly from the development – by Newtonians – of classical infinitesimal analysis which was a crucial precondition of its success.)

Thus the methodology of scientific research programmes accounts for the *relative autonomy of theoretical science*: a historical fact whose rationality cannot be explained by the earlier falsificationists. Which problems scientists working in powerful research programmes rationally choose, is determined by the positive heuristic of the programme rather than by

psychologically worrying (or technologically urgent) anomalies. The anomalies are listed but shoved aside in the hope that they will turn, in due course, into corroborations of the programme. Only those scientists have to rivet their attention on anomalies who are either engaged in trial-and-error exercises or who work in a degenerating phase of a research programme when the positive heuristic ran out of steam. (All this, of course, must sound repugnant to naive falsificationists who hold that once a theory is 'refuted' by experiment (by *their* rule book), it is irrational (and dishonest) to develop it further: one has to replace the old 'refuted' theory by a new, unrefuted one.)

NOTES

* From *Criticism and the Growth of Knowledge* (pp. 91–138), ed. by I. Lakatos and A. Musgrave. Copyright © 1970. Reprinted by permission of Cambridge University Press.

This paper is a considerably improved version of my (1968b) and a crude version of my (1970). Some parts of the former are here reproduced without change with the permission of the Editor of the *Proceedings of the Aristotelian Society*. In the preparation of the new version I received much help from Tad Beckman, Colin Howson, Clive Kilmister, Larry Laudan, Eliot Leader, Alan Musgrave, Michael Sukale, John Watkins and John Worrall.

[1] The main contemporary protagonist of the ideal of 'probable truth' is Rudolf Carnap. For the historical background and a criticism of this position, cf. Lakatos (1968a).

[2] The main contemporary protagonists of the ideal of 'truth by consensus' are Polanyi and Kuhn. For the historical background and a criticism of this position, cf. Musgrave (1969a), Musgrave (1969b) and Lakatos (1970).

[3] Indeed he introduces his (1962) by arguing against the 'development-by-accumulation' idea of scientific growth. But his intellectual debt is to Koyré rather than to Popper. Koyré showed that positivism gives bad guidance to the historian of science, for the history of physics can only be understood in the context of a succession of 'metaphysical' research programmes. Thus scientific changes are connected with vast cataclysmic metaphysical revolutions. Kuhn develops this message of Burtt and Koyré and the vast success of his book was partly due to his hard-hitting, direct criticism of justificationist historiography – which created a sensation among ordinary scientists and historians of science whom Burtt's, Koyré's (or Popper's) message has not yet reached. But, unfortunately, his message had some authoritarian and irrationalist overtones.

[4] Cf. e.g. Watkins's and Feyerabend's contributions to *Criticism and the Growth of Knowledge*.

[5] Justificationists repeatedly stressed this asymmetry between singular factual statements and universal theories. Cf. e.g. Popkin's discussion of Pascal in Popkin (1968), p. 14 and Kant's statement to the same effect as quoted in the new *motto* of the third 1969 German edition of Popper's *Logik der Forschung*. (Popper's choice of this time-honoured cornerstone of elementary logic as a *motto* of the new edition of his classic shows his main concern: to fight *probabilism*, in which this asymmetry becomes irrelevant; for probabilists theories may become almost as well established as factual propositions.)

[6] Indeed, even some of these few shifted, following Mill, the rather obviously insoluble problem of inductive proof (of universal from particular propositions) to the slightly less obviously insoluble problem of proving *particular* factual propositions from other *particular* factual propositions.

[7] The founding fathers of probabilism were intellectualists; Carnap's later efforts to build up an empiricist brand of probabilism failed. Cf. my (1968a), p. 367 and also p. 361, footnote 2.

[8] For a detailed discussion, cf. my (1968a), especially pp. 353ff.

[9] Russell (1943), p. 683. For a discussion of Russell's justificationism, cf. my (1962), especially pp. 167 ff.

[10] Medawar (1967), p. 144.

[11] This discussion already indicates the vital importance of a demarcation between provable factual and unprovable theoretical propositions for the dogmatic falsificationist.

[12] "*Criteria of refutation* have to be laid down beforehand: it must be agreed which observable situations, if actually observed, mean that the theory is refuted" (Popper [1963], p. 38, Footnote 3).

[13] Quoted in Popper (1934), Section 85, with Popper's comment: "I fully agree".

[14] Braithwaite (1953), pp. 367–8. For the 'incorrigibility' of Braithwaite's observed facts, cf. his (1938). While in the quoted passage Braithwaite gives a forceful answer to the problem of scientific objectivity, in another passage he points out that "except for the straightforward generalizations of observable facts... complete refutation is no more possible than is complete proof" ([1953], p. 19).

[15] For these assumptions and their criticism, cf. Popper (1934), Sections 4 and 10. It is because of this assumption that – following Popper – I call this brand of falsificationism 'naturalistic'. Popper's 'basic propositions' should not be confused with the basic propositions discussed in this section; cf. *below*, note 42.

It is important to point out that these two assumptions are also shared by many justificationists who are not falsificationists: they may add to experimental proofs 'intuitive proofs' – as did Kant – or 'inductive proofs' – as did Mill. Our falsificationist accepts experimental proofs *only*.

[16] The empirical basis of a theory is the set of its potential falsifiers: the set of those observational propositions which may disprove it.

[17] Incidentally, Galileo also showed – with the help of his optics – that if the moon was a faultless crystal ball, it would be invisible (Galileo [1632]).

[18] True, most psychologists who turned against the idea of justificationist sensationalism did so under the influence of pragmatist philosophers like William James who denied the possibility of any sort of objective knowledge. But, even so, Kant's influence through Oswald Külpe, Franz Brentano and Popper's influence through Egon Brunswick and Donald Campbell played a role in the shaping of modern psychology; and if psychology ever vanquishes psychologism, it will be due to an increased understanding of the Kant-Popper mainline of objectivist philosophy.

[19] Cf. Popper (1934), Section 29.

[20] It seems that the first philosopher to emphasize this might have been Fries in 1837 (cf. Popper [1934], Section 29, Footnote 3). This is of course a special case of the general thesis that logical relations, like probability or consistency, refer to *propositions*. Thus, for instance, the proposition 'nature is consistent' is false (or, if you wish, meaningless), for nature is not a proposition (or a conjunction of propositions).

[21] Incidentally, even this is questionable.

[22] As Popper put it: "No conclusive disproof of a theory can ever be produced", those who wait for an infallible disproof before eliminating a theory will have to wait for ever and "will never benefit from experience" ([1934], Section 9).
[23] Kant and his English follower, Whewell, both realized that all scientific propositions, whether *a priori* or *a posteriori*, are equally theoretical; but both held that they are equally provable. Kantians saw clearly that the propositions of science are theoretical in the sense that they are not written by sensations on the *tabula rasa* of an empty mind, nor deduced or induced from such propositions. A factual proposition is only a special kind of theoretical proposition. In this Popper sided with Kant against the empiricist version of dogmatism. But Popper went a step further: in his view the propositions of science are not only theoretical but they are all also *fallible*, conjectural for ever.
[24] If the tiny conjectural planet were out of the reach even of the biggest *possible* optical telescopes, he might try some quite novel instrument (like a radiotelescope) in order to enable him to 'observe it', that is, to ask Nature about it, even if only indirectly. (The new 'observational' theory may itself not be properly articulated, let alone severely tested, but he would care no more than Galileo did.)
[25] At least not until a new research programme supersedes Newton's programme which happens to explain this previously recalcitrant phenomenon. In this case, the phenomenon will be unearthed and enthroned as a 'crucial experiment'.
[26] Popper asks: "What kind of clinical responses would refute to the satisfaction of the analyst not merely a particular diagnosis but psychoanalysis itself?" ([1963], p. 38, Footnote 3.) But what kind of observation would refute to the satisfaction of the Newtonian not merely a particular version but Newtonian theory itself?
[27] (*Added in press*:) This '*ceteris paribus*' clause must not normally be interpreted as a separate premise.
[28] Incidentally, we might persuade the dogmatic falsificationist that his demarcation criterion was a very naive mistake. If he gives it up but retains his two basic assumptions, he will have to ban theories from science and regard the growth of science as an accumulation of proven basic statements. This indeed is the final stage of classical empiricism after the evaporation of the hope that facts can prove or at least disprove theories.
[29] Cf. Popper (1934), Chapter VIII.
[30] For a *much* stronger case, cf. *below*, Sect. 3.
[31] This demarcation – and terminology – is due to Popper; cf. especially his (1934), Section 19 and his (1945), Chapter 23 and Footnote 3 to Chapter 25.
[32] No version of conservative activism explained why Newton's *gravitational* theory should be invulnerable; Kantians restricted themselves to the explanation of the tenacity of Euclidean geometry and Newtonian *mechanics*. About Newtonian *gravitation* and *optics* (or other branches of science) they had an ambiguous, and occasionally inductivist position.
[33] I do not include Hegel among 'revolutionary *activists*'. For Hegel and his followers change in conceptual frameworks is a predetermined, inevitable process, where individual creativity or rational criticism plays no essential role. Those who run ahead are equally at fault as those who stay behind in this 'dialectic'. The clever man is not he who creates a better 'prison' or who demolishes critically the old one, but the one who is always in step with history. Thus dialectic accounts for change without criticism.
[34] Cf. Whewell's (1837), (1840) and (1858).
[35] Cf. especially Poincaré (1891) and (1902); Milhaud (1896); Le Roy (1899) and (1901). It was one of the chief philosophical merits of conventionalists to direct the limelight to

the fact that any theory can be saved by 'conventionalist stratagems' from refutations. (The term 'conventionalist stratagem' is Popper's; cf. the critical discussion of Poincaré's conventionalism in his [1934], especially Sections 19 and 20.)

[36] Poincaré first elaborated his conventionalism only with regard to geometry (cf. his [1891]). Then Milhaud and Le Roy generalized Poincaré's idea to cover all branches of accepted physical theory. Poincaré's (1902) starts with a strong criticism of the Bergsonian Le Roy against whom he defends the empirical (falsifiable or 'inductive') character of all physics *except for* geometry and mechanics. Duhem, in turn, criticized Poincaré: in his view there was a possibility of overthrowing even Newtonian mechanics.

[37] The *loci classici* are Duhem's (1905) and Popper's (1934). Duhem was not a *consistent* revolutionary conventionalist. Very much like Whewell, he thought that conceptual changes are only *preliminaries* to the final – if perhaps distant – 'natural classification': "The more a theory is perfected, the more we apprehend that the logical order in which it arranges experimental laws is the reflection of an ontological order". In particular, he refused to see Newton's mechanics *actually* 'crumbling' and characterized Einstein's relativity theory as the manifestation of a "frantic and hectic race in pursuit of a novel idea" which "has turned physics into a real chaos where logic loses its way and commonsense runs away frightened" (Preface – of 1914 – to the second edition of his [1905]).

[38] Duhem (1905), Chapter VI, Section 10.

[39] Popper (1934), Section 30.

[40] *In this section I discuss the 'naive' variant of Popper's methodological falsificationism. Thus, throughout the section 'methodological falsificationism' stands for 'naive methodological falsificationism'.*

[41] Popper (1934), Section 27.

[42] *Op. cit.* Section 28. For the non-basicness of these methodologically 'basic' statements, cf. e.g. Popper (1934) *passim* and Popper (1959a), p. 35, Footnote *2.

[43] Cf. Popper (1934), end of Section 26 and also his (1968c), pp. 291–2.

[44] Cf. Popper (1963), p. 390.

[45] Indeed, Popper carefully puts 'observational' in quotes; cf. his (1934), Section 28.

[46] This demarcation plays a role both in the *first* and in the *fourth* type of decisions of the methodological falsificationist.

[47] For a fascinating discussion, cf. Feyerabend (1969).

[48] One wonders whether it would not be better to make a break with the terminology of naturalistic falsificationism and rechristen observational theories '*touchstone theories*'.

[49] Cf. Popper (1934), Section 22. Many philosophers overlooked Popper's important qualification that a basic-statement has no power to refute anything without the support of a well-corroborated falsifying hypothesis.

[50] Cf. Popper (1934), Section 30.

[51] Popper (1963), p. 387.

[52] Popper (1934), Section 30; also cf. Section 29: 'The Relativity of Basic Statements'.

[53] Popper (1957), p. 134. Popper, in other places, emphasizes that his method cannot 'ensure' the survival of the fittest. Natural selection may go wrong: the fittest may perish and monsters survive.

[54] Popper (1935).

[55] Popper (1934), Section 82.

[56] Popper (1934), Section 82.

[57] This kind of methodological 'falsification' is, unlike dogmatic falsification (disproof), a pragmatic, methodological idea. But then what exactly should we mean by it?

Popper's answer – which I am going to discard – is that methodological 'falsification' indicates an "urgent need of replacing a falsified hypothesis by a better one" (Popper [1959a], p. 87, Footnote *1). This shift is an excellent illustration of the process I described in my (1963–4) whereby critical discussion shifts the original *problem* without necessarily changing the old *terms*. The byproducts of such processes are *meaning-shifts*.

[58] The demarcation criterion of the dogmatic falsificationist was: a theory is 'scientific' if it has an empirical basis.

[59] Incidentally, Popper, in his (1934), does not seem to have seen this point clearly. He writes: "Admittedly, it is possible to interpret the concept of an *observable event* in a psychologistic sense. But I am using it in such a sense that it might just as well be replaced by 'an event involving position and movement of macroscopic physical bodies'". ([1934], Section 28.) In the light of our discussion, for instance, we may regard a positron passing through a Wilson chamber at time t_0 as an 'observable' event, in spite of the non-macroscopic character of the positron.

[60] Popper (1934), Section 68. Indeed, this methodological falsificationism is the philosophical basis of some of the most interesting developments in modern statistics. The Neyman-Pearson approach rests completely on methodological falsificationism. Also cf. Braithwaite (1953), Chapter VI. (Unfortunately, Braithwaite reinterprets Popper's demarcation criterion as separating meaningful from meaningless rather than scientific from non-scientific propositions.)

[61] Popper (1933).

[62] Popper (1933).

[63] Popper (1957), p. 133.

[64] For a discussion of this important concept of Popperian methodology, cf. my (1968a), pp. 397 ff.

[65] Popper (1934), Section 9.

[66] *Ibid.*

[67] *Ibid.*

[68] The problem of '*controlled experiment*' may be said to be nothing else but the problem of arranging experimental conditions in such a way as to minimize the risk involved in such decisions.

[69] This type of decision belongs, in an important sense, to the same category as the first decision: it demarcates, by decision, problematic from unproblematic knowledge.

[70] Our exposition shows clearly the complexity of the decisions needed to define the 'empirical content' of a theory – that is, the set of its potential falsifiers. 'Empirical content' depends on our *decision* as to which are our 'observational theories' and which anomalies are to be promoted to counterexamples. If one attempts to compare the empirical content of different scientific theories in order to see which is 'more scientific', then one will get involved in an enormously complex and therefore hopelessly arbitrary system of decisions about their respective classes of 'relatively atomic statements' and their 'fields of application'. (For the meaning of these (very) technical terms, cf. Popper [1934], Section 38.) But such comparison is possible only when one theory supersedes another (cf. Popper, [1959a], p. 401, Footnote 7). And even then, there may be difficulties (which would not, however, add up to irremediable 'incommensurability').

[71] This was suggested by J. O. Wisdom: cf. his (1963).

[72] For instance: 'All metals have a solvent'; or 'There exists a substance which can turn all metals into gold'. For discussions of such theories, cf. especially Watkins (1957) and Watkins (1960).

[73] Russell (1943), p. 683.

[74] I am sure that some will welcome methodological falsificationism as an 'existentialist' philosophy of science.
[75] Neurath (1935), p. 356.
[76] Hempel (1952), p. 621. Agassi, in his (1966), follows Neurath and Hempel, especially pp. 16ff. It is rather amusing that Agassi, in making this point, thinks that he is taking up arms against "the whole literature concerning the methods of science".

Indeed, many scientists were fully aware of the difficulties inherent in the "confrontation of theory and facts". (Cf. Einstein [1949], p. 27.) Several philosophers sympathetic to falsificationism emphasized that "the process of refuting a scientific hypothesis is more complicated than it appears to be at first sight" (Braithwaite [1953], p. 20). But only Popper offered a constructive, rational solution.
[77] Hempel (1952), p. 622. Hempel's crisp "theses on empirical certainty" do nothing but refurbish Neurath's – and some of Popper's – old arguments (against Carnap, I take it); but deplorably, he does not mention either his predecessors or his adversaries.
[78] Neurath (1935).
[79] Popper (1934), Section 26.
[80] Neurath's (1935) shows that he never grasped Popper's simple argument.
[81] I am using here 'verisimilitude' in Popper's sense: the difference between the truth content and falsity content of a theory. For the risks involved in estimating it, cf. my (1968a), especially pp. 395 ff.
[82] I tried to develop such a general theory of criticism in my (1970).
[83] The falsification of theories depends on the high degree of corroboration of the *ceteris paribus* clause. This however is not always the case. This is why the methodological falsificationist may advise us to rely on our 'scientific instinct' (Popper [1934], Section 18, Footnote 2) or 'hunch' (Braithwaite [1953], p. 20).
[84] Agassi (1959); he calls Popper's idea of science '*scientia negativa*' (Agassi [1968]).
[85] It should be mentioned here that the Kuhnian sceptic is still left with what I would call the '*scientific sceptic's dilemma*': any scientific sceptic will still try to explain changes in beliefs and will regard his own psychology as a theory which is more than simple belief, which, in some sense, is 'scientific'. Hume, while trying to show up science as a mere system of beliefs with the help of his stimulus-response theory of learning, never raised the problem of whether his theory of learning applies also to his own theory of learning. In contemporary terms, we might well ask, does the popularity of Kuhn's philosophy indicate that people recognize its *truth*? In this case it would be refuted. Or does this popularity indicate that people regarded it as an attractive new fashion? In this case, it would be 'verified'. But would Kuhn like *this* 'verification'?
[86] Feyerabend who contributed probably more than anybody else to the spread of Popper's ideas, seems now to have joined the enemy camp. Cf. his intriguing (1970).
[87] I use 'prediction' in a wide sense that includes 'postdiction'.
[88] *For a detailed discussion of these acceptance and rejection rules and for references to Popper's work*, cf. my (1968a), pp. 375–90.
[89] Molière, for instance, ridiculed the doctors of his *Malade Imaginaire*, who offered the *virtus dormitiva* of opium as the answer to the question as to why opium produced sleep. One might even argue that Newton's famous dictum *hypotheses non fingo* was really directed against *ad hoc* explanations – like his own explanation of gravitational forces by an aether-model in order to meet Cartesian objections.
[90] Incidentally, Duhem agreed with Bernard that experiments alone – without simplicity considerations – can decide the fate of theories in physiology. But in physics, he argued, they cannot ([1905], Chapter VI, Section 1).

[91] Koestler correctly points out that only Galileo created the myth that the Copernican theory was simple (Koestler [1959], p. 476); in fact, "the motion of the earth [had not] done much to simplify the old theories, for though the objectionable equants had disappeared, the system was still bristling with auxiliary circles" (Dreyer [1906], Chapter XIII).

[92] Popper (1934), Sections 19 and 20. I have discussed in some detail – under the heads 'monster-barring', 'exception-barring', 'monster-adjustment' – such stratagems as they appear in informal, quasi-empirical mathematics; cf. my (1963–4).

[93] If I already know P_1: 'Swan A is white', P_ω: 'All swans are white' represents no progress, because it may only lead to the discovery of such further similar facts as P_2: 'Swan B is white'. So-called 'empirical generalizations' constitute no progress. A *new* fact must be improbable or even impossible in the light of previous knowledge.

[94] The appropriateness of the term 'problemshift' for a series of theories rather than of problems may be questioned. I chose it partly because I have not found a more appropriate alternative – 'theoryshift' sounds dreadful – partly because theories are always problematical, they never solve all the problems they have set out to solve. Anyway, in the second half of the paper, the more natural term 'research programme' will replace 'problemshifts' in the most relevant contexts.

[95] Indeed, in the original manuscript of my (1968a) I wrote : "A theory without excess corroboration has no excess explanatory power; *therefore, according to Popper, it does not represent growth and therefore it is not 'scientific'; therefore, we should say, it has no explanatory power*" (p. 386). I cut out the italicized half of the sentence under pressure from my colleagues who thought it sounded too eccentric. I regret it now.

[96] Popper's conflation of 'theories' and 'series of theories' prevented him from getting the basic ideas of sophisticated falsificationism across more successfully. His ambiguous usage led to such confusing formulations as "Marxism [as the core of a series of theories or of a 'research programme'] is irrefutable" and, at the same time, "Marxism [as a particular conjunction of this core and some specified auxiliary hypotheses, initial conditions and a *ceteris paribus* clause] has been refuted." (Cf. Popper [1963].)

Of course, there is nothing wrong in saying that an isolated, single theory is 'scientific' if it represents an advance on its predecessor, as long as one clearly realizes that in this formulation we appraise the theory as the outcome of – and in the context of – a certain historical development.

[97] Popper (1945), Vol. II, p. 233. Popper's more sophisticated attitude surfaces in the remark that "concrete and practical consequences can be *more* directly tested by experiment" (*ibid.*, my italics).

[98] Popper (1934), Section 30.

[99] "In most cases we have, before falsifying a hypothesis, another one up our sleeves" (Popper [1959a], p. 87, Footnote *1). But, as our argument shows, we *must* have one. Or, as Feyerabend put it: "The best criticism is provided by those theories which can replace the rivals they have removed" ([1965], p. 227). He notes that in *some* cases "alternatives will be quite indispensable for the purpose of refutation" (*ibid.* p. 254). But according to our argument *refutation without an alternative shows nothing but the poverty of our imagination in providing a rescue hypothesis.*

[100] Cf. my (1968a), pp. 387ff.

[101] In the distorting mirror of naive falsificationism, new theories which replace old refuted ones, are themselves born unrefuted. Therefore they do not believe that there is a relevant difference between anomalies and crucial counterevidence. For them, anoma-

ly is a dishonest euphemism for counterevidence. But in actual history new theories are born refuted: they inherit many anomalies of the old theory. Moreover, frequently it is *only* the new theory which dramatically predicts that fact which will function as crucial counterevidence against its predecessor, while the 'old' anomalies may well stay on as 'new' anomalies.

All this will be still clearer when we introduce the idea of 'research programme'.

[102] *Sophisticated falsificationism adumbrates a new theory of learning.*

[103] It is clear that the theory T' may have excess corroborated empirical content over another theory T even if both T and T' are refuted. Empirical content has nothing to do with truth or falsity. Corroborated contents can also be compared irrespective of the refuted content. Thus we may see the rationality of the elimination of Newton's theory in favour of Einstein's, even though Einstein's theory may be said to have been born – like Newton's – 'refuted'. We have only to remember that 'qualitative confirmation' is a euphemism for 'quantitative disconfirmation'. (Cf. my [1968a], pp. 384–6.)

[104] Cf. Popper (1934), Section 85, p. 279 of the 1959 English translation.

[105] It is true that a certain type of *proliferation of rival theories* is allowed to play an accidental heuristic role in falsification. In many cases falsification heuristically "depends on [the condition] that sufficiently many and sufficiently different theories are offered" (Popper [1940]). For instance, we may have a theory T which is apparently unrefuted. But it may happen that a new theory T', inconsistent with T, is proposed which equally fits the available facts: the differences are smaller than the range of observational error. In such cases the inconsistency prods us into improving our 'experimental techniques', and thus refining the 'empirical basis' so that either T or T' (or, incidentally, both) can be falsified: "We need [a] new theory in order to find out where the old theory was deficient" (Popper [1963], p. 246). But the role of this proliferation is *accidental* in the sense that, once the empirical basis is refined, the fight is between this refined empirical basis and the theory T under test; the rival theory T' acted only as a *catalyst.*

[106] Also cf. Feyerabend (1965), pp. 254–5.

[107] Popper (1959a), p. 87, Footnote *1.

[108] Popper (1934), Section 30.

[109] [*Added in press:*] Possibly it would be better in future to abandon these terms altogether, just as we have abandoned terms like 'inductive (or experimental) proof'. Then we may call (naive) 'refutations' anomalies, and (sophisticatedly) 'falsified' theories 'superseded' ones. Our 'ordinary' language is impregnated not only by 'inductivist' but also by falsificationist dogmatism. A reform is overdue.

[110] For a defence of this theory of 'learning from experience', cf. Agassi (1969).

[111] *These remarks show that 'learning from experience' is a normative idea; therefore all purely 'empirical' learning theories miss the heart of the problem.*

[112] Cf. Leibnitz (1678). The expression in brackets shows that Leibnitz regarded this criterion as second best and thought that the best theories are those which are proved. Thus Leibnitz's position – like Whewell's – is a far cry from fully fledged sophisticated falsificationism.

[113] Mill (1843), Vol. II, p. 23.

[114] This was J. S. Mill's argument (*ibid.*). He directed it against Whewell, who thought that 'consilience of inductions' or successful prediction of improbable events *verifies* (that is, *proves*) a theory. (Whewell [1858], pp. 95–6.) No doubt, *the basic mistake both in Whewell's and in Duhem's philosophy of science is their conflation of predictive power and proven truth. Popper separated the two.*

[115] Keynes (1921), p. 305. But cf. my (1968a), p. 394.

[116] This is Whewell's critical comment on an *ad hoc* auxiliary hypothesis in Newton's theory of light (Whewell [1857], Vol. II, p. 317).

[117] In the terminology of my (1968a), this theory was '*ad hoc*$_1$' (cf. my [1968a], p. 389, Footnote 1); the example was originally suggested to me by Paul Feyerabend as a paradigm of a *valuable ad hoc theory*.

[118] In the terminology of my (1968a), this theory was not '*ad hoc*$_1$', but it was '*ad hoc*$_2$' (cf. my [1968a], p. 389, Footnote 1). For a simple but artificial illustration, see *ibid.* p. 387, Footnote 2.

[119] *We can formulate this condition with striking clarity only in terms of the methodology of research programmes to be explained in § 3: we retain a syntactically metaphysical theory as the 'hard core' of a research programme as long as its associated positive heuristic produces a progressive problemshift in the 'protective belt' of auxiliary hypotheses.*

[120] This phenomenon was described in a beautiful paper by Whewell (1851); but he could not explain it methodologically. Instead of recognizing the victory of the *progressive* Newtonian programme over the *degenerating* Cartesian programme, he thought this was the victory of proven truth over falsity. For details cf. my (1973): for a general discussion of the demarcation between progressive and degenerating reduction cf. Popper (1969).

[121] Popper (1934), Section 22.

[122] Cf. e.g. Popper (1959a), p. 107, Footnote *2.

[123] This is argued in Popper (1934), Section 29.

[124] Agassi claims that this example shows that we may "stick to the hypothesis in the face of known facts in the hope that the facts will adjust themselves to theory rather than the other way round" ([1966]. p. 18) .But *how* can facts 'adjust themselves'? Under which *particular* conditions should the theory win? Agassi gives no answer.

[125] The decision to use some monotheoretical model is clearly vital for the naive falsificationist to enable him to reject a theory on the *sole* ground of experimental evidence. *It is in line with the necessity for him to divide sharply, at least in a test-situation, the body of science into two: the problematic and the unproblematic. It is only the theory he decides to regard as problematic which he articulates in his deductive model of criticism.*

[126] Let me here answer a possible objection: "Surely we do not need Nature to tell us that a set of theories is *inconsistent*. Inconsistency – unlike falsehood – can be ascertained without Nature's help". But Nature's actual 'NO' in a monotheoretical methodology takes the form of a fortified 'potential falsifier', that is a sentence which, in this way of speech, we claim Nature had uttered and which is the *negation of our theory*. Nature's actual 'INCONSISTENCY' in a pluralistic methodology takes the form of a 'factual' statement couched in the light of one of the theories involved, which we claim Nature had uttered and which, if added to our proposed theories, yields an *inconsistent system*.

[127] For instance, in our earlier example some may try to replace the gravitational theory with a new one and others may try to replace the radio-optics by a new one: we choose the way which offers the more spectacular growth, the more progressive problemshift.

[128] Criticism does not *assume* a fully articulated deductive structure: it creates it. (Incidentally, this is the main message of my [1963–4].)

[129] A classical example of this pattern is Newton's relation to Flamsteed, the first Astronomer Royal. For instance, Newton visited Flamsteed on 1 September 1694, when working full time on his lunar theory; told him to reinterpret some of his data since they contradicted his own theory; and he explained to him exactly how to do it. Flamsteed

obeyed Newton and wrote to him on 7 October: "Since you went home, I examined the observations I employed for determining the greatest equations of the earth's orbit, and considering the moon's places at the times... I find that (*if, as you intimate, the earth inclines on that side the moon then is*) you may abate abt 20″ from it..." Thus Newton constantly criticized and corrected Flamsteed's observational theories. Newton taught Flamsteed, for instance, a better theory of the refractive power of the atmosphere; Flamsteed accepted this and corrected his original 'data'. One can understand the constant humiliation and slowly increasing fury of this great observer, having his data criticized and improved by a man who, on his own confession, made no observations himself: it was this feeling – I suspect – which led finally to a vicious personal controversy.

[130] The same applies to the third type of decision. If we reject a stochastic hypothesis only for one which, in our sense, supersedes it, the exact form of the 'rejection rules' becomes *less* important.

[131] Popper (1945), Vol. II, Chapter 23, p. 218.

[132] Agassi is then wrong in his thesis that "observation reports may be accepted as false and hence the problem of the empirical basis is thereby disposed of" (Agassi [1966], p. 20).

[133] One may point out that the negative and positive heuristic gives a rough (implicit) definition of the 'conceptual framework' (and consequently of the language). The recognition that the history of science is the history of research programmes rather than of theories may therefore be seen as a partial vindication of the view that the history of science is the history of conceptual frameworks or of scientific languages.

[134] Popper (1934), Sections 11 and 70. I use 'metaphysical' as a technical term of naive falsificationism: a contingent proposition is 'metaphysical' if it has no 'potential falsifiers'.

[135] Watkins (1958). Watkins cautions that "the logical gap between statements and prescriptions in the metaphysical-methodological field is illustrated by the fact that a person may reject a [metaphysical] doctrine in its fact-stating form while subscribing to the prescriptive version of it" (*Ibid.* pp. 356–7).

[136] For this Cartesian research programme, cf. Popper (1958) and Watkins (1958), pp. 350–1.

[137] Laplace (1796), Livre IV, Chapter ii.

[138] The actual hard core of a programme does not actually emerge fully armed like Athene from the head of Zeus. It develops slowly, by a long, preliminary process of trial and error. In this paper this process is not discussed.

[139] For *real* examples, cf. my (1973).

[140] The 'refutation' was each time successfully diverted to 'hidden lemmas'; that is, to lemmas emerging, as it were, from the *ceteris paribus* clause.

[141] If a scientist (or mathematician) has a positive heuristic, he refuses to be drawn into observation. He will "lie down on his couch, shut his eyes and forget about the data". (Cf. my [1963–64], especially pp. 300ff., where there is a detailed case study of such a programme.) Occasionally, of course, he will ask Nature a shrewd question: he will then be encouraged by Nature's *YES*, but not discouraged by its *NO*.

[142] Reichenbach, following Cajori, gives a different explanation of what delayed Newton in the publication of his *Principia*: "To his disappointment he found that the observational results disagreed with his calculations. Rather than set any theory, however beautiful, before the facts, Newton put the manuscript of his theory into his drawer. Some twenty years later, after new measurements of the circumference of the

earth had been made by a French expedition, Newton saw that the figures on which he had based his test were false and that the improved figures agreed with his theoretical calculation. It was only after this test that he published his law... The story of Newton is one of the most striking illustrations of the method of modern science" (Reichenbach [1951], pp. 101–2). Feyerabend criticizes Reichenbach's account (Feyerabend [1965], p. 229), but does not give an alternative *rationale*.

[143] For a further discussion of Newton's research programme, cf. my (1973).

[144] For this point cf. Truesdell (1960).

[145] Soddy's contribution to Prout's programme or Pauli's to Bohr's (old quantum theory) programme are typical examples of such creative shifts.

[146] A 'verification' is a corroboration of excess content in the expanding programme. But, of course, a 'verification' does not *verify* a programme: it shows only its heuristic power.

[147] Cf. my (1963–4), pp. 324–30. Unfortunately in 1963–4 I had not yet made a clear terminological distinction between theories and research programmes, and this impaired my exposition of a research programme in informal, quasi-empirical mathematics. There are fewer such shortcomings in my (1974).

BIBLIOGRAPHY

Agassi, J.: 1959, 'How are Facts Discovered?', *Impulse* **3**, 2–4.

Agassi, J.: 1962, 'The Confusion between Physics and Metaphysics in the Standard Histories of Sciences', in the *Proceedings of the Tenth International Congress of the History of Science*, 1964, Vol. 1, pp. 231–8.

Agassi, J.: 1964, 'Scientific Problems and Their Roots in Metaphysics', in Bunge (ed.), *The Critical Approach to Science and Philosophy*, 1964, pp. 189–211.

Agassi, J.: 1966, 'Sensationalism', *Mind* **75**, 1–24.

Agassi, J.: 1968, 'The Novelty of Popper's Philosophy of Science', *International Philosophical Quarterly* **8**, 442–63.

Agassi, J.: 1969, 'Popper on Learning from Experience', in Rescher (ed.), *Studies in the Philosophy of Science*, 1969.

Ayer, A. J.: 1936, *Language, Truth and Logic*, 1936; second edition 1946.

Bartley, W. W.: 1968, 'Theories of Demarcation between Science and Metaphysics', in Lakatos and Musgrave (eds.), *Problems in the Philosophy of Science*, 1968, pp. 40–64.

Braithwaite, R.: 1938, 'The Relevance of Psychology to Logic', *Aristotelian Society Supplementary Volumes* **17**, 19–41.

Braithwaite, R.: 1953, *Scientific Explanation*, 1953.

Carnap, R.: 1932–3, 'Über Protokollsätze', *Erkenntnis* **3**, 215–28.

Carnap, R.: 1935, Review of Popper's (1934), *Erkenntnis* **5**, 290–4.

Dreyer, J. L. E.: 1906, *History of the Planetary Systems from Thales to Kepler*, 1906.

Duhem, P.: 1906, *La Théorie Physique, Son Objet et Sa Structure*, 1905. English translation of the second (1914) edition: *The Aim and Structure of Physical Theory*, 1954.

Einstein, A.: 1949, 'Autobiographical Notes', in Schilpp (ed.), *Albert Einstein, Philosopher-Scientist*, Vol. 1, pp. 2–95.

Feyerabend, P. K.: 1959, 'Comments on Grünbaum's "Law and Convention in Physical Theory"', in Feigl and Maxwell (eds.), *Current Issues in the Philosophy of Science*, 1961, pp. 155–61.

Feyerabend, P. K.: 1965, 'Reply to Criticism', in Cohen and Wartofsky (eds.), *Boston Studies in the Philosophy of Science*, Vol. II, pp. 223–61.

Feyerabend, P. K.: 1968–9, 'On a Recent Critique of Complementarity', *Philosophy of Science* **35**, 309–31 and 36, 82–105.

Feyerabend, P. K.: 1969, 'Problems of Empiricism II', in Colodny (ed.), *The Nature and Function of Scientific Theory*, 1969.

Feyerabend, P. K.: 1970, 'Against Method', *Minnesota Studies for the Philosophy of Science* **4**, 1970.

Galileo, G.: 1632, *Dialogo dei Massimi Sistemi*, 1632.

Grünbaum, A.: 1959a, 'The Falsifiability of the Lorentz-Fitzgerald Contraction Hypothesis', *British Journal for the Philosophy of Science* **10**, 48–50.

Grünbaum, A.: 1959b, 'Law and Convention in Physical Theory', in Feigl and Maxwell (eds.), *Current Issues in the Philosophy of Science*, 1961, pp. 40–155.

Grünbaum, A.: 1960, 'The Duhemian Argument', *Philosophy of Science* **2**, 75–87.

Grünbaum, A.: 1966, 'The Falsifiability of a Component of a Theoretical System', in Feyerabend and Maxwell (eds.), *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*, 1966, pp. 273–305.

Grünbaum. A.: 1969, 'Can We Ascertain the Falsity of a Scientific Hypothesis?', *Studium Generale* **22**, 1061–93.

Hempel, C. G.: 1937, Review of Popper's (1934), *Deutsche Literaturzeitung*, 1937, pp. 309–14.

Hempel, C. G.: 1952, 'Some Theses on Empirical Certainty', *The Review of Metaphysics* **5**, 620–1.

Keynes, J. M.: 1921, *A Treatise on Probability*, 1921.

Koestler, A.: 1959, *The Sleepwalkers*, 1959.

Kuhn, T. S.: 1962, *The Structure of Scientific Revolutions*, 1962.

Kuhn, T. S.: 1965, 'Logic of Discovery or Psychology of Research', pp. 1–23.

Lakatos, I.: 1962, 'Infinite Regress and the Foundations of Mathematics', *Aristotelian Society Supplementary Volume* **36**, 155–84.

Lakatos, I.: 1963–4, 'Proofs and Refutations', *The British Journal for the Philosophy of Science* **14**, 1–25, 120–39, 221–43, 296–342.

Lakatos, I.: 1968a, 'Changes in the Problem of Inductive Logic', in Lakatos (ed.), *The Problem of Inductive Logic*, 1968, pp. 315–417.

Lakatos, I.: 1968b, 'Criticism and the Methodology of Scientific Research Programmes', in *Proceedings of the Aristotelian Society* **69**, 149–86.

Lakatos, I.: 1971, 'Popper zum Abgrenzungs- und Induktionsproblem', in H. Lenk (ed.), *Neue Aspekte der Wissenschaftstheorie*, 1971; the English version to appear under the title 'Popper on Demarcation and Induction' in Schilpp (ed.), *The Philosophy of Sir Karl Popper*.

Lakatos, I.: 1972a, 'History of Science and its Rational Reconstructions', in R. C. Buck and R. S. Cohen (eds.), *Boston Studies in the Philosophy of Science*, Vol. 8. Reidel Publishing House, 1972, pp. 91–135.

Lakatos, I.: 1972b, 'Replies to Critics', in R. C. Buck and R. S. Cohen (eds.), *Boston Studies in the Philosophy of Science*, Vol. 8. Reidel Publishing House, 1972, pp. 174–82.

Lakatos, I.: 1973, *The Changing Logic of Scientific Discovery*, 1973.

Lakatos, I.: 1974, *Proofs and Refutations and Other Essays in the Philosophy of Mathematics*, 1974.

Laplace, P.: 1796, *Exposition du Système du Monde*, 1796.

Laudan, L.: 1965, 'Grünbaum on "The Duhemian Argument"', *Philosophy of Science* **32**, 295–9.

Leibnitz, G. W.: 1678, Letter to Conring, 19.3.1678.

Le Roy, E.: 1899, 'Science et Philosophie', *Revue de Métaphysique et de Morale* **7**, 375–425, 503–62, 706–31.

Le Roy, E.: 1901, 'Un Positivisme Nouveau', *Revue de Métaphysique et de Morale* **9**, 138–53.

Medawar, P. B.: 1967, *The Art of the Soluble*, 1967.

Medawar, P. B.: 1969, *Induction and Intuition in Scientific Thought*, 1969.

Milhaud, G.: 1896, 'La Science Rationnelle', *Revue de Métaphysique et de Morale* **4**, 280–302.

Mill, J. S.: 1843, *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*, 1843.

Musgrave, A.: 1968, 'On a Demarcation Dispute', in Lakatos and Musgrave (eds.), *Problems in the Philosophy of Science*, 1968, pp. 78–88.

Musgrave, A.: 1969a, *Impersonal Knowledge*, Ph. D. Thesis, University of London, 1969.

Musgrave, A.: 1969b, Review of Ziman's 'Public Knowledge: An Essay Concerning the Social Dimensions of Science', in *The British Journal for the Philosophy of Science* **20**, 92–4.

Musgrave, A.: 1973, 'The Objectivism of Popper's Epistemology', in Schilpp (ed.), *The Philosophy of Sir Karl Popper*, 1973.

Nagel, E.: 1967, 'What is True and False in Science: Medawar and the Anatomy of Research', *Encounter* **29**, 68–70.

Neurath, O.: 1935, 'Pseudorationalismus der Falsifikation', *Erkenntnis* **5**, pp. 353–65.

Poincaré, J. H.: 1891, 'Les géométries non euclidiennes', *Revue Générale des Sciences Pures et Appliquées* **2**, 769–74.

Poincaré, J. H.: 1902, *La Science et l'Hypothèse*, 1902.

Polanyi, M.: 1958, *Personal Knowledge, Towards a Post-critical Philosophy*, 1958.

Popkin, R.: 1968, 'Scepticism, Theology and the Scientific Revolution in the Seventeenth Century', in Lakatos and Musgrave (eds.), *Problems in the Philosophy of Science*, 1968, pp. 1–28.

Popper, K. R.: 1933, 'Ein Kriterium des empirischen Charakters theoretischer Systeme', *Erkenntnis* **3**, 426–7.

Popper, K. R.: 1934, *Logik der Forschung*, 1935 (expanded English edition, Popper [1959a]).

Popper, K. R.: 1935, 'Induktionslogik und Hypothesenwahrscheinlichkeit', *Erkenntnis* **5**, 170–2; published in English in his (1959a), pp. 315–17.

Popper, K. R.: 1940, 'What is Dialectic?', *Mind*, N. S. **49**, 403–26; reprinted in Popper 1963, pp. 312–35.

Popper, K. R.: 1945, *The Open Society and Its Enemies*, I-II, 1945.

Popper, K. R.: 1957a, 'The Aim of Science', *Ratio* **I**, 24–35.

Popper, K. R.: 1957b, *The Poverty of Historicism*, 1957.

Popper, K. R.: 1958, 'Philosophy and Physics'; published in *Atti del XII Congresso Internazionale di Filosofia*, Vol. 2, 1960, pp. 363–74.

Popper, K. R.: 1959a, *The Logic of Scientific Discovery*, 1959.

Popper, K. R.: 1959b, 'Testability and "ad-Hocness" of the Contraction Hypothesis', *British Journal for the Philosophy of Science* **10**, p. 50.

Popper, K. R.: 1963, *Conjectures and Refutations*, 1963.

Popper, K. R.: 1965, 'Normal Science and its Dangers', pp. 51–8.

Popper, K. R.: 1968a, 'Epistemology without a Knowing Subject', in Rootselaar and Staal (eds.), *Proceedings of the Third International Congress for Logic, Methodology and Philosophy of Science*, Amsterdam, 1968, pp. 333–73.

Popper, K. R.: 1968b, 'On the Theory of the Objective Mind', in *Proceedings of the XIV International Congress of Philosophy* **1** (1968), 25–53.
Popper, K. R.: 1968c, 'Remarks on the Problems of Demarcation and Rationality', in Lakatos and Musgrave (eds.), *Problems in the Philosophy of Science*, 1968, pp. 88–102.
Popper, K. R.: 1969, 'A Realist View of Logic, Physics and History', in Yourgrau and Breck (eds.), *Physics, Logic and History*, 1969.
Quine, W. V.: 1953, *From a Logical Point of View*, 1953.
Reichenbach, H.: 1951, *The Rise of Scientific Philosophy*, 1951.
Russell, B.: 1914, *The Philosophy of Bergson*, 1914.
Russell, B.: 1943, 'Reply to Critics', in Schilpp (ed.), *The Philosophy of Bertrand Russell*, 1943, pp. 681–741.
Russell, B.: 1946, *History of Western Philosophy*, 1946.
Truesdell, C.: 1960, 'The Program Toward Rediscovering the Rational Mechanics in the Age of Reason', *Archive of the History of Exact Sciences* **I**, 3–36.
Watkins, J.: 1957, 'Between Analytic and Empirical', *Philosophy* **32**, 112–31.
Watkins, J.: 1958, 'Influential and Confirmable Metaphysics', *Mind*, N.S. **67**, 344–65.
Watkins, J.: 1960, 'When are Statements Empirical?', *British Journal for the Philosophy of Science* **10**, 287–308.
Watkins, J.: 1968, 'Hume, Carnap and Popper', in Lakatos (ed.), *The Problem of Inductive Logic*, 1968, pp. 271–82.
Whewell, W.: 1837, *History of the Inductive Sciences, from the Earliest to the Present Time*, Three volumes, 1837.
Whewell, W.: 1840, *Philosophy of the Inductive Sciences, Founded upon their History*, Two volumes, 1840.
Whewell, W.: 1851, 'On the Transformation of Hypotheses in the History of Science', *Cambridge Philosophical Transactions* **9**, 139–47.
Whewell, W.: 1858, *Novum Organon Renovatum*. Being the second part of the philosophy of the inductive sciences, Third edition, 1858.
Whewell, W.: 1860, *On the Philosophy of Discovery, Chapters Historical and Critical*, 1860.
Wisdom, J. T.: 1963, 'The Refutability of "Irrefutable" Laws', *The British Journal for the Philosophy of Science* **13**, 303–6.

ADOLF GRÜNBAUM

IS IT *NEVER* POSSIBLE TO FALSIFY A HYPOTHESIS IRREVOCABLY?*

In his book *The Aim and Structure of Physical Theory*, Duhem denied the feasibility of crucial experiments in physics. Said he:

> ... the physicist can never subject an isolated hypothesis to experimental test but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; *but the experiment does not designate which one should be changed* (my italics)[1].

Duhem illustrates and elaborates this contention by means of examples from the history of optics. And in each of these cases, he maintains that "If physicists had attached some value to this task",[2] any one component hypothesis of optical theory such as the corpuscular hypothesis (or so-called emission hypothesis) could have been preserved in the face of seemingly refuting experimental results such as those yielded by Foucault's experiment. According to Duhem, this continued espousal of the component hypothesis could be justified by "shifting the weight of the experimental contradiction to some other proposition of the commonly accepted optics".[3] Here Duhem is maintaining that the refutation of a component hypothesis H is at least usually no more certain than its verification could be.

In terms of the notation H and A which we have been using, Duhem is telling us that we could blame an experimentally false consequence C of the total optical theory T on the falsity of A while upholding H. In making this claim, Duhem is quite clear that the falsity of *A* no more *follows* from the experimental falsity of C than does the falsity of H. But his point is that this fact does not logically prevent us from postulating that A is false while H is true. And he is telling us that altogether the pertinent empirical facts *allow* us to reject A as false. Hence we are under no deductive logical constraint to infer the falsity of H. In questioning Dicke's purported refutation of the GTR, I gave sanction to this Duhemian contention in this instance precisely on the grounds that Dicke's auxiliary A is *not* known to be true with certainty.

In contemporary philosophy of science, a *generalized* version of

Duhem's thesis with its ramifications has been attributed to Duhem and has been highly influential. I shall refer to this elaboration of Duhem's philosophical legacy in present-day philosophy of science as 'the D-thesis'. But in doing so, my concern is with the philosophical credentials of this legacy, not with whether this attribution to Duhem himself can be uniquely sustained exegetically as against rival interpretations given by Duhem scholars. But I should remark that at least one such scholar, L. Laudan, has cited textual evidence which casts some doubt on this attribution.[4] The present philosophical appraisal is intended to supersede some parts of my earlier published critique of the D-thesis. In his Pittsburgh Doctoral Dissertation, Philip Quinn has pointed out that the version of the D-thesis with which I am concerned can be usefully stated in the form of two subtheses D1 and D2, and has argued that Laudan's attributional doubts are warranted only with respect to D2 but not with respect to D1.

The two subtheses are:

D1. No constituent hypothesis H of a wider theory can *ever* be sufficiently isolated from some set or other of auxiliary assumptions so as to be separately falsifiable observationally. H is here understood to be a constituent of a wider theory in the sense that no observational consequence can be deduced from H alone.

It is a corollary of this subthesis that *no* such hypothesis H *ever* lends itself to a crucially falsifying experiment any more than it does to a crucially verifying test.

D2. In order to state the second subthesis D2, we let T be a theory of *any* domain of empirical knowledge, and we let H be *any* of its component subhypotheses, while A is the collection of the remainder of its subhypotheses. Also, we assume that the observationally testable consequence O entailed by the conjunction $H \cdot A$ is taken to be empirically false, because the observed findings are taken to have yielded a result O′ *incompatible* with O. Then D2 asserts the following: For all potential empirical findings O′ of this kind, there exists at least one suitably revised set of auxiliary assumptions A′ such that the conjunction of H with A′ *can be held to be true and explains O′*. Thus D2 claims that H can be held to be true *and* can be used to explain O′ no matter what O′ turns out to be, i.e., *come what may.*

Note that if D2 did not assert that A′ can be held to be true in the face of the evidence no less than H, then H could not be claimed to explain O′

via A′. For premises which are already *known* to be false are not scientifically acceptable as bases for explanation.[5] Hence the part of D2 which asserts that H and A′ can each be held to be *true* presupposes that either they could not be separately falsified or that neither of them has been separately falsified.

In my prior writings on the D-thesis, I made three main claims concerning it:

(1) There are quite trivial senses in which D1 and D2 are uninterestingly true and in which no one would wish to contest them.[6]

(2) In its non-trivial form, D2 has not been demonstrated.[7]

(3) D1 is false, as shown by counterexamples from physical geometry.[8]

Since then, Gerald Massey has called my attention to yet another defect of D2 which any proponent of that thesis would presumably endeavor to remedy. Massey has pointed out that, as it stands, D2 attributes *universal* explanatory relevance and power to any one component hypothesis H. For let O* be *any* observationally testable statement *whatever* which is compatible with $\sim$O, while O′ is the conjunction $\sim O \cdot O^*$. Assume that $(H \cdot A) \rightarrow O$. Then D2 asserts the existence of an auxiliary A′ such that the theoretical conjunction $H \cdot A'$ explains the putative observational finding $\sim O \cdot O^*$. As Joseph Camp has suggested, the proponent of D2 might reply that A′ itself may potentially explain O* *without* H, even though H is essential for explaining $\sim$O via A′. But the advocate of D2 has no guarantee that he can circumvent the difficulty in this way.

Instead, he might perhaps wish to require that O′ must pertain to *the same kind of phenomena* as O, thereby ruling out 'extraneous' findings O*. Yet even if he can articulate such a restriction or provide a viable alternative to it, there is the following further difficulty noted by Massey: D2 gratuitously asserts the existence of a *deductive* explanation for any event whatever. This existential claim is gratuitous. For there *may* be individual occurrences (in the domain of quantum phenomena or elsewhere) which cannot be explained deductively, because of the irreducibly statistical character of its pertinent laws.

Our governing concern here is the question: 'Is there *any* component hypothesis H whatever whose falsity we can ascertain?' I shall try to

answer this question by giving reasons for now *qualifying* my erstwhile charge that D1 is false. Hence I shall modify the *third* of my earlier contentions about the D-thesis. But before doing so, I must give my reasons for *not* also retracting either of the first two of these contentions in response to the critical literature which they have elicited. These reasons will occupy a number of the pages that follow. The first of my earlier claims was that D1 and D2 are each true in trivial senses which are respectively exemplified by the following two examples, which I had given.[9]

i. Suppose that someone were to assert the presumably false empirical hypothesis H that 'Ordinary buttermilk is highly toxic to humans'. Then the English sentence expressing this hypothesis could be 'saved from refutation' in the face of the observed wholesomeness of ordinary buttermilk by making the following change in the theoretical system constituted by the hypothesis: changing the rules of English usage so that the intension of the term 'ordinary buttermilk' is that of the term 'arsenic' as customarily understood. In this Pickwickian sense, D1 could be sustained in the case of this particular H.

ii. In an endeavor to justify D2, let someone propose the use of an A′ which is itself of the form

$$\sim H \vee O'.$$

In that case, it is certainly true in standard systems of logic that

$$(H \cdot A') \rightarrow O'.$$

Let us now see by reference to these two examples why I regard them as exemplifications of trivially true versions of D1 and D2 respectively.

i. *D1*

Here H was the hypothesis that 'ordinary buttermilk is highly toxic to humans'. When the proponent of D1 was challenged to save this H from refutation, what he did 'save' was *not* the proposition H but the *sentence* H expressing it, as reinterpreted in the following respect: Only the term 'ordinary buttermilk' was given a new semantical usage, and *no constraint was imposed on its new usage other than that the ensuing reinterpretation turn the sentence H into a true proposition.*

If one does countenance such *unbridled* semantical instability of some of the theoretical language in which H is stated, then one can indeed

thereby uphold D1 in the form of Quine's epigram: "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system."[10] But, in that case, D1 turns into a thoroughly unenlightening truism.

I took pains to point out, however, that the commitments of D1 can also be trivially or uninterestingly fulfilled by semantical devices far more sophisticated and restrictive than the unbridled reinterpretation of some of the vocabulary in H. As an example of such a trivial fulfillment of D1 by devices whose feasibility is itself not at all trivial in other respects, I cited the following:

> ... suppose we had two particular substances I_1 and I_2 which are isomeric with each other. That is to say, these substances are composed of the same elements in the same proportions and with the same molecular weight but the arrangement of the atoms within the molecule is different. Suppose further that I_1 is not at all toxic while I_2 is highly toxic, as in the case of two isomers of trinitrobenzene. [At this point, I gave a footnote citation of 1,3,5-trinitrobenzene and of 1,2,4-trinitrobenzene respectively.] Then if we were to call I_1 'aposteriorine' and asserted that 'aposteriorine is highly toxic', this statement H could also be trivially saved from refutation in the face of the evidence of the wholesomeness of I_1 by the following device: only *partially* changing the meaning of 'aposteriorine' so that its intension is the second, highly toxic isomer I_2, thereby leaving the chemical 'core meaning' of 'aposteriorine' intact. To avoid misunderstanding of my charge of triviality, let me point out precisely what I regard as trivial here. The preservation of H from refutation in the face of the evidence by a *partial* change in the meaning of 'aposteriorine' is trivial in the sense of being only a *trivial* fulfillment of *the expectations raised by the D-thesis* [D 1]. But, in my view, *the possibility as such of preserving H by this particular kind of change in meaning is* not at all trivial. For this possibility as such reflects a fact about the world: the existence of isomeric substances of radically different degrees of toxicity (allergenicity)![11]

Mindful of this latter kind of example, I emphasized that a construal of D1 which allows itself to be sustained by this kind of alteration of the intension of 'trinitrobenzene' is no less trivial *in the context of the expectations raised by the D-thesis* than one which rests its case on calling arsenic 'buttermilk'.[12] And hence I was prompted to conclude that "a necessary condition for the nontriviality of Duhem's thesis is that the theoretical language be semantically stable in the relevant respects [i.e., with respect to the vocabulary used in H]"[13].

Mary Hesse took issue with this conclusion in her review of the essay in which I made these claims. There she wrote:

> ... it is not clear ... that 'semantic stability' *is* always required when a hypothesis is nontrivially saved in face of undermining evidence. The law of conservation of momentum

is in a sense saved in relativistic mechanics, and yet the usage of 'mass' is changed – it becomes a function of velocity instead of a constant property. But further argument along these lines is idle without more detailed analysis of what it is for a hypothesis to be the 'same', and what is involved in 'semantic stability'.[14]

This criticism calls for several comments.

(a) Hesse calls for a "more detailed analysis of what it is for a hypothesis to be the 'same', and what is involved in 'semantic stability'." To this I say that the *primary* onus for providing that more detailed analysis falls on the shoulders of the Duhemian. For he wishes to claim that his thesis D1 is interestingly true. And we saw that if he is to make such a claim, he must surely *not* be satisfied with the mere retention of the *sentence* H in some interpretation or other.

Fortunately, both the proponent and the critic of D1 can avail themselves of an apparatus of distinctions proposed by Peter Achinstein to confer specificity on what it is for the vocabulary of H to remain semantically stable.

Achinstein introduces semantical categories for describing the various possible relationships between properties of X and the term 'X'. And he appeals to the normal, standard or typical scientific *use* of a term 'X' at a given time as distinct from special uses.[15] He writes:

... I must introduce the concept of relevance and speak of a property as relevant for being an X. By this I mean that if an item is known to possess certain properties and lack others, the fact that the item possesses (or lacks) the property in question normally will count, at least to some extent, in favor of (or against) concluding that it is an X; and if it is known to possess or lack sufficiently many properties of certain sorts, the fact that the item possesses or lacks the property in question may justifiably be held to settle whether it is an X.[16]...

Two distinctions are now possible. The first is between positive and negative relevance. If the fact that an item has P tends to count more in favor of concluding that it is an X than the fact that it lacks P tends to count against it, P can be said to have more positive than negative relevance for X. The second distinction is between semantical and nonsemantical relevance and is applicable only to certain cases of relevance.

Suppose one is asked to justify the claim that the reddish metallic element of atomic number 29, which is a good conductor and melts at 1083°C, is copper. One reply is that such properties tend to count in and of themselves, to some extent, toward classifying something as copper. By this I mean that an item is correctly classifiable as copper solely in virtue of having such properties; they are among the properties which constitute a final court of appeal when considering matters of classification; such properties are, one might say, intrinsically copper-making ones. Suppose, on the other hand, one is asked to justify the claim that the substance constituting about 10^{-4} percent of the igneous rocks in the earth's crust, that is mined in Michigan, and that was used by the ancient Greeks, is copper. Among the possible replies is *not* that such properties tend

to count in and of themselves, to some extent, toward something's being classifiable as copper – that is, it is not true that something is classifiable as copper solely in virtue of having such properties. These properties do not constitute a final court of appeal when considering matters of classification. They are not intrinsically copper-making ones. Rather, the possession of such properties (among others) counts in favor of classifying something as copper solely because it allows one to infer that the item possesses other properties such as being metallic and having the atomic number 29, properties that are intrinsically copper-making ones, in virtue of which it is classifiable as copper.[17]...

Suppose $P^1, \ldots, P_n$ constitutes some set of relevant properties of X. If the properties in this set tend to count in and of themselves, to some extent, toward an item's being classifiable as an X, I shall speak of them as *semantically relevant* for X. If the possession of properties by an item tends to count toward an X-classification solely because it allows one to infer that the item possesses properties of the former sort, I shall speak of such properties as *nonsemantically relevant* for X. This distinction is not meant to apply to all relevant properties of X, for there will be cases on or near the borderline not clearly classifiable in either way.[18]

... in the latter part of the eighteenth century, with the systematic chemical nomenclature of Bergman and Lavoisier, the chemical composition of compounds began to be treated as semantically relevant; and ... chemical composition ... provided the basis for classification of compounds.

I have used the labels semantical and nonsemantical relevance because X's semantically relevant properties have something to do with the meaning or use of the term 'X' in a way that X's nonsemantically relevant properties do not.[19]...

Suppose you learn the semantically relevant properties of items denoted by the term 'X.' Then you will know those properties a possession of which by actual and hypothetical substances in and of itself tends to count in favor of classifying them as ones to which the term 'X' is applicable.[20]... Properties semantically relevant for X... include those that are logically necessary or sufficient....

Consider a term 'X' and the properties or conditions semantically relevant for X. It is perfectly possible that there be two different theories in which the term 'X' is used, where the same set of semantically relevant properties of X (or conditions for X) are presupposed in each theory (even though other properties attributed to X by these theories, properties not semantically relevant for X, might be different). If so, the term 'X' would not mean something different in each theory.[21]

Let us now employ Achinstein's distinctions to characterize a semantically stable use of the *sentence* H. I would say that one is engaged in a *semantically stable use* of the term 'X', if and only if no changes are made in the membership of the set of properties which are semantically relevant to being an item denoted by the term 'X'. And similarly for the semantically stable use of the various terms 'X_1' ($i = 1, 2, 3 \ldots n$) constituting the vocabulary of the sentence H in which a particular hypothesis (proposition) is expressed, even though these terms will, of course, not be confined to substance words or to the three major types of terms treated by Achinstein.[22] By the same token, if the entire sentence H is used in a semantically stable manner, then the hypothesis H has remained the same

in the face of other changes in the total theory. Moreover, employing different terminology, I dealt with the concrete case of geometry and optics in earlier publications to point out the following: when the rejection of a certain presumed law of optics leads to a change in the membership of the properties *non*semantically relevant to a geometrical term 'X', this change does *not* itself make for a semantic instability in the use of 'X'.[23] I shall develop the latter point further below after discussing Hesse's objection.

In the case of my example of the two isomers of trinitrobenzene, it is clear from the very names of the two isomers that the molecular structure is semantically relevant or even logically necessary to being 1, 2, 4-trinitrobenzene as distinct from 1, 3, 5-trinitrobenzene or conversely. It will be recalled that presumably only the *former* of these two isomers is toxic. The Duhemian should be concerned to be able to uphold the hypothesis '1, 3, 5-trinitrobenzene is toxic' (H) in a semantically stable manner. Hence, if he is to succeed in doing so, his use of the word '1, 3, 5-trinitrobenzene' must leave all of the semantically relevant properties of 1, 3, 5-trinitrobenzene unchanged. And thus it would constitute only a trivial fulfillment of D1 in this case to adopt a *new use* of '1, 3, 5-trinitrobenzene' which suddenly confers *positive* semantic relevance on the particular properties constituting the molecular structure of 1, 2, 4-trinitrobenzene. I do say that upholding H by such a 'meaning-switch' is trivial *for the purposes of D1*. But I do *not* thereby affirm at all the biochemical triviality of the fact which makes it possible to uphold H in this particular fashion. For I would be the first to grant that the existence of isomers of trinitrobenzene differing greatly in toxicity is biochemically significant.

(b) Hesse believes that Einstein's replacement of the Newtonian law of conservation of momentum by the relativistic one is a case in which "a hypothesis is non-trivially saved in face of undermining evidence" amid a violation of semantic stability. And she views this case as a counter-example to my claim that semantic stability is a necessary condition for the non-trivial *fulfillment of D1*. The extension of my remarks about the trinitrobenzene example to this case will now serve to show why I do not consider this objection cogent: it rests on a conflation of being non-trivial in some respect or other with being non-trivial vis-à-vis D1.

One postulational base of the special relativistic dynamics of particles combines *formal homologues* of the two principles of the conservation of

mass and momentum with the kinematical Lorentz transformations.[24] It is then shown that it is possible to satisfy the two formal conservation principles such that they are Lorentz-covariant, only on the assumption that the mass of a particle depends on its velocity. And the exact form of that velocity-dependence is derived via the requirement that the conservation laws go over into the classical laws for moderate velocities, i.e., that the relativistic mass m assume the value of the Newton mass m_0 for vanishing velocity.[25] This latter fact certainly makes it non-trivial and useful *for mechanics* to use the word 'mass' in the case of both Newtonian and relativistic mass. And hence Einstein's *formal* retention of the conservation principles is certainly *not* an instance of an unbridled reinterpretation of them. Yet despite the interesting common feature of the term 'mass', and the formal homology of the conservation principles, the two theories disagree here.

Using Achinstein's concept of mere *relevance*, we can say that the Newtonian and relativity theories disagree here as to the membership of the set of properties *relevant* to 'mass'. For it is clear that in Newton's theory, velocity-*in*dependence is positively *relevant*, in Achinstein's sense, to the term 'mass'. And it is likewise clear that velocity-*dependence* is similarly positively relevant in special relativistic dynamics. What is unclear is whether velocity *in*dependence and dependence respectively are *semantically* or *non*semantically relevant. Thus, it is unclear whether relativistic dynamics modified the Newtonian conservation principles in semantically relevant ways or preserved them semantically while modifying only the properties *non*semantically relevant to 'mass' and 'velocity'.

Let us determine the bearing of each of these two possibilities on the force of Hesse's purported counterexample to my claim that semantic stability is a necessary condition for the non-triviality of the D-thesis.

In the event of semantic relevance, semantic stability has been violated in a manner already illustrated by my trinitrobenzene example. For in that event, Newton's theory can be said to assert the conservation principles in an interpretation which assigns to the abstract word 'mass' the value m_0 as its denotatum, whereas relativity theory rejects these principles as generally false in that interpretation. But a *non*-trivial fulfillment of D1 here would have required the retention of the hypothesis of conservation of momentum in an interpretation which preserves *all* of the properties semantically relevant to 'mass', and to 'velocity' for that matter. Since this

requirement is not met on this construal, the *formal* relativistic retention of this conservation principle cannot qualify as a nontrivial fulfillment of D1. But the success of Einstein's particular semantical reinterpretation is, of course, highly illuminating in other respects.

On the other hand, suppose that relativistic dynamics has modified only the properties which are *non*-semantically relevant to 'mass' and 'velocity'. In that event, its retention of the conservation hypothesis is indeed non-trivial vis-à-vis D1. But on that alternative construal, the relativistic introduction of a velocity-dependence of mass would *not* involve a violation of semantic stability. And then this retention cannot furnish Hesse with a viable counterexample to my claim that semantic stability is necessary for non-triviality vis-à-vis D1.

Finally, suppose that here we are confronted with one of Achinstein's borderline cases such that the distinction between semantic and non-semantic relevance does not apply in the case of mass to the relevant property of velocity-dependence. In that case, we can characterize the transition from the Newtonian to the relativistic momentum conservation law with respect to 'mass' as merely a repudiation of the Newtonian claim that velocity-*in*dependence is positively relevant in favor of asserting that velocity-dependence is thus relevant. And then this transition cannot be adduced as a retention of a hypothesis H which violates semantic stability while being non-trivial vis–à–vis D1.

Since there are borderline cases between semantic and non-semantic relevance, it is interesting that Duhem has recently been interpreted as denying that the distinction between them is ever scientifically pertinent. C. Giannoni reads Duhem as maintaining that actual scientific practice always accords the same semantic role to *all* relevant properties.[26] Thus, C. Giannoni writes:

> Duhem is concerned primarily with quantities which are the subject of derivative measurement rather than fundamental measurement.... Quantities which are measured by other means than the means originally used in introducing the concept are derivative relative to this method of measurement. For example, length can be fundamentally measured by using meter sticks, but it also can be measured by sending a light beam along the length and back and measuring the time which it takes to complete the round trip. We can then calculate the length by finding the product of the velocity of light (c) and the time. Such a method of measurement is dependent not only on the measurement of time as is the first type of derivative quantity, but also on the law of nature that the velocity of light is a constant equal to c.[27]
>
> Duhem further substantiates his view... by noting that when several methods are

available for measuring a certain property, no one method is taken as the absolute criterion relative to which the other methods are derivative in the second sense of derivative measurement noted above [i.e., in Achinstein's sense of being nonsemantically relevant]. Each method is used as a check against the others.[28]

If this thesis of *semantic parity* among *all* of the relevant properties were correct, then one might argue that a semantically stable use of the sentence H would simply not be possible in conjunction with each of two different auxiliary hypotheses A and A′ of the following kind: A and A′ contain *incompatible* law-like statements each of which pertains to one or more entities that are *also* designated by terms in the sentence H. In earlier publications, I used physical geometry and optics as a test case for these claims. There I considered the actual use of geometrical and optical language in standard relativity physics[29] as well as a hypothetical Duhemian modification of the optics of special relativity.[30] Let me now develop the import of these considerations. We shall see that it runs counter to the thesis of universal semantic *parity* and sustains the feasibility of semantic stability.

The Michelson-Morley experiment involves a comparison of the round-trip times of light along equal closed spatial paths in different directions of an inertial system. As is made explicit in the account of this experiment in standard treatises, the spatial distances along the arms of the interferometer *are measured by rigid rods.*[31] The issue which arose from this experiment was conceived to be the following: Are the round-trip times *along the equal spatial paths* in an inertial frame generally *unequal*, as claimed by the aether theory, or equal as asserted by special relativity? Hence both parties to the dispute agreed that, to within a certain accuracy, equal numerical verdicts furnished by rigid rods were positively semantically relevant to the geometrical relation term 'spatially congruent' (as applied to line segments).[32] But the statement of the dispute shows that the *positive* relevance of the equality of the round-trip times of light to *spatial* congruence was made *contingent* on the particular *law of optics* which would be borne out by the Michelson-Morley experiment. Let X be the relational property of being congruent as obtaining among intervals of space. And let Y be the relational property of being traversed by light in equal round-trip times, as applied to spatial paths as well. Then we can say that the usage of 'X' at the turn of the current century was such that the relation Y was only *non*semantically relevant to the relation term 'X', whereas

identical numerical findings of *rigid rods* – to be called 'RR' – were held to be positively semantically relevant to it.

Relativity physics then added the following results pertinent to the kind of relevance which can be claimed for Y: (1) Contrary to the situation in inertial frames, in the socalled 'time-orthogonal' *non*-inertial frames the round-trip times of light in different directions along spatially *equal* paths are generally unequal.[33] (2) On a disk rotating uniformly in an inertial system I, there is generally an *inequality* among the transit times required by light departing from the same point to traverse a given closed polygonal path in opposite senses. Thus, suppose that two light signals are *jointly* emitted from a point P on the periphery of a rotating disk and are each made to traverse that periphery in opposite directions so as to return to P. Then the two oppositely directed light pulses are indeed said to traverse *equal spatial distances*, i.e., the relation X is asserted for their paths on the strength of the obtaining of the relation RR. But the one light pulse which travels in the direction *opposite* to that of the disk's rotation with respect to frame I will travel a shorter path in I, and hence will return to the disk point P earlier than the light pulse traveling in the direction in which the disk rotates. Hence the round-trip times of light for these *equal* spatial paths will be *unequal*.[34] (3) Whereas the spatial geometry yielded by a rigid rod on a stationary disk is Euclidean, such a rod yields a hyperbolic non-Euclidean geometry on a rotating disk.[35]

What is the significance of these several results? We see that space intervals which are RR are called 'congruent' ('X') in reference frames in which Y is relevant to their being *non*-X, no less than RR intervals are called 'X' in those frames in which Y is relevant to their being X! Contrary to Giannoni's general contention, this important case from geometry and optics exhibits the pertinence of the distinction between semantic and non-semantic relevance. Here RR is positively semantically relevant to being X. But Y has only *non*semantical relevance to X, as shown by the fact that while Y is relevant to X in inertial systems, it is relevant to *non*-X in at least two kinds of non-inertial systems. And this refutes the generalization which Giannoni attributes to Duhem, viz., that there is parity of positive *semantic* relevance among properties such as RR and Y, so long as there are instances in which they are all merely positively *relevant* to X.

This lack of *semantic* parity between Y and RR with respect to 'X'

allows a semantically *stable* geometrical use of 'X' on both the stationary and rotating disks, unencumbered by the *incompatibility* of the optical laws which relate the property Y to X in these two different reference frames. Indeed, by furnishing a *tertium comparationis*, this semantic stability confers *physical* interest on the contrast between the incompatible laws of optics in the two disks, and on the contrast between their two incompatible spatial geometries! For in *both* frames, RR is alike centrally semantically relevant to being X, i.e., in each of the two frames, rigid rod coincidences serve alike to determine what space intervals are assigned equal measures $ds = \sqrt{(g_{ik} dx^i dx^k)}$. This sameness of RR, which makes for the semantic stability of 'X', need not, however, comprise a sameness of the correctional physical laws that enable us to allow computationally for the thermal and other deviations from rigidity.

I used certain formulations of relativity theory as a basis for the preceding account of the properties semantically and nonsemantically relevant to 'X' ('spatially congruent'). But this account does not, of course, gainsay the legitimacy of alternative uses of 'X' that would issue in alternative, albeit physically equivalent, formulations. For example, the relativistic interpretation of the upshot of the Michelson-Morley experiment could alternatively have been stated as follows: In any inertial system, distances which are equal in the metric based on light-propagation are equal also in the metric which is based on rigid measuring rods.[36] I would say that this particular alternative formulation places Y on a par with RR as a *candidate* for being semantically relevant to 'X'. Moreover, I have strongly emphasized elsewhere that we are indeed free to formulate certain physical theories alternatively so as to make Y rather than RR semantically relevant to 'X'.[37]

So much for matters pertaining to trivial fulfillments of D1.

ii. *D2*

It will be recalled that my example of a trivial fulfillment of D2 involved the use of an A′ which is itself of the form $\sim H \vee O'$. In that example, the triviality did *not* arise from a violation of semantic stability. Instead, the fulfillment of D2 by the formal validity of the statement $[H \cdot (\sim H \vee O')] \rightarrow O'$ is only trivial because the entailment holds independently of the specific assertive content of H, unless a restrictive kind of entailment is used as that of the system E of Anderson and Belnap. No matter what H

happens to be about, and no matter whether H is substantively relevant to the specific content of O′ or not, O′ will be entailed by H via the particular A′ used here. Hence H does *not* serve to explain via A′ the facts asserted by O′, thereby satisfying D2 here only trivially, *if at all.*

As Philip Quinn has shown,[38] one of the fallacies which vitiates J. W. Swanson's purported syntactical proof that D2 holds interestingly[39] is the failure to take cognizance of this particular kind of semantic trivialization or even outright falsity of D2.

This brings us to the second of my previously published claims. The latter was that in its non-trivial form, D2 has not been demonstrated by means of D1. Having dealt with several *necessary* conditions for the non-triviality of A′, I therefore went on to comment on a possible *sufficient* condition by writing:

> I am unable to give a formal and completely general *sufficient* condition for the *non*-triviality of A′. And, so far as I know, neither the originator nor any of the advocates of the D-thesis [D2] have ever shown any awareness of the need to circumscribe the class of *non*trivial revised auxiliary hypotheses A′ so as to render the D-thesis [D2] interesting. I shall therefore assume that the proponents of the D-thesis intend it to stand or fall on the kind of A′ which we would all recognize as *non*trivial *in any given case*, a kind of A′ which I shall symbolize by A'_{nt}.[40]

And I added that D2 was *undemonstrated*, since it does *not* follow from D1 that

$$(\exists A'_{nt})\,[(H \cdot A'_{nt}) \rightarrow O'].$$

Mary Hesse discusses my charge that D2 is a non-sequitur, as well as my erstwhile objections to D1, which I shall qualify below. And she addresses herself to a particular example of mine which was of the following kind: the *original* auxiliary hypothesis A ingredient in the conjunction H·A is highly confirmed quite separately from H, but the conjunction H·A does entail a presumably incorrect observational consequence. Apropos of this example, she writes:

> ... Grünbaum admits that... "A is only more or less highly confirmed" (p. 289) – a significant retreat from the claim of truth. He thinks nevertheless that it is crucial that confirmation of A can be *separated* from that of H. But surely this is not sufficient for his purpose. For as long as it is not empirically demonstrable, but only likely, that A is true, it will always be possible to reject A in order to save H.... Which is where the Duhem thesis came in. It must be concluded that D [the D-thesis] still withstands Grünbaum's assaults....[41]

It is not clear how Hesse wants us to construe her phrase "in order to save H", when she tells us that "it will always be possible to reject A in order to save H". If she intends that phrase to convey merely the same as "with a view to attempting to save H", then her objection pertains only to my critique of D1, which is not now at issue. And in that case, she cannot claim to have shown that the D-thesis "still withstands Grünbaum's assaults" with respect to my charge of non-sequitur against D2. But if she interprets "in order to save H" in the sense of D2 as meaning the same as "with the assurance that there exists an A'_{nt} that permits H to explain O′", then I claim that she is fallaciously deducing D2 from the assumed truth of D1. I quite agree that if D1 is assumed to be true *and* IF THERE EXISTS AN A'_{nt} such that H does explain O′ via A'_{nt}, then D1 would guarantee the feasibility of upholding A'_{nt} or H (but not necessarily both) as true. But this fact does not suffice at all to establish that there *exists* such an A'_{nt}! Thus, my charge of non-sequitur against the purported deduction of D2 from D1 does *not* rest on the complaint that the Duhemian cannot assure *a priori* being able to marshal supporting evidence *for* A'_{nt}.

Thus, if Hesse's criticisms were intended, in part, to invalidate my charge of non-sequitur against D2, then I cannot see that they have been successful. And since Philip Quinn has shown[42] that D2, in turn, does *not* entail D1, I maintain that D1 and D2 are logically independent of one another.

This concludes the statement of my reasons for not retracting either of the first two of my three erstwhile claims, which were restated early in this section. Hence we are now ready to consider whether there are any known counterexamples to D1.

Thus, our question is: Is there any component hypothesis H of some theory T or other such that H can be sufficiently isolated so as to lend itself to separate observational falsification? It is understood here that H *itself* does not have any observational consequences. The example which I had adduced to answer this question affirmatively in earlier publications was drawn from physical geometry. I now wish to re-examine my claim that the example in question is a counterexample to D1. In order to do so, let us consider an arbitrary surface S. And suppose that the Duhemian wants to uphold the following component hypothesis H about S: *if* lengths $ds = \sqrt{(g_{ik}dx^i dx^k)}$ are assigned to space intervals by means of *rigid* unit rods, then Euclidean geometry is the metric geo-

metry which prevails physically on the given surface S. Before investigating whether and how this hypothesis H might be falsified, we must be mindful of an important assumption which is ingredient in the antecedent of its if-clause.

That antecedent tells us that the numbers furnished by *rigid* unit rods are to be centrally semantically relevant to the interval measures ds of the theory. Thus, it is a necessary condition for the *consistent* use of rigid rods in assigning lengths to space intervals that any collection of two or more initially coinciding unit solid rods of whatever chemical constitution can thereafter be used *interchangeably* everywhere in the region S, *unless* they are subjected to independently designatable perturbing influences. Thus, the assumption is made here that there is a concordance in the coincidence behavior of solid rods such that no inconsistency would result from the subsequent interchangeable use of initially coinciding unit rods which remain *unperturbed* or 'rigid' in the specified sense. In short, there is concordance among rigid rods such that all rigid rods *alike* yield the same metric ds and thereby the same geometry. Einstein stated the relevant empirical law of concordance as follows:

> All practical geometry is based upon a principle which is accessible to experience, and which we will now try to realize. We will call that which is enclosed between two boundaries, marked upon a practically-rigid body, a tract. We imagine two practically-rigid bodies, each with a tract marked out on it. These two tracts are said to be 'equal to one another' if the boundaries of the one tract can be brought to coincide permanently with the boundaries of the other. We now assume that:
>
> If two tracts are found to be equal once and anywhere, they are equal always and everywhere.
>
> Not only the practical geometry of Euclid, but also its nearest generalisation, the practical geometry of Riemann and therewith the general theory of relativity, rest upon this assumption.[43]

I shall refer to the empirical assumption just formulated by Einstein as 'Riemann's concordance assumption', or, briefly, as 'R'. Note that the assumption R predicates the concordance among rods of different chemical constitution on their being rigid or free from perturbing influences. Perturbing influences are exemplified by inhomogeneities of temperature and by the presence of electric, magnetic and gravitational fields. Thus, two unit rods of different chemical constitution which initially coincide when at the same standard temperature will generally experience a destruction of their initial coincidence if brought to a place at

which the temperature is different, and the amount of their thermal elongation will depend on their chemical constitution. By the same token, a wooden rod will shrink or sag more in a gravitational field than a steel one. Since such perturbing forces produce effects of different magnitude on different kinds of solid rods, Reichenbach has called them 'differential forces'. The perturbing influences which qualify as differential are linked to designatable sources, and their presence is certifiable in ways other than the fact that they issue in the destruction of the initial coincidence of chemically different rods. But the set of perturbing influences is open-ended, since physics cannot be presumed to have discovered *all* such influences in nature by now. This open-endedness of the class of perturbing forces must be borne in mind if one asserts that the measuring rods in a certain region of space are free from perturbing influences or 'rigid'.

It is clear that if our surface S is indeed free from perturbing influences, then it follows from the assumption R stated by Einstein that *any* two rods of different chemical constitution which initially coincide in S will coincide everywhere else in S, independently of their respective paths of transport. This result has an important bearing on whether it might be possible to falsify the putative Duhemian hypothesis H that the geometry G of our surface S is Euclidean.

For suppose now that S does satisfy the somewhat idealized condition of actually being macroscopically free from perturbing influences. Let us call the auxiliary assumption that S is thus free from perturbing influences 'A'. Then the conjunction H·A entails that measurements carried out on S should yield the findings required by Euclidean geometry. Among other things, this conjunction entails, for example, that the ratio of the periphery of a circle to its diameter should be π on S. But suppose that the surface S is actually a sphere rather than a Euclidean plane and that the measurements carried out on S with presumedly rigid rods yield various values significantly *less* than π for this ratio, depending on the size of the circle. How then is the Duhemian going to uphold his hypothesis H that if lengths ds on S are measured with rigid rods, the geometry G of S will be Euclidean so that all circles on S will exhibit the ratio π independently of their size?

Clearly, he will endeavor to do so by denying the initial auxiliary hypothesis A, and asserting instead the following: the rods in S are not rigid after all but are being subjected to perturbing influences constituted by

differential forces. Obviously, just as in the case of Dicke's refutation of Einstein's theory, the falsification of H itself requires that the truth of A be established, so that Duhem would not be able to deny A. Suppose that we seek to establish A on the strength of the following further experimental findings: However different in chemical appearance, all rods which are found to coincide initially in S preserve their coincidences everywhere in S to within experimental accuracy independently of their paths of transport. How well does this finding establish the truth of A, i.e., that S is *free* from all perturbing influences which would act differentially on rods of diverse chemical constitution?

Let us use the letter 'C' to denote the statement that whatever their chemical constitution, any and all rods invariably preserve their initial coincidences under transport in S. Then we can say that Riemann's concordance assumption R states the following: If A is true of *any* region S, then C is true of that region S. Clearly, therefore, C follows from the conjunction A·R, but *not* from A alone. Hence if the observed preservation of coincidence or concordance in S establishes anything here, what it does establish at best is the conjunction A·R rather than A alone, since I would *not* wish to argue against the Duhemian holist that any finding C which serves inductively to establish a conjunction A·R must be held to establish likewise each of the conjuncts separately.[44] This fact does not, however, give the Duhemian a basis for an objection to my attempt to establish A. To see this, recall that R must be assumed by the Duhemian if his claim that S is Euclidean is to have the physical significance which is asserted by his hypothesis H. Since the Duhemian wants to uphold H, he could not contest R but only A. Hence, in challenging my attempt to establish A, the Duhemian will *not* be able to object that the observed concordance of the rods in S could establish only the conjunction A·R rather than establishing A *itself*. The issue is therefore one of establishing A. And for the reasons given in Einstein's statement of R above, R is assumed both by the Duhemian, who claims that our surface S is a Euclidean plane, and by the anti-Duhemian, who maintains that S is a sphere.

The Duhemian can argue that the observed concordance cannot conclusively establish A for the following two reasons: (1) He can question whether C can establish A, even if the *apparent* concordance is taken to establish the *universal* statement C indubitably, and (2) he can challenge the initial inference from the seeming concordance to the physical truth

of C. Let me first articulate each of these two Duhemian objections in turn. Thereafter, I shall comment critically on them.

Objection 1

The Duhemian grants that the mere presence of a *single* perturbing influence of significant magnitude would have produced differential effects on chemically different rods. Thus, he admits that the presence of a *single* such influence would be incompatible with C. But he goes on to point out that C *is* compatible with the joint presence of several perturbing influences of possibly unknown physical origin which just *happen* to act as follows: different effects produced on the rods by each one of these perturbing influences are of just the right magnitude to combine into the same *total* deformation of each of the rods. And all being deformed *alike* by the collective action of differential forces *emanating from physical sources*, the rods behave in accord with C. In short, the Duhemian claims that such a *superposition* of differential effects is compatible with C and that it therefore cannot be ruled out by C. Hence he concludes that C cannot be held to establish the truth of A beyond question. Since this Duhemian objection is based on the possibility of conjecturing the specified kind of superposition, I shall refer to it as "the superposition objection".[45]

Objection 2

This second Duhemian objection calls attention to the fact that in inferring C from the *apparent* concordance, we have made the following several inductive inferences: (i) We have taken differences in chemical *appearance* to betoken actual differences in chemical constitution, (ii) we have taken apparent coincidence to betoken actual coincidence, at least to within a certain accuracy, and (iii) we have inductively inferred the *invariable* preservation of coincidence by *all* kinds of rods everywhere in S from only a limited number of trials. The first two of these three inferences invoke collateral hypotheses. For they rest on presumed laws of psychophysics and neurophysiology to the following effect: certain appearances, which are contents of the awareness of the human organism, are lawfully correlated with the objective physical presence of the respective states of affairs which they are held to betoken. Hence let us say that the first two of these three inferences infer a conclusion of the form 'Physical item P

has the property Q' from the premise 'P *appears* to be Q', or that they infer *being* Q from *appearing* to be Q.

These objections prompt two corresponding sets of comments.

1. It is certainly true that A is *not* entailed by the conjunction R·C. For R is the conditional statement 'If A, then C'. Since A does not follow deductively from C·R, the premise C·R *cannot* be held to rule out the rival superposition conjecture DEDUCTIVELY. Thus, given R, C *deductively allows* the superposition conjecture as a *conceivable* alternative to A. Nevertheless, we shall now see that in the context of R, the inductive confirmation of A by C is so enormously high that A can be reregarded as *well-nigh established* by R·C. Since my impending inductive argument will make use of Bayes' theorem, it will be unconvincing to those who question the applicability of that theorem to the probability of a *hypothesis*. For example, Imre Lakatos registered an objection in this vein.[46]

I shall adopt the non-standard probability notation employed by H. Reichenbach[47] and W. C. Salmon[48] to let 'P(L, M)' mean 'the probability *from* L to M' or 'the probability of M, given L'. In the usual notations, the two arguments L and M are reversed.

Recalling that R says 'If A, then C' (for any region S), we wish to evaluate the probability P(R·C, A) of A, given that R·C. To do so, we use a special form of Bayes' theorem[49] and write

$$P(R\cdot C,A)=\frac{P(R,A)}{P(R,C)}\times P(R\cdot A,C).$$

C follows *deductively* from R·A. Hence C is guaranteed by R·A, so that

$$P(R\cdot A,C)=1.$$

Thus, our formula shows that our desired posterior probability P(R·C,A) will be very high, if the ratio P(R,A)/P(R,C) is itself close to 1, as indeed it will now turn out to be. Needless to say, the *ratio* of these probabilities can be close to 1, even though they are individually quite small.

We can compare the two probabilities in this ratio with respect to magnitude without knowing their individual magnitudes. To effect this comparison, we first consider the conditional probability of C, if we are given that R is true while A is *false*. We can now see that this probability $P(R\cdot\sim A,C)$ is very low, whereas we recall that $P(R\cdot A,C)=1$. For sup-

pose that A is false, i.e., that S *is* subject to differential forces. Then C can hold only in the *very rare* event that there *happens* to be just the right superposition. Incidentally in the case of our surface S, there is no evidence at all for the existence of the *physical sources* to which the Duhemian needs to attribute his superposed differential effects. Hence we are confronted here with a situation in which the Duhemian wants C to hold under conditions in which superposed differential forces are actually operative, although there is no independent evidence whatever for them.[50]

Since $P(R \cdot \sim A,C)$ is very low, we can say that the probability $P(R,C)$ of concordance among the chemically different rods is *only slightly greater, if at all,* than the probability $P(R, A)$ of the occurrence of a region S which is free from differential forces. And this comparison among these probabilities holds here, notwithstanding the fact that there may be as yet undiscovered kinds of differential forces in nature, as I stressed above. It follows that $P(R \cdot C,A)$ is close to 1. And since

$$P(R \cdot C, \sim A) = 1 - P(R \cdot C,A),$$

the conditional probability of $\sim A$ is close to zero.[51]

Thus, the Duhemian's first objection does not detract, and indeed may not be intended to detract from the fact that A is *well-nigh* established by $R \cdot C$.

2. The second objection noted that being Q is inferred inductively from appearing to be Q in the case of chemical difference and of coincidence. Furthermore, it called attention to the inductive inference of the *universal* statement C from only a limited number of test cases. What is the force of these remarks?

Let me point out that the claim of mere chemical *difference* does *not* require the definite identification of the particular chemical constitution of each of the rods as, say, iron or wood.[52] Now suppose that the apparent chemically relevant differences among most of the tested rods are striking. And note that they can be further enriched by bringing additional rods to S. Would it then be helpful to the Duhemian to postulate that all of these prima facie differences in chemical appearance mask an underlying chemical identity? If he wishes to avoid resorting to the fantastic superposition conjecture, the Duhemian might be tempted to postulate that the great differences in chemical appearance belie a true crypto-identity of chemical constitution. For by postulating his crypto-identity,

he could hope to render the observed concordance among the rods unproblematic for $\sim$A *without* the superposition conjecture. Crypto-identity might make superposition dispensable because the Duhemian might then be able to assert that S is subject to only *one* kind of differential force.[53] But there is no reason to think that the crypto-identity conjecture is any more probable than the incredible superposition conjecture. And either conjecture may try a working scientist's patience with philosophers.[54]

As for the inductive inferences of physical coincidence from apparent coincidence, the degree of accuracy to which it can be certified inductively is, of course, limited and macroscopic. But what would it avail the Duhemian to assume in lieu of superposition the imperceptibly slight differential deformations which are compatible with the limited accuracy of the observations of coincidence? Surely the imprecision involved here does not provide adequate scope to modify A sufficiently to be able to reconcile the substantial observed deviations of the circular ratios from π with the Euclidean H that he needs to save.

What of the inductive inference of the universal statement C from only a limited number of test cases? Note that C need not be established in its full universality in order to create inductive difficulties for the superposition conjecture and to provide strong support for A. For as long as all the chemically different rods actually tested satisfy C, the Duhemian must make his superposition conjecture or crypto-identity assumption credible with respect to *these* rods. Otherwise, he could not reconcile the measurements of circular ratios furnished by them on S with his Euclidean H.

The Duhemian might point out that I have inferred A inductively in a two-step inductive inference, which first infers at least instances of the physical claim C from apparent coincidences of prima facie different kinds of rods, and then proceeds to infer A inductively from these instances of C, coupled with R. And he might object that the relation of inductive support is not transitive. It is indeed the case that the relation of inductive support fails to be transitive.[55] But my inductive inference of A does not require an intermediate step via C, for cases of *apparent* coincidence, no less than cases of *physical* coincidence, are correlated with A·R, and R is no more in question here than before. Thus, given R, the inference can proceed to A directly via the assumption that most cases of apparent coincidence of prima facie chemically different rods are corre-

lates of cases of A.[56] And my inductive inference of A is thus predicated on this correlation.

In conclusion, we can now try to answer two questions: (i) To what extent has our verification of A been separate from the assumption of H?, and (ii) to what extent has H *itself* been falsified? When I speak here of a hypothesis as having been 'falsified' in asking the latter question, I mean that the *presumption* of its falsity has been established, *not* that its falsity has been established with certainty or irrevocably. This construal of 'falsify', is of course, implicit in the presupposition of my question that there are different degrees of falsification or differences in the extent to which a hypothesis can be falsified.

(i) Our verification of A did proceed in the context of the assumption of R. And while we saw that R is ingredient in the Euclidean H of our example, as specified, R is similarly ingredient in the rival hypothesis that our surface S is not a Euclidean plane but a spherical surface. Thus, our verification of A was separate from the assumption of the *distinctive* physical content of the particular Duhemian H.

(ii) Duhem attributed the inconclusiveness of the falsification of a component hypothesis H to the legitimacy of denying instead any of the collateral hypotheses A which enter into any test of H. Our analysis has shown that *the denial of A is legitimate precisely to the extent that its* VERIFICATION *suffers from inductive uncertainty*. Moreover, in each of our examples of attempted falsification, the inconclusiveness is attributable *entirely* to the inductive uncertainty besetting the following two *verifications*: the verification of A, and the verification of the so-called observation statement which entails the *falsity* of the conjunction H·A. In short, the inconclusiveness of the falsification of a component H derives wholly from the inconclusiveness of verification. And the falsification of H itself is inconclusive or revocable in the sense that the falsity of H is *not* a *deductive* consequence of premises *all* of which can be known to be true with certainty.

Hence if the falsification of H denied by Duhem's D1 is construed as *irrevocable*, then I agree with Mary Hesse[57] that my geometrical example does not qualify as a counterexample to D1. But I continue to claim that it does so qualify, if one requires only the *very strong presumption* of falsity. And to the extent that my geometrical example does falsify the A'_{nt} which D2 invokes in conjunction with the H of that example, the

example also refutes D2's claim that A'_{nt} can justifiably be held to be true.

Subject to an important *caveat* to be issued presently, I maintain, therefore, that there are cases in which we can establish a strong presumption of the falsity of a component hypothesis, although we cannot falsify H in these cases beyond any and all possibility of subsequent rehabilitation. Thus, I emphatically do allow for the following possibility, though not likelihood, in the case of an H which has been falsified in my merely presumptive sense: A daring and innovative scientist who continues to entertain H, albeit as part of a new research program which he envisions as capable of vindicating H, *may* succeed in incorporating H in a theory so subsequently fruitful and well-confirmed as to retroactively alter our assessment of the initial falsification of H. And my *caveat* is that my conception of the falsification of H as establishing the strong presumption of falsity is certainly *not* tantamount to a stultifying injunction to any and all imaginative scientists to cease entertaining H forthwith, whenever such falsification obtains at a given time! Nor is this conception of falsification to be construed as being committed to the historically false assertion that no inductively *un*warranted and daring continued espousal of H has ever been crowned with success in the form of subsequent vindication.

Appendix

One might ask: Are there no *other* cases at all in which H can be justifiably rejected as *irrevocably* falsified by observation? On the basis of a schema given by Philip Quinn, it may *seem* that *if* we can ignore such uncertainty as attaches to our observations when testing H, then this question can be answered affirmatively for reasons to be stated presently. When asserting the prima facie existence of this kind of irrevocable falsification as a fact of logic, Quinn recognized that its relevance to actual concrete cases in the pursuit of empirical science is at best very limited. Specifically, Quinn invites consideration of the following kind of logical situation. Let the observation statement O_3 be pairwise *incompatible* with each of the two observation statements O_1 and O_2, so that $O_3 \rightarrow \sim O_1$, and $O_3 \rightarrow \sim O_2$. And suppose further that

$$[(H \cdot A) \rightarrow O_1] \cdot [(H \cdot \sim A) \rightarrow O_2].$$

Assume also that observations made to test H are taken to establish

the truth of O_3. Then we can deduce via *modus tollens* that

$$\sim(H \cdot A) \cdot \sim(H \cdot \sim A).$$

Using the law of excluded middle to assert $A \vee \sim A$, we can write

$$(A \vee \sim A) \cdot [\sim(H \cdot A) \cdot \sim(H \cdot \sim A)].$$

The application of the distributive law and the omission of one of the two conjuncts in the brackets from each of the resulting disjuncts yields

$$[A \cdot \sim(H \cdot A)] \vee [\sim A \cdot \sim(H \cdot \sim A)].$$

Using the principle of the complex constructive dilemma, we can deduce $\sim H \vee \sim H$ and hence $\sim H$.

Although the assumed truth of the observation statement O_3 allows us to deduce the falsity of H in this kind of case, its relevance to empirical science is at best very limited for the following reasons.

In most, if not all, cases of actual science, the conjunction $H \cdot \sim A$ will not be rich enough to yield an observational consequence O_2, if the conjunction $H \cdot A$ does yield an observational consequence O_1. The reason is that the mere denial of A is not likely to be sufficiently specific in content. What, for example, is the observational import of conjoining the *denial* of Darwin's theory of evolution to a hypothesis H concerning the age of the earth? Thus this feature at least severely limits the relevance of this second group of logically possible cases to actual science.

Furthermore, the certainty of the conclusion that H is false in this kind of case rests on certainty that the observational statement O_3 is true. Such certainty is open to question. But *if* we can ignore our doubts on this score, then it would seem that H can be held to have been falsified beyond any possibility of subsequent rehabilitation. Yet, alas, John Winnie has pointed out that despite appearances to the contrary, the H in Quinn's schema does not qualify as a component hypothesis after all.

For Winnie has noted that Quinn's basic premise $[(H \cdot A) \rightarrow O_1] \cdot [(H \cdot \sim A) \rightarrow O_2]$ permits the deduction of

$$H \rightarrow (O_1 \vee O_2).$$

Thus, if $O_1 \vee O_2$ can be held to qualify as an observationally testable statement, then we find that H entails it *without* the aid of the auxiliary A or of $\sim A$. But in that case, Winnie notes that H does not qualify as a

constituent of a wider theory in the sense specified in our statement of D1 above.[58]

NOTES

* From Chapter 17 of *Philosophical Problems of Space and Time*, second, enlarged ed., pp. 585–629. Published by D. Reidel, Boston and Dordrecht, 1974 as Vol. XII in the *Boston Studies in the Philosophy of Science* (ed. by R. S. Cohen and M. Wartofsky). In this form the essay also appeared in M. Mandelbaum (ed.), *Observation and Theory in Science*, Johns Hopkins Press, Baltimore, 1971. Reprinted by permission.

[1] Pierre Duhem, *The Aim and Structure of Physical Theory*, Princeton University Press, Princeton, 1954, p. 187.

[2] *Ibid.*

[3] *Ibid.*, p. 186.

[4] Laurens Laudan, 'On the Impossibility of Crucial Falsifying Experiments: Grünbaum on "The Duhemian Argument"', *Philosophy of Science* **32** (1965), 295–99.

[5] For a discussion of the epistemic requirement that explanatory premises must *not* be known to be false, see Ernest Nagel, *The Structure of Science*, Harcourt, Brace and World, New York, 1961, pp. 42–43.

[6] A. Grünbaum, 'The Falsifiability of a Component of a Theoretical System', in *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*, P. K. Feyerabend and G. Maxwell, eds., University of Minnesota Press, Minneapolis, 1966, pp. 276–80.

[7] *Ibid.*, pp. 280–81.

[8] *Ibid.*, pp. 283–95; and A. Grünbaum, *Geometry and Chronometry in Philosophical Perspective*, University of Minnesota Press, Minneapolis, 1968, Chap. III, pp. 341–51. In Chap. III, Section 9.2, pp. 351–69 of the latter book, I present a counterexample to H. Putnam's *particular* geometrical version of D2. For a brief summary of Putnam's version, see Note 55.

[9] Grünbaum, 'The Falsifiability of a Component of a Theoretical System', pp. 277–78.

[10] W. V. O. Quine, *From a Logical Point of View* (revised ed.), Harvard University Press, Cambridge, Mass., 1961, pp. 43 and 41n.

[11] Grünbaum, 'The Falsifiability of a Component of a Theoretical System', pp. 279–80.

[12] *Ibid.*, p. 279.

[13] *Ibid.*, p. 278.

[14] Mary Hesse, *The British Journal for the Philosophy of Science* **18** (1968), 334.

[15] Peter Achinstein, *Concepts of Science*, Johns Hopkins Press, Baltimore, 1968, pp. 3ff.

[16] *Ibid.*, p. 6.

[17] *Ibid.*, pp. 7–8.

[18] *Ibid.*, pp. 8–9.

[19] *Ibid.*, p. 9.

[20] *Ibid.*, p. 35.

[21] *Ibid.*, p. 101.

[22] *Ibid.*, p. 2.

[23] See pp. 143–44 of A. Grünbaum, *Philosophical Problems of Space and Time*, second, enlarged ed., D. Reidel Publ. Co., Boston and Dordrecht, 1974; and A. Grünbaum, *Geometry and Chronometry...*, pp. 314–17.

[24] See, for example, Tolman, *Relativity, Thermodynamics and Cosmology*, pp. 42–45.

[25] See P. G. Bergmann, *Introduction to the Theory of Relativity*, Prentice-Hall, New York, 1946, pp. 86–88.
[26] C. Giannoni, 'Quine, Grünbaum, and The Duhemian Thesis', *Nous* **1** (1967), 288.
[27] *Ibid.*, pp. 286–87.
[28] *Ibid.*, p. 288. The example which Giannoni then goes on to cite from Duhem is one in which a very *weak* electric current was running through a battery and where, therefore, one or more indicators may fail to register its presence. In that case, the current will still be said to flow if one or another indicator yields a positive response. This particular case may well be a borderline one as between semantic and nonsemantic relevance.
[29] Grünbaum, *Geometry and Chronometry...*, pp. 314–17.
[30] See pp. 143–44 of Grünbaum, *Philosophical Problems of Space and Time.*
[31] E.g., in Bergmann, *Introduction to the Theory of Relativity*, pp. 23–26.
[32] Since the aether-theoretically expected time difference in the second order terms is only of the order of 10^{-15} second, allowance had to be made in practice for the absence of a corresponding accuracy in the measurement of the equality of the two arms. This is made feasible by the fact that, on the aether theory, the effect of any discrepancy in the lengths of the two arms should *vary*, on account of the earth's motion, as the apparatus is rotated. For details, see Bergmann, *Introduction to the Theory of Relativity*, pp. 24–26, and J. Aharoni, *The Special Theory of Relativity*, Oxford University Press, Oxford, 1959, pp. 270–73. Indeed, slightly unequal arms are needed to produce neat interference fringes.
[33] Grünbaum, *Geometry and Chronometry in Philosophical Perspective*, pp. 316–17.
[34] *Ibid.*, pp. 314–15.
[35] See, for example, C. Møller, *Theory of Relativity*, Oxford University Press, Oxford, 1952, Ch. VIII, Section 84.

For a rebuttal to John Earman's objections to ascribing a spatial geometry to the rotating disk, see the detailed account of the status of metrics in A. Grünbaum, 'Space, Time and Falsifiability', Part I, *Philosophy of Science* **37** (1970), 562–65 (Chapter 16 of Grünbaum, *Philosophical Problems of Space and Time*, pp. 542–545).
[36] This formulation is given in E. T. Whittaker, *From Euclid to Eddington*, Cambridge University Press, London, 1949, p. 63.
[37] Grünbaum, *Geometry and Chronometry..., passim.*
[38] Philip Quinn, 'The Status of the D-Thesis', Section III, *Philosophy of Science* **36**, (1969), No. 4.
[39] J. W. Swanson, 'On the D-Thesis', *Philosophy of Science* **34** (1967), 59–68.
[40] Grünbaum, 'The Falsifiability of a Component of a Theoretical System', p. 278.
[41] Hesse, *The British Journal for the Philosophy of Science* **18**, 334–35.
[42] University of Pittsburgh Doctoral Dissertation (1970).
[43] A. Einstein, 'Geometry and Experience', in *Readings in the Philosophy of Science*, in H. Feigl and M. Brodbeck (eds.), Appleton-Century-Crofts, New York, 1953, p. 192.
[44] By the same token, if *alternatively* the initial coincidence of the rods had been observed to *cease conspicuously* in the course of transport, the latter observations could *not* be held to falsify A in isolation from R. This inconclusiveness with respect to the falsity of A itself would obtain in the face of the *apparent* discordances, even if the latter are taken as *indubitably* falsifying the claim C of universal physical concordance. For C is entailed by A·R rather than by A alone.
[45] This kind of objection is raised by L. Sklar, 'The Falsifiability of Geometric

Theories', *The Journal of Philosophy* **64** (1967), 247–53. Swanson, 'On the D-Thesis', presents a model universe for which he conjectures superposition of differential effects *in name only*. For, as Philip Quinn has noted (in 'The Status of the D-Thesis'), Swanson effectively adopts the convention that even in the *absence* of perturbing influences, all initially coinciding rods will be held to be *non*-rigid by being assigned lengths that *vary alike* with their positions and/or orientations. Thus, Swanson renders the superposition conjecture physically empty by ignoring the crucial requirement that the alleged differential effects are held to emanate individually from physical sources. Swanson's model universe will be discussed further in Note [54].

[46] See Imre Lakatos and Alan Musgrave (eds.), *Criticism and Growth of Knowledge*, Cambridge University Press, New York, 1970, p. 187. But if he countenances entertaining the superposition conjecture $\sim$A on the grounds that A cannot be said to be highly confirmed, how will this $\sim$A escape being a *metaphysical* proposition?

[47] H. Reichenbach, *The Theory of Probability*, University of California Press, Berkeley, 1949.

[48] W. C. Salmon, *The Foundations of Scientific Inference*, University of Pittsburgh Press, Pittsburgh, 1967, p. 58.

[49] For this form of the theorem, see Reichenbach, *Theory of Probability*, p. 91, Equation (6).

[50] This lack of evidence is here construed as obtaining at the given time. An alternative construal in which there avowedly would *never* be any such evidence at all is tantamount to an admission that A is true after all, or that $\sim$A is physically empty. See Philip Quinn ['The Status of the D-Thesis',] for a discussion of the import of this point for L. Sklar's *particular* superposition objection.

[51] H. Putnam claims to have shown in 'An Examination of Grünbaum's Philosophy of Geometry', in *Philosophy of Science*, The Delaware Seminar, Vol. 2, B. Baumrin (ed.), Interscience Publishers, New York, 1963, pp. 247–55, that a superposition of so-called gravitational, electromagnetic, and interactional differential forces can always be invoked here to assert $\sim$A. In particular, he is committed to the following: Suppose A is assumed for a region S (say, on the basis of C) and that a particular metric tensor g_{ik} then results from measurements conducted in S with rods that are presumed to be rigid in virtue of A. Then, says Putnam, we can *always* assert $\sim$A of S by reference to the specified three kinds of forces. And we can do so in such a way that rods corrected for the effects of these forces would yield any desired different metric tensor g'_{ik}. The latter tensor would enable the Duhemian to assert his chosen geometric H of S.

But I have demonstrated elsewhere in detail [Grünbaum, *Geometry and Chronometry in Philosophical Perspective*, Ch. III, Section 9] that the so-called gravitational, electromagnetic and interactional forces which Putnam wishes to invoke in order to assert $\sim$A do *not* qualify *singly* as differential. Hence Putnam's proposed scheme cannot help the Duhemian.

[52] For some pertinent details, see Grünbaum, 'The Falsifiability of a Component of a Theoretical System', pp. 286–87.

[53] It is not obvious that the Duhemian could make this assertion in this context with impunity: even in the presence of only one kind of differential force, there may be rods of one and the same chemical constitution which exhibit the hysteresis behavior of ferromagnetic materials in the sense that their coincidences at a given place will depend on their paths of transport. Cf. Grünbaum, *Geometry and Chronometry*..., p. 359.

[54] Swanson, 'On the D-Thesis', pp. 64–65, maintains that when the Duhemian contests A in my geometric example, he would *not* invoke a crypto-*identity* conjecture, as I have

him do, but rather a conjecture of a crypto-*difference* among prima facie chemically *identical* rods. And that crypto-difference might arise from a kind of undiscovered isomerism. He says:

It is not the assertion of the chemical *difference* of two rods that the D-theorist would claim to be theory-laden. Grünbaum is certainly right in showing that the D-theorist could not argue for *that*. Rather, it is the assertion of their *identity* that would be theory-laden. For suppose that according to a given chemical theory T_1, two rods a and b were found to be chemically identical to one another.... We can now construe T_1 as some sort of pre-'isomeric' chemistry... and then go on to imagine a post-'isomeric' T_2 such that upon the fulfillment of certain tests... it is found that $a \neq b$.... But if identity of a and b is thus theory-laden, it does little good to argue that non-identical rods can be clearly discriminated.

Swanson then considers a model universe of three rods, two of which are prima facie chemically identical though crypto-different, while being pairwise objectively and perceptibly different from the third. And he then introduces two (ghostlike) perturbing influences whose *superposition* imparts the same total deformation to each of the three rods. Thus, in the context of his superposition model, Swanson attaches significance to the fact that the first two of his three rods, a_1 and a_2, are crypto-different.

But nothing in his argument from superposition depends on the hypothesis that a_1 and a_2 are crypto-different instead of identical, nor does his superposition model gain anything from that added hypothesis. Surely the differential character of his two perturbing influences would not be gainsaid by the mere supposition that a_1 and a_2 are chemically identical and that either of the two perturbing influences therefore affects them alike. If there is superposition issuing in the same total deformation, as contrasted with the absence of differential perturbations, then chemically identical rods can exhibit the same total deformation no less than conspicuously different or crypto-different ones! Far from lending added plausibility to the superposition conjecture, Swanson's introduction of the hypothesis of crypto-difference merely compounds the inductive felonies of the two.

In fact, Swanson overlooked the non-sequitur which I did commit apropos of crypto-identity in my 1966 paper. As is clear from the present essay, the Duhemian could couple his denial of A with the superposition conjecture *rather than* with the hypothesis of crypto-identity. It was therefore incorrect on my part to assert in 1966 that, when denying A, the Duhemian 'must' assert crypto-identity (see 'The Falsifiability of a Component of a Theoretical System', p. 287, *item (3)*).

[55] For a lucid explanation of the *non*-transitivity (as distinct from intransitivity) of the relation of inductive support and of its ramifications, see W. C. Salmon, 'Consistency, Transitivity, and Inductive Support', *Ratio* **7** (1965), 164–68.

[56] This claim of direct inductive inferability is made here only with respect to the specifics of this example. And it is certainly not intended to deny the existence of other situations, such as those discussed by R. C. Jeffrey, *The Logic of Decision*, McGraw-Hill, New York, 1965, pp. 155–56, in which a perceptual experience can reasonably prompt us to entertain a variety of relevant propositions.

[57] Hesse, *The British Journal for the Philosophy of Science* **18**, 334.

[58] I am indebted to Philip Quinn and Laurens Laudan for reading a draft of this paper and suggesting some improvements in it. I also wish to thank Wesley Salmon and Allen Janis for some helpful comments and references.

PAUL K. FEYERABEND

THE RATIONALITY OF SCIENCE

(From 'Against Method')

1. Discovery and justification; observation and theory*

Let us now use the material of the [earlier sections of 'Against Method'] to throw light on the following features of contemporary empiricism: first, the distinction between a context of discovery and a context of justification; second, the distinction between observational terms and theoretical terms; third, the problem of incommensurability.

One of the objections which may be raised against the preceding discussion is that it has confounded two contexts which are essentially separate, viz. a context of discovery and a context of justification. *Discovery* may be irrational and need not follow any recognized method. *Justification*, on the other hand, or, to use the Holy Word of a different school, *criticism*, starts only after the discoveries have been made and proceeds in an orderly way. Now, if the example given here and the examples I have used in earlier papers show anything, then they show that the distinction refers to a situation that does not arise in practice at all. And, if it does arise, it reflects a temporary stasis of the process of research. Therefore, it should be eliminated as quickly as possible.

Research at its best is an *interaction* between new theories which are stated in an explicit manner and older views which have crept into the observation language. It is not a one-sided *action* of the one upon the other. Reasoning within the context of justification, however, presupposes that one side of this pair, viz. observation, has frozen, and that the principles which constitute the observation concepts are preferred to the principles of a newly invented point of view. The former feature indicates that the discussion of principles is not carried out as vigorously as is desirable; the latter feature reveals that this lack of vigor may be due to some unreasonable and perhaps not even explicit preference. But is it wise to be dominated by an inarticulate preference of this kind? Is it wise to make it the raison d'être of a distinction that separates two entirely different modes of research? Or should we not rather demand that our methodolo-

gy treat explicit and implicit assertions, doubtful and intuitively evident theories, known and unconsciously held principles, in exactly the same way, and that it provide means for the discovery and the criticism of the latter? Abandoning the distinction between a context of discovery and a context of justification is the first step toward satisfying this demand.

Another distinction which is clearly related to the distinction between discovery and justification is the distinction between *observational terms and theoretical terms*. It is now generally admitted that the distinction is not as sharp as it was thought to be only a few decades ago. It is also admitted, in complete agreement with Neurath's original views, that *both* theories *and* observation statements are open to criticism. Yet the distinction is still held to be a useful one and is defended by almost all philosophers of science. But what is its point? Nobody will deny that the sentences of science can be classified into long sentences and short sentences, or that its statements can be classified into those which are intuitively obvious and others which are not. But nobody will put particular weight on these distinctions, or will even mention them, *for they do not now play any role in the business of science*. (This was not always so. Intuitive plausibility, for example, was once thought to be a most important guide to the truth; but it disappeared from methodology the very moment intuition was replaced by experience.) Does experience play such a role in the business of science? Is it as essential to refer to experience as it was once thought essential to refer to intuition? Considering what has been said earlier, I think that these questions must be answered in the negative. True – much of our thinking *arises* from experience, but there are large portions which do not arise from experience at all but are firmly grounded on intuition, or on even deeper lying reactions. True – we often test our theories by experience, but we equally often *invert* the process; we *analyze* experience with the help of more recent views and we *change* it in accordance with these views (see the preceding discussion of Galileo's procedure). Again, it is true that we often rely on experience in a way that suggests that we have here a solid foundation of knowledge, but such reliance turns out to be just a psychological quirk, as is shown whenever the testimony of an eyewitness or of an expert crumbles under cross-examination. Moreover, we equally firmly rely on general principles so that even our most solid *perceptions* (and not only our *assumptions*) become indistinct and ambiguous when they clash with these prin-

ciples. The symmetry between observation and theory that emerges from such remarks is perfectly reasonable. Experience, just as our theories, contains natural interpretations which are abstract and even metaphysical ideas. For example, it contains the idea of an observer-independent existence. It is incontestable that these abstractions, these speculative ideas, are connected with sensations and perceptions. But, first of all, this does not give them a privileged position, unless we want to assert that perception is an infallible authority. And, secondly, it is quite possible to altogether *eliminate* perception from all the essential activities of science.... All that remains is that some of our ideas are *accompanied* by strong and vivid psychological processes, 'sensations', while others are not. This, however, is just a peculiarity of human existence which is as much in need of examination as is anything else.

Now, if we want to be 'truly scientific' (dreaded words!), should we then not regard the theses "experience is the foundation of our knowledge" and "experience helps us to discover the properties of the external world" as (very general) hypotheses? And must these hypotheses not be examined just like any other hypothesis, *and perhaps even more vigorously*, as so much depends on their truth? Furthermore, will not such an examination be rendered impossible by a method that either justifies or criticizes "on the basis of experience"? These are some of the questions which arise in connection with the customary distinctions between observation and theory, discovery and justification. None of them is really new. They are known to philosophers of science, and are discussed by them at length. But the inference that the distinction between theory and observation has now ceased to be relevant either is not drawn or is explicitly rejected.[1] Let us take a step forward, and let us abandon this last remainder of dogmatism in science!

1. Rationality again

Incommensurability, which I shall discuss next, is closely connected with the question of the rationality of science. Indeed, one of the most general objections, either against the *use* of incommensurable theories or even against the idea that *there are* such theories to be found in the history of science, is the fear that they would severely restrict the efficacy of traditional, nondialectical *argument*. Let us, therefore, look a little more closely at the critical *standards* which, according to some people, consti-

tute the content of a 'rational' argument. More especially, let us look at the standards of the Popperian school with whose ratiomania we are here mainly concerned.

Critical rationalism is either a meaningful idea or a collection of slogans (such as 'truth'; 'professional integrity'; 'intellectual honesty') designed to intimidate yellow-bellied opponents (who has the fortitude, or even the insight, to declare that Truth might be unimportant, and perhaps even undesirable?).

In the former case it must be possible to produce rules, standards, restrictions which permit us to separate critical behavior (thinking, singing, writing of plays) from other types of behavior so that we can *discover* irrational actions and *correct* them with the help of concrete suggestions. It is not difficult to produce the standards of rationality defended by the Popperian school.

These standards are standards of *criticism*: rational discussion consists in the attempt to criticize, and not in the attempt to prove, or to make probable. Every step that protects a view from criticism, that makes it safe, or 'well founded', is a step away from rationality. Every step that makes it more vulnerable is welcome. In addition it is recommended that ideas which have been found wanting be abandoned, and it is forbidden to retain them in the face of strong and successful criticism unless one can present a suitable counterargument. Develop your ideas so that they can be criticized; attack them relentlessly; do not try to protect them, but exhibit their weak spots; and eliminate them as soon as such weak spots have become manifest – these are some of the rules put forth by our critical rationalists.

These rules become more definite and more detailed when we turn to the philosophy of science, and especially to the philosophy of the natural sciences.

Within the natural sciences criticism is connected with experiment and observation. The content of a theory consists in the sum total of those basic statements which contradict it; it is the class of its potential falsifiers. Increased content means increased vulnerability; hence theories of large content are to be preferred to theories of small content. Increase of content is welcome; decrease of content is to be avoided. A theory that contradicts an accepted basic statement must be given up. Ad hoc hypotheses are forbidden – and so on and so forth. A science, however, that

accepts the rules of a critical empiricism of this kind will develop in the following manner.

We start with a *problem* such as the problem of the planets at the time of Plato. This problem is not merely the result of *curiosity*, it is a *theoretical result*, it is due to the fact that certain *expectations* have been disappointed: On the one hand it seemed to be clear that the stars must be divine; hence one expects them to behave in an orderly and lawful manner. On the other hand one cannot find any easily discernible regularity. The planets, to all intents and purposes, move in a quite chaotic fashion. How can this fact be reconciled with the expectation and with the principles that underlie the expectation? Does it show that the expectation is mistaken? Or have we failed in our analysis of the facts? This is the problem.

It is important to see that the elements of the problem are not simply *given*. The 'fact' of irregularity, for example, is not accessible without further ado. It cannot be discovered by just anyone who has healthy eyes and a good mind. It is only through a certain expectation that it becomes an object of our attention. Or, to be more accurate: this fact of irregularity *exists* because there is an expectation of regularity. After all, the term 'irregularity' makes sense only if we have a rule. In our case the rule (which is a more specific part of the expectation that has not yet been mentioned) asserts circular motion with constant angular velocity. The fixed stars agree with this rule and so does the sun if we trace its path relative to the fixed stars. The planets do not obey the rule, neither directly, with respect to the earth, nor indirectly, with respect to the fixed stars.

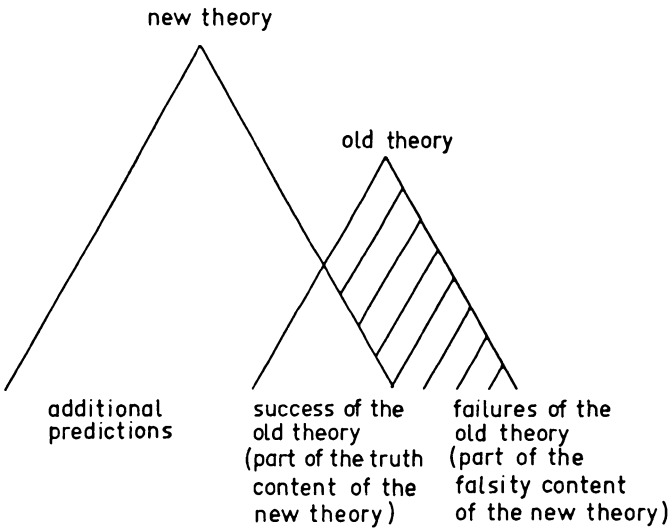
(In the case just discussed the rule is formulated explicitly, and it can be discussed. This need not be the case. Recognizing a color as red is made possible by deep-lying assumptions concerning the structure of our surroundings and recognition does not occur when these assumptions cease to be available.)

To sum up this part of the Popperian doctrine: Research starts with a problem. The problem is the result of a conflict between an expectation and an observation which in turn is constituted by the expectation. It is clear that this doctrine differs from the doctrine of inductivism where objective facts mysteriously enter a passive mind and leave their traces there. It was prepared by Kant, by Dingler, and, in a very different manner, by Hume.

Having formulated a problem one tries to *solve* it. Solving a problem

means inventing a theory that is relevant, falsifiable (to a larger degree than any alternative solution), but not yet falsified. In the case mentioned above (planets at the time of Plato) the problem was to find circular motions of constant angular velocity for the purpose of saving the planetary phenomena. It was solved by Eudoxos.

Next comes the *criticism* of the theory that has been put forth in the attempt to solve the problem. Successful criticism removes the theory *once and for all* and creates a new problem, viz. to explain (a) why the theory has been successful so far; (b) why it failed. Trying to solve *this* problem we need a new theory that produces the successful consequences of the older theory, denies its mistakes, and makes additional predictions



not made before. These are some of the *formal conditions* which *a suitable successor of a refuted theory* must satisfy. Adopting the conditions one proceeds, by conjectures and refutations, from less general theories to more general theories and expands the content of human knowledge. More and more facts are *discovered* (or constructed with the help of expectations) and are then *connected* in a reasonable manner. There is no guarantee that man will solve every problem and replace every theory that has been refuted with a successor satisfying the formal conditions.

The invention of theories depends on our talents and other fortuitous circumstances, such as a satisfactory sex life. But as long as these talents hold out the accompanying scheme is a correct account of the growth of a knowledge that satisfies the rules of critical rationalism.

Now, at this point we may raise two questions:

1. Is it *desirable* to live in accordance with the rules of a critical rationalism?

2. Is it *possible* to have both a science as we know it and these rules?

As far as I am concerned the first question is far more important than the second. True – science and other depressing and narrow-minded institutions play an important part in our culture and they occupy the center of interest of most philosophers. Thus the ideas of the Popperian school were obtained by generalizing solutions for methodological and epistemological problems. Critical rationalism arose from the attempt to solve Hume's problem and to understand the Einsteinian revolution, and it was then extended to politics, and even to the conduct of one's private life (Habermas and others therefore seem to be justified in calling Popper a positivist). Such a procedure may satisfy a *school philosopher* who looks at life through the spectacles of his own specific problems and recognizes hatred, love, happiness only to the extent to which they occur in these problems. But if we consider the interests of *man* and, above all, the question of his freedom (freedom from hunger, despair, from the tyranny of constipated systems of thought, *not* the academic 'freedom of the will'), then we are proceeding in the worst possible fashion.

For is it not possible that science as we know it today (the science of critical rationalism that has been freed from all inductive elements) or a 'search for the truth' in the style of traditional philosophy will create a monster? Is it not possible that it will harm man, turn him into a miserable, unfriendly, self-righteous mechanism without charm and without humor? "Is it not possible", asks Kierkegaard, "that my activity as an objective [or a critico-rational] observer of nature will weaken my strength as a human being?"[2] I suspect the answer to all these questions must be affirmative and I believe that a reform of the sciences that makes it more anarchistic and more subjective (in Kierkegaard's sense) is therefore urgently needed. But this is not what I want to discuss in the present essay. Here I shall restrict myself to the second question and I shall ask: is it possible to have both a science as we know it and the rules of a criti-

cal rationalism as just described? And to *this* question the answer seems to be a resounding *no*.

To start with we have seen, though rather briefly, that the actual development of institutions, ideas, practices, and so on often does not start from a problem but rather from some irrelevant activity, such as playing, which, as a side effect, leads to developments which later on can be interpreted as solutions to unrealized problems. Are such developments to be excluded? And if we *do* exclude them, will this not considerably reduce the number of our adaptive reactions and the quality of our learning process?

Secondly, we have seen that a strict principle of falsification, or a 'naive falsificationism' as Imre Lakatos calls it, combined with the demand for maximum testability and non-adhocness would wipe out science as we know it, and would never have permitted it to start. This has been realized by Lakatos who has set out to remedy the situation.[3] His remedy is not mine, it is not anarchism. His remedy consists in slight modification of the 'critical standards' he adores. (He also tries to show, with the help of amusing numerological considerations, that it is already foreshadowed in Popper.)

According to naive falsificationism, a theory is judged, i.e., either accepted or condemned, as soon as it is introduced into the discussion. Lakatos gives a theory time, he permits it to develop, to show its hidden strength, and he judges it only 'in the long run'. The 'critical standards' *he* employs provide for an interval of hesitation. They are applied 'with hindsight'. If the theory gives rise to interesting new developments, if it engenders 'progressive problem shifts', then it may be retained despite its initial vices. If on the other hand the theory leads nowhere, if the ad hoc hypotheses it employs are not the starting point but the end of all research, if the theory seems to kill the imagination and to dry up every resource of speculation, if it creates 'degenerating problem shifts', i.e., changes which terminate in a dead end, then it is time to give it up and to look for something better.

Now it is easily seen that standards of this kind have practical force only if they are combined with a *time limit*. What looks like a degenerating problem shift may be the beginning of a much longer period of advance, so – how long are we supposed to wait? But if a time limit *is* introduced, then the argument against the more conservative point of view, against

'naive falsificationism', reappears with only a minor modification. For if you can wait, then why not wait a little longer? Besides there are theories which for centuries were accompanied by degenerating problem shifts until they found the right defenders and returned to the stage in full bloom. The heliocentric theory is one example. The atomic theory is another. We see that the new standards which Lakatos wants to defend either are *vacuous* – one does not know when and how to apply them – or else can be *criticized* on grounds very similar to those which led to them in the first place.

In these circumstances one can do one of the following two things. One can stop appealing to permanent standards which remain in force throughout history, and govern every single period of scientific development and every transition from one period to another. Or one can retain such standards as a *verbal ornament*, as a memorial to happier times when it was still thought possible to run a complex and catastrophic business like science by a few simple and 'rational' rules. It seems that Lakatos wants to choose the second alternative.

Choosing the second alternative means abandoning permanent standards *in fact*, though retaining them *in words*. *In fact* Lakatos's position now is identical with the position of Popper as summarized in the marvelous (because self-destructive) Appendix i/15 of the fifth edition of the *Open Society*.[4] According to Popper, we do not "need any... definite frame of reference for our criticism", we may revise even the most fundamental rules and drop the most fundamental demands if the need for a different measure of excellence should arise.[5] Is such a position irrational? Yes and no. Yes, because there no longer exists a single set of rules that will guide us through all the twists and turns of the history of thought (science), either as participants or as historians who want to reconstruct its course. One can of course *force* history into a pattern, but the results will always be poorer and less interesting than were the actual events. No, because each particular episode is rational in the sense that some if its features can be explained in terms of reasons which were either accepted at the time of its occurrence or invented in the course of its development. Yes, because even these local reasons which change from age to age are never sufficient to explain *all* the important features of a particular episode. One must add accidents, prejudices, material conditions, e.g., the existence of a particular type of glass in one country and

not in another for the explanation of the history of optics, the vicissitudes of married life (Ohm!), superficiality, pride, oversight, and many other things, in order to get a complete picture. No, because, transported into the climate of the period under consideration and endowed with a lively and curious intelligence, we might have had still more to say; we might have tried to overcome accidents, and to 'rationalize' even the most whimsical sequence of events. But, and now I come to a decisive point for the discussion of incommensurability, how is the transition from certain standards to other standards to be achieved? More especially, what happens to our standards, as opposed to our theories, during a period of revolution? Are they changed in the manner suggested by Mill by a critical discussion of alternatives, or are there processes which defy a rational analysis? Well, let us see!

That standards are not always adopted on the basis of argument has been emphasized by Popper himself. Children, he says, "learn to imitate others... and so learn to look upon standards of behavior as if they consisted of fixed, 'given' rules... and such things as sympathy and imagination may play an important role in this development".[6] Similar considerations apply to those grownups who want to continue learning, and who are intent on expanding both their knowledge and their sensibility. This we have already discussed in Section 1. Popper also admits that new standards may be discovered, invented, accepted, imparted to others in a very irrational manner. But, he points out, one can criticize them *after* they have been adopted, and it is *this* possibility which keeps our knowledge rational. "What, then, are we to trust?" he asks after a survey of possible sources for standards.[7]

> What are we to accept? The answer is: whatever we accept we should trust only tentatively, always remembering that we are in possession, at best, of partial truth (or rightness), and that we are bound to make at least some mistake or misjudgement somewhere – not only with respect to facts but also with respect to the adopted standards; secondly, we should trust (even tentatively) our intuition only if it has been arrived at as the result of many attempts to use our imagination; of many mistakes, of many tests, of many doubts, *and of searching criticism.*

Now this reference to tests and to criticism, which is supposed to guarantee the rationality of science, and, perhaps, of our entire life, may be either to *well-defined procedures* without which a criticism or test cannot be said to have taken place, or to a purely *abstract* notion, so that it is

left *to us* to fill it now with this, and now with that concrete content. The first case has just been discussed. In the second case we have again but a verbal ornament. The questions asked in the last paragraph but one remain unanswered in either case.

In a way even this situation has been described by Popper, wo says that "rationalism is necessarily far from comprehensive or self-contained."[8] But our present inquiry is not whether *there are* limits to our reason; the question is *where* these limits are *situated.* Are they outside the sciences so that science itself remains entirely rational; or are irrational changes an essential part even of the most rational enterprise that has been invented by man? Does the historical phenomenon 'science' contain ingredients which defy a rational analysis, although they may be described with complete clarity in psychological or sociological terms? Can the abstract aim to come closer to the truth be reached in an entirely rational manner, or is it perhaps inaccessible to those who decide to rely on argument only? These are the problems which were raised, first by Hegel and then, in quite different terms, by Kuhn. They are the problems I wish to discuss.

In discussing these further problems, Popper and Lakatos reject considerations of sociology and psychology, or as Lakatos expresses himself, 'mob psychology', and assert the rational character of *all* science. According to Popper, it is possible to arrive at a judgment as to which of two theories is closer to the truth, even if the theories should be separated by a catastrophic upheaval such as a scientific or other revolution. (A theory is closer to the truth than another theory if the class of its true consequences, its truth content, exceeds the truth content of the latter without an increase of falsity content.) According to Lakatos, the apparently unreasonable features of science occur only in the material world and in the world of (psychological) thought; they are absent from the "world of ideas, from Plato's and Popper's 'third world'". It is in this third world that the growth of knowledge takes place, and that a rational judgment of all aspects of science becomes possible.

Now in regard to this convenient flight into higher regions, it must be pointed out that the scientist is, unfortunately, dealing with the world of matter and of psychological (i.e., subjective) thought also. It is *mainly* this material world he wants to change and to influence. And the rules which create order in the third world will most likely be entirely inappropriate for creating order in the brains of living human beings (unless these

brains and their structural features are put in the third world also, a point that does not become clear from Popper's account[9]). The numerous deviations from the straight and rather boring path of rationality which one can observe in actual science may well be *necessary* if we want to achieve progress with the brittle and unreliable material (instruments; brains; assistants; etc.) at our disposal.

However, there is no need to pursue this objection further. There is no need to argue that science as we know it may differ from its third-world shadow *in precisely those respects* which make progress possible.[10] For the Popperian model of an approach to the truth breaks down even if we confine ourselves to ideas entirely. It breaks down because there are incommensurable theories.

2. Incommensurability

Scientific investigation, says Popper, *starts* with a problem, and it proceeds by *solving* it.

This characterization does not take into account that problems may be wrongly formulated, that one may inquire about properties of things or processes which later research declares to be nonexistent. Problems of this kind are not *solved*, they are *dissolved* and removed from the domain of legitimate inquiry. Examples are the problem of the absolute velocity of the earth, the problem of the trajectory of an electron in an interference pattern, or the important problem whether incubi are capable of producing offspring or whether they are forced to use the seeds of men for that purpose.[11]

The first problem was dissolved by the theory of relativity which denies the existence of absolute velocities. The second problem was dissolved by the quantum theory which denies the existence of trajectories in interference patterns. The third problem was dissolved, though much less decisively so, by modern (i.e., post-sixteenth century) psychology and physiology as well as by the mechanistic cosmology of Descartes.

Now changes of ontology such as those just described are often accompanied by conceptual changes.

The discovery that certain entities do not exist may force the scientist to redescribe the events, processes, observations which were thought to be manifestations of them and were therefore described in terms assuming

their existence. Or, rather, it may force him to use new *concepts* as the older *words* will remain in use for a considerable time. Thus the term 'possessed' which was once used for giving a causal description of the behavioral peculiarities connected with epilepsy was retained, but it was voided of its devilish connotations.

An interesting development occurs when the faulty ontology is *comprehensive*, that is, when its elements are thought to be present in every process in a certain domain. In *this* case *every* description inside the domain must be changed and must be replaced by a different statement (or by no statement at all). Classical physics is a case in point. It has developed a comprehensive terminology for describing the most fundamental mechanical properties of our universe, such as shapes, speeds, and masses. The conceptual system connected with this terminology assumes that the properties *inhere* in objects and that they change only if one interferes with the objects, not otherwise. The theory of relativity teaches us, at least in one of its interpretations, that there are no such inherent properties in the world, neither observable, nor unobservable, and it produces an entirely new conceptual system for description inside the domain of mechanics. This new conceptual system does not just *deny* the existence of classical states of affairs, it does not even permit us to *formulate statements* expressing such states of affairs (there is no arrangement in the Minkowski diagram that corresponds to a classical situation). It does not, and cannot, share a single statement with its predecessor. As a result the formal conditions for a suitable successor of a refuted theory (it has to repeat the successful consequences of the older theory, deny its false consequences, and make additional predictions) cannot be satisfied in the case of relativity versus classical physics and the Popperian scheme of progress breaks down. It is not even possible to connect classical statements and relativistic statements by an *empirical hypothesis*.[12] Formulating such a connection would mean formulating statements of the type "whenever there is possession by a demon there is discharge in the brain" which perpetuate rather than eliminate the older ontology. Comprehensive theories of the kind just mentioned are therefore completely disjointed, or *incommensurable*. The existence of incommensurable theories provides another difficulty for critical rationalism (and, a fortiori, for its more positivistic predecessors). We shall discuss this difficulty by discussing and refuting objections against it.

It was pointed out that progress may lead to a complete replacement of statements (and perhaps even of descriptions) in a certain domain. More especially, it may replace certain natural interpretations by others.... Galileo replaces the idea of the operative character of all motion by his relativity principle in order to accommodate the new views of Copernicus. It is entirely natural to proceed in this way. A cosmological theory such as the heliocentric theory, or the theory of relativity, or the quantum theory (though the last one only with certain restrictions) makes assertions about the world as a whole. It applies to observed and to unobserved (unobservable, 'theoretical') processes. It can therefore demand to be used always, and not only on the theoretical level. Now such an adaption of observation to theory, and this is the gist of the *first objection*, removes conflicting observation reports and saves the theory in an ad hoc manner. Moreover, there arises the *suspicion* that observations which are interpreted in terms of a new theory can no longer be used to refute that theory. It is not difficult to reply to these points.

As regards the objection we point out, in agreement with what has been said before..., that an inconsistency between theory and observation may reveal a fault of our *observational terminology* (and even of our sensations) so that it is quite natural to change this terminology, to adapt it to the new theory, and to see what happens. Such a change gives rise, and should give rise, to new auxiliary subjects (hydrodynamics, theory of solid objects, optics in the case of Galileo) which may more than compensate for the empirical content lost by the adaptation. And as regards the suspicion we must remember that the predictions of a theory depend on its postulates, the associated grammatical rules, *as well as* on initial conditions while the meaning of the 'primitive' notions depends on the postulates (and the associated grammatical rules) only.[13] In those rare cases, however, where a theory *entails* assertions about possible initial conditions,[14] we can refute it with the help of *self-inconsistent observation reports* such as "object A does not move on a geodesic" which, if analyzed in accordance with the Einstein-Infeld-Hoffmann account reads "singularity α which moves on a geodesic does not move on a geodesic".

The *second objection* criticizes the interpretation of science that brings about incommensurability. To deal with it we must realize that the question "are two particular comprehensive theories, such as classical celestial mechanics (CM) and the special theory of relativity (SR) incommen-

surable?" is not a complete question. Theories can be interpreted in different ways. They will be commensurable in some interpretations, incommensurable in others. Instrumentalism, for example, makes commensurable all those theories which are related to the same observation language and are interpreted on its basis. A realist, on the other hand, wants to give a unified account, both of observable and of unobservable matters, and he will use the most abstract terms of whatever theory he is contemplating for that purpose.[15] This is an entirely natural procedure. SR, so one would be inclined to say, does not just invite us to rethink *unobserved* length, mass, duration; it would seem to entail the relational character of all lengths, masses, durations, whether observed or unobserved, observable or unobservable.

Now, and here we only repeat what was said not so long ago, extending the concepts of a new theory, T, to all its consequences, observational reports included, may change the interpretation of these consequences to such an extent that they disappear from the consequence classes either of earlier theories or of the available alternatives. These earlier theories and alternatives will then all become incommensurable with T. The relation between SR and CM is a case in point. The concept of length as used in SR and the concept of length as presupposed in CM are different concepts. Both are *relational* concepts, and very complex relational concepts at that (just consider determination of length in terms of the wave length of a specified spectral line). But relativistic length, or relativistic shape, involves an element that is absent from the classical concept and is in principle excluded from it. It involves the relative velocity of the object concerned in some reference system. It is of course true that the relativistic scheme very often yields numbers which are practically identical with the *numbers* obtained from CM, but this does not make the *concepts* more similar. Even the case $c \rightarrow \infty$ (or $v \rightarrow O$) which yields identical predictions cannot be used as an argument for showing that the concepts must coincide, at least in this special case. Different magnitudes based on different concepts may give identical values on their respective scales without ceasing to be different magnitudes. The same remark applies to the attempt to identify classical mass with relativistic *rest* mass.[16] This conceptual disparity, if taken seriously, infects even the most 'ordinary' situations. The relativistic concept of a certain *shape*, such as the shape of a table, or of a certain temporal sequence, such as my saying 'Yes', will

differ from the corresponding classical concept also. It is therefore futile to expect that sufficiently long derivations may eventually return us to the older ideas.[17] The consequence classes of SR and CM are not related in any way. A comparison of content and a judgment of verisimilitude cannot be made.[18]

The situation becomes even clearer when we use the Marzke-Wheeler interpretation of SR. For it can be easily shown that the methods of measurement provided by these authors, while perfectly adequate in a relativistic universe, either collapse or give nonsensical results in a classical world (length, for example, is no longer transitive, and in some coordinate systems it may be impossible to assign a definite length to *any* object).

We are now ready to discuss the *second* and most popular objection against incommensurability. This objection proceeds from the version of realism described above. "A realist", we said, "will want to give a unified account, both of observable and of unobservable matters, and he will use the most abstract terms of whatever theory he is contemplating for his purpose". He will use such terms in order either to *give* meaning to observation sentences or else to *replace* their customary interpretation. (For example, he will use the ideas of SR in order to replace the customary CM-interpretation of everyday statements about shapes, temporal sequences, and so on.) Against this, it is pointed out that theoretical terms receive their interpretation by being connected with a preexisting observation language, or with another theory that has already been connected with such an observation language, and that they are devoid of content without such connection. Thus Carnap asserts[19] that

> [t]here is no independent interpretation for L_T [the language in terms of which a certain theory, or a certain world view, is formulated]. The system T [the axioms of the theory and the rules of derivation] is in itself an uninterpreted postulate system. [Its] terms ... obtain only an indirect and incomplete interpretation by the fact that some of them are connected by the [correspondence] rules C with observation terms....

Now, if theoretical terms have no "independent interpretation", then surely they cannot be used for correcting the interpretation of the observation statements, which is the one and only source of their meaning. It follows that realism as described here is an impossible doctrine.

The guiding idea behind this very popular objection is that new and abstract languages cannot be introduced in a direct way, but must be first

connected with an already existing, and presumably stable, observational idiom.[20]

This guiding idea is refuted at once by noting the way in which children learn to speak and in which anthropologists and linguists learn the unknown language of a newly discovered tribe.

The first example is instructive for other reasons also, for incommensurability plays an important role in the early months of human development. As has been suggested by Piaget and his school[21] the child's perception develops through various stages before it reaches its relatively stable adult form. In one stage objects seem to behave very much like afterimages,[22] and they are treated as such. In this stage the child follows the object with his eyes until it disappears, and he does not make the slightest attempt to recover it, even if this would require but a minimal physical (or intellectual) effort, an effort, moreover, that is already within the child's reach. There is not even a tendency to search; and this is quite appropriate, 'conceptually' speaking. For it would indeed be nonsensical to 'look for' an afterimage. Its 'concept' does not provide for such an operation.

The arrival of the concept and of the perceptual image of material objects changes the situation quite dramatically. There occurs a drastic reorientation of behavioral patterns, and, so one may conjecture, of thought. Afterimages, or things somewhat like them, still exist, but they are now difficult to find and must be discovered by special methods. (The earlier visual world therefore *literally disappears*.) Such special methods proceed from a new conceptual scheme (afterimages occur in *humans*, not in the outer physical world, and are tied to them) and cannot lead back to the exact phenomena of the previous stage (these phenomena should therefore be called by a different name, such as 'pseudo-afterimages'). Neither afterimages nor pseudo-afterimages are given a special position in the new world. For example, they are not treated as 'evidence' on which the new notion of a material object is supposed to rest. Nor can they be used to *explain* this notion: afterimages arise *togeher with it*, and are absent from the minds of those who do not yet recognize material objects. And pseudo-afterimages *disappear* as soon as such recognition takes place. It is to be admitted that every stage possesses a kind of observational 'basis' to which one pays special attention and from which one receives a multitude of suggestions. However, this basis (i) *changes* from stage to stage;

and (ii) is *part* of the conceptual apparatus of a given stage; it is *not* its one and only source of interpretation.

Considering developments such as these, one may suspect that the family of concepts centering upon 'material object' and the family of concepts centering upon 'pseudo-afterimage' are incommensurable in precisely the sense that is at issue here. Is it reasonable to expect that conceptual and perceptual changes of this kind occur in childhood only? Should we welcome the fact, if it is a fact, that an adult is stuck with a stable perceptual world and an accompanying stable conceptual system which he can modify in many ways, but whose general outlines have forever become immobilized? Or is it not more realistic to assume that fundamental changes, entailing incommensurability, are still possible, and that they should be encouraged lest we remain forever excluded from what might be a higher stage of knowledge and of consciousness?... Besides, the question of the mobility of the adult stage is at any rate an empirical question, which must be attacked by *research* and which cannot be settled by methodological *fiat*. The attempt to break through the boundaries of a given conceptual system and to escape the reach of 'Popperian spectacles' (Lakatos) is an essential part of such research (and should be an essential part of any interesting life).[23]

Looking now at the second element of the refutation, anthropological field work, we see that what is anathema here (and for very good reasons) is still a fundamental principle for the contemporary representatives of the philosophy of the Vienna Circle. According to Carnap, Feigl, Nagel, and others, the terms of a theory receive their interpretation in an indirect fashion, by being related to a different conceptual system which is either an older theory or an observation language.[24] This older theory, this observation language, is not adopted because of its theoretical excellence. It cannot possibly be: the older theories are usually refuted. It is adopted because it is "used by a certain language community as a means of communication".[25] According to this method, the phrase 'having much larger relativistic mass than...' is partially *interpreted* by first connecting it with some *prerelativistic* terms (classical terms, common-sense terms), which are 'commonly understood' (presumably, as the result of previous teaching in connection with crude weighing methods), and it is *used* only after such connection has given it a well-defined meaning.

This is even worse than the once quite popular demand to clarify doubt-

ful points by translating them into Latin. For while Latin was chosen because of its precision and clarity, and also because it was conceptually richer than the slowly evolving vulgar idioms,[26] the choice of an observation language or of an older theory as a basis for interpretation is justified by saying that they are 'antecedently understood': the choice is based on sheer *popularity*. Besides, if *prerelativistic* terms which are pretty far removed from reality (especially in view of the fact that they come from an incorrect theory implying a nonexistent ontology) can be taught ostensively, for example, with the help of crude weighing methods (and one must assume that they can be so taught, or the whole scheme collapses), then why should one not introduce the *relativistic* terms *directly*, and *without* assistance from the terms of some other idiom? Finally, it is but plain common sense that the teaching or the learning of new and unknown languages must not be contaminated by external material. Linguists remind us that a perfect translation is never possible, even if one is prepared to use complex contextual definitions. This is one of the reasons for the importance of *field work* where new languages are learned *from scratch*, and for the rejection, as inadequate, of any account that relies on 'complete' or 'partial' translation. *Yet just what is anathema in linguistics is taken for granted by logical empiricism*, a mythical 'observation language' replacing the English of the translators. Let us commence field work in this domain also, and let us study the language of new theories not in the definition factories of the double language model, but in the company of those metaphysicians, theoreticians, playwrights, courtesans who have constructed new world views! This finishes my discussion of the guiding principle behind the second objection against realism and the possibility of incommensurable theories.

Another point that is often made is that there exist *crucial experiments* which refute one of two allegedly incommensurable theories and confirm the other (example: the Michelson-Morley experiment, the variation of the mass of elementary particles, the transverse Doppler effect, are said to refute CM and confirm SR). The answer to this problem is not difficult either: adopting the point of view of relativity, we find that the experiments, *which of course will now be described in relativistic terms*, using the relativistic notions of length, duration, speed, and so on, are relevant to the theory. And we also find that they support the theory. Adopting CM (with, or without an ether), we again find that the experiments, which are

now described in the very different terms of classical physics, i.e., roughly in the manner in which Lorentz described them, are relevant. But we also find that they *undermine* CM, i.e., the conjunction of classical electrodynamics and of CM. Why should it be necessary to possess terminology that allows one to say that it is the *same* experiment which confirms one theory and refutes the other? But did we not ourselves use such terminology? Well, for one thing it should be easy though somewhat laborious to express what was just said *without* asserting identity. Secondly, the identification is of course not contrary to our thesis, for we are now not *using* the terms of either relativity or classical physics, as is done in a test, but are *referring* to them and their relation to the physical world. The language in which *this* discourse is carried out can be classical, or relativistic, or ordinary. It is no good insisting that scientists act as if the situation were much less complicated. If they act that way, then they are either instrumentalists (see above) or mistaken (many scientists are nowadays interested in *formulas*, while the subject here is *interpretations*). It is also possible that being well acquainted with both CM and SR, they change back and forth between these theories with such speed that they seem to remain within a single domain of discourse.

It is also said that by admitting incommensurability into science we can no longer decide whether a new view explains what it is supposed to explain, or whether it does not wander off into different fields.[27] For example, we would not know whether a newly invented physical theory is still dealing with problems of space and time or whether its author has not by mistake made a biological assertion. But there is no need to possess such knowledge. For once the fact of incommensurability has been admitted, the question which underlies the objection does not arise. Conceptual progress often makes it impossible to ask certain questions and to explain certain things; thus we can no longer ask for the absolute velocity of an object, at least as long as we take relativity seriously. Is this a serious loss for science? Not at all! Progress was made by the very same 'wandering off into different fields' whose undecidability now so greatly exercises the critic: Aristotle saw the world as a super *organism*, as a *biological* entity, while one essential element of the new science of Descartes, Galileo, and their followers in medicine and in biology is its exclusively *mechanistic* outlook. Are such developments to be forbidden? And if they are not, what, then, is left of the complaint?

A closely connected objection starts from the notion of *explanation* or *reduction* and emphasizes that this notion presupposes continuity of concepts; other notions could be used for starting exactly the same kind of argument. (Relativity is supposed to explain the valid parts of classical physics; hence it cannot be incommensurable with it!) The reply is again obvious. As a matter of fact it is a triviality for anyone who has only the slightest acquaintance with the Hegelian philosophy: why should the relativist be concerned with the fate of classical mechanics except as part of a historical exercise? There is only *one* task we can legitimately demand of a theory, and it is that it should give us a correct account of the world, i.e., of the totality of facts *as seen through its own concepts.* What have the principles of explanation got to do with this demand? Is it not reasonable to assume that a point of view such as the point of view of classical mechanics that has been found wanting in various respects, that gets in difficulty *with its own facts* (see above, on crucial experiments), and must therefore be regarded as self-inconsistent (another application of Hegelian principles!), cannot have entirely adequate concepts? Is it not equally reasonable to try replacing its concepts with those of a more promising cosmology? Besides, why should the notion of explanation be burdened by the demand for conceptual continuity? This notion has been found to be too narrow before (demand of derivability), and it had to be widened so as to include partial and statistical connections. Nothing prevents us from widening it still further and admitting, say, 'explanations by equivocation'.

Incommensurable theories, then, can be *refuted* by reference to their own respective kinds of experience, i.e., by discovering the *internal contradictions* from which they are suffering (in the absence of commensurable alternatives these refutations are quite weak, however[28]). Their *content* cannot be compared, nor is it possible to make a judgment of *verisimilitude* except within the confines of a particular theory. None of the methods which Popper (or Carnap, or Hempel, or Nagel) want to use for rationalizing science can be applied, and the one that *can* be applied, refutation, is greatly reduced in strength. What remains are esthetic judgments, judgments of taste, and our own subjective wishes.[29] Does this mean that we are ending up in subjectivism? Does this mean that science has become arbitrary, that it has become an element of the general relativism which so much exercises the conscience of some philosophers? Well, let us see.

3. THE CHOICE BETWEEN COMPREHENSIVE IDEOLOGIES

To start with, it seems to me that an enterprise whose human character can be seen by all is preferable to one that looks 'objective' and impervious to human actions and wishes.[30] The sciences, after all, are our own creation, including all the severe standards they seem to impose on us. It is good to be constantly reminded of this fact. It is good to be constantly reminded of the fact that science as we know it today is not inescapable, and that we can construct a world in which it plays no role whatever. (Such a world, I venture to suggest, would be more pleasant to behold than the world we live in today, both materially and intellectually.) What better reminder is there than the realization that the choice between theories which are sufficiently general to yield a comprehensive world view and which are empirically disconnected may become a matter of taste? *That the choice of a basic cosmology may become a matter of taste*?

Secondly, matters of taste are not completely beyond the reach of argument. Poems, for example, can be compared in grammar, sound structure, imagery, rhythm, and can be evaluated on such a basis (cf. Ezra Pound on progress in poetry[31]). Even the most elusive mood can be analyzed *and should be analyzed* if the purpose is to present it in a manner that either can be enjoyed or increases the emotional, cognitive, perceptual, etc., inventory of the reader. Every poet who is worth his salt compares, improves, argues until he finds the correct formulation of what he wants to say.[32] Would it not be marvelous if this free and entertaining[33] process played a role in the sciences also?

Finally, there are more pedestrian ways of explaining the same matter which may be somewhat less repulsive to the tender ears of a professional philosopher of science. One may consider the *length* of derivations leading from the principles of a theory to its observation language, and one may also draw attention to the number of *approximations* made in the course of the derivation. All derivations must be standardized for this purpose so that unambiguous judgments of length can be made. (This standardization concerns the form of the derivation, it does not concern the *content*.) Smaller length and smaller number of approximations would seem to be preferable. It is not easy to see how this requirement can be made compatible with the demand for simplicity and generality which, so it seems, would tend to increase both parameters. However that may

be, there are many ways open to us once the fact of incommensurability is understood, and taken seriously.

4. Conclusion

The idea that science can and should be run according to some fixed rules, and that its rationality consists in agreement with such rules, is both unrealistic and vicious. It is *unrealistic*, since it takes too simple a view of the talents of men and of the circumstances which encourage, or cause, their development. And it is *vicious*, since the attempt to enforce the rules will undoubtedly erect barriers to what men might have been, and will reduce our humanity by increasing our professional qualifications. We can *free* ourselves from the idea and from the power it may possess over us (i) by a detailed study of the work of revolutionaries such as Galileo, Luther, Marx, or Lenin; (ii) by some acquaintance with the Hegelian philosophy and with the alternative provided by Kierkegaard; (iii) by remembering that the existing separation between the sciences and the arts is artificial, that it is a side effect of an idea of professionalism one should eliminate, that a poem or a play can be intelligent as well as informative (Aristophanes, Hochhuth, Brecht), and a scientific theory pleasant to behold (Galileo, Dirac), and that we can change science and make it agree with our wishes. We can turn science from a stern and demanding mistress into an attractive and yielding courtesan who tries to anticipate every wish of her lover. Of course, it is up to us to choose either a dragon or a pussycat as our companion. So far mankind seems to have preferred the latter alternative: "The more solid, well defined, and splendid the edifice erected by the understanding, the more restless the urge of life... to escape from it into freedom". We must take care that we do not lose our ability to make such a choice.

NOTES

* Sections 11 through 15 of 'Against Method: Outline of an Anarchistic Theory of Knowledge', by Paul K. Feyerabend. From Volume IV, *Minnesota Studies in the Philosophy of Science*, University of Minnesota Press, Minneapolis. © Copyright 1970 by the University of Minnesota. Reprinted by permission.

[1] "Neurath fails to give... rules [which distinguish empirical statements from others] and thus unwittingly throws empiricism overboard." K. R. Popper, *The Logic of Scientific Discovery*, Basic Books, New York, 1959, p. 97.

[2] *Papirer*, ed. by P. A. Heiberg (Copenhagen, 1909), VII, Part I, see A, No. 182. Cf. also Sections 7ff of my forthcoming paper 'Abriss einer anarchistischen Erkenntnislehre'.
[3] 'Criticism and the Methodology of Scientific Research Programs', in *Criticism and the Growth of Knowledge*, ed. by I. Lakatos and A. Musgrave, North-Holland, Amsterdam, 1969. Quotations are from the typescript of the paper which Lakatos distributed liberally before its publication. In this typescript the reference is mostly to Popper. Had Lakatos been as careful with acknowledgments as he is when the Spiritual Property of the Popperian Church is concerned, he would have pointed out that his liberalization which sees knowledge as a *process* is indebted to Hegel.
[4] Popper, *The Open Society and Its Enemies*, pp. 388ff.
[5] *Ibid.*, p. 390. Cf. also Note 28.
[6] *Ibid.* Cf. Note 22 and the corresponding text.
[7] *Ibid.*, p. 391.
[8] *Ibid.*, p. 231.
[9] I am referring here to the following two papers: 'Epistemology Without a Knowing Subject', in Bob Van Rootselaar and J. F. Staal (eds.), *Logic, Methodology and a Knowledge of Science*, Vol. III, North-Holland, Amsterdam, 1968, as well as 'On the Theory of the Objective Mind'. In the first paper, *birdnests* are assigned to the 'third world' (p. 341) and an interaction is assumed between them and the remaining worlds. They are assigned to the third world *because of their function.* But then stones and rivers can be found in this third world too, for a bird may sit on a stone, or take a bath in a river. As a matter of fact, everything that is noticed by some organism will be found in the third world, which will therefore contain the whole material world and all the mistakes mankind had made. It will also contain 'mob psychology'.
[10] Cf. again 'Problems of Empiricism, Part II'.
[11] Cf. *Malleus Maleficarum*, trans. by Montague Summers (Pushkin Press, London, 1928), Part II, question I, Chapter IV: "Here follows the way whereby witches copulate with those Devils known as Incubi", second item, as to the acts, "whether it is always accompanied with the injection of semen received from some other man". The theory goes back to St. Thomas Aquinas.
[12] It is of course possible to establish correlations between the *sentences* of the two theories, but one must realize that the elements of the correlation, when interpreted, cannot be both meaningful, or both true: if relativity is true, then classical descriptions are either always false or are always nonsensical. Continued use of classical sentences must therefore be regarded as an abbreviation for sentences of the following kind: "Given conditions C, the classical sentence S was uttered by a classical physicist whose sense organs are in order, and who understands his physics" – and sentences of this kind, if taken together with certain psychological assumptions, can be used for a test of relativity. However, the *statements* which are expressed by these sentences are part of the *relativistic* framework, for they use relativistic terms. This situation is overlooked by Lakatos who argues as if classical terms and relativistic terms can be combined at will and who infers from this assumption the nonexistence of incommensurability.
[13] This became clear to me in a discussion with Mr. L. Briskman, in Professor Watkins's seminar at the London School of Economics.
[14] This seems to occur in certain versions of the general theory of relativity. Cf. A. Einstein, L. Infeld, and B. Hoffmann, 'The Gravitational Equations and the Problem of Motion', *Annals of Mathematics* **39** (1938), 65, and Sen, *Fields and/or Particles*, pp. 19ff.
[15] This consideration has been raised into a principle by Bohr and Rosenfeld, *Kgl.*

Danske Videnskab. Sekskab, Mat. Fys. Medd., 12, No. 8 (1933), and, more recently, by Robert F. Marzke and John A. Wheeler, 'Gravitation as Geometry I', in Chiu and Hoffmann (eds.), *Gravitation and Relativity*, p. 48: "every proper theory should provide in and by itself its own means for defining the quantities with which it deals. According to this principle, classical general relativity should admit to calibrations of space and time that are altogether free of any reference to [objects which are external] to it such as rigid rods, inertial clocks, or atomic clocks [which involve] the quantum of action."

[16] For this point and further arguments see A. S. Eddington, *The Mathematical Theory of Relativity*, Cambridge University Press, Cambridge, 1963, p. 33. The more general problem of concepts and numbers has been treated by Hegel, *Logik*, I, Das Mass.

[17] This takes care of an objection which Professor J. W. N. Watkins has raised on various occasions.

[18] For further details, especially concerning the concept of mass, the function of 'bridge laws' or 'correspondence rules', and the two-language model, see Section iv of 'Problems of Empiricism'. It is clear that, given the situation described in the text, we cannot derive classical mechanics from relativity, not even approximately. For example, we cannot derive the classical law of mass conservation from relativistic laws. The possibility of connecting the *formulas* of the two disciplines in a manner that might satisfy a pure mathematician, or an instrumentalist, is, however, not excluded. For an analogous situation in the case of quantum mechanics see Section 3 of my paper 'On a Recent Critique of Complementarity'. See also Section 2 of the same article for more general considerations.

[19] R. Carnap, 'The Methodological Character of Theoretical Concepts', *Minnesota Studies in the Philosophy of Science*, Vol. I, ed. by H. Feigl and M. Scriven (University of Minnesota Press, Minneapolis, 1956), p. 47.

[20] An even more conservative principle is sometimes used when discussing the possibility of languages with a logic different from our own: "Any allegedly new possibility must be capable of being fitted into, or understood in terms of, our present conceptual or linguistic apparatus". B. Stroud, 'Conventionalism and the Indeterminacy of Translations', *Synthese*, 1968, p. 173.

[21] As an example the reader is invited to consult J. Piaget, *The Construction of Reality in the Child*, Basic Books, New York, 1954.

[22] *Ibid.*, pp. 5ff.

[23] For the condition of research formulated in the last sentence see Section 8 of 'Reply to Criticism', *Boston Studies in the Philosophy of Science*, Vol. II, ed. by Cohen and Wartofsky. For the role of observation see Section 7 of the same article. For the application of Piaget's work to physics and, more especially, to the theory of relativity see the appendix of Bohm, *The Special Theory of Relativity*. Bohm and Schumacher have also carried out an analysis of the various informal structures which underlie our theories. One of the main results of their work is that Bohr and Einstein argued from incommensurable points of view. Seen in this way the case of Einstein, Podolsky, and Rosen cannot refute the Copenhagen Interpretation and it cannot be refuted by it either. The situation is, rather, that we have two theories, one permitting us to formulate EPR, the other not providing the machinery necessary for such a formulation. We must find independent means for deciding which one to adopt. For further comments on this problem see Section 9 of my 'On a Recent Critique of Complementarity'.

[24] For what follows cf. also my review of Nagel's *Structure of Science* on pp. 237-249 of the *British Journal for the Philosophy of Science* 6 (1966), 237-249.

[25] Carnap, 'The Methodological Character of Theoretical Concepts', p. 40. Cf. also

C. G. Hempel, *Philosophy of Natural Science*, Prentice-Hall, Englewood Cliffs, N.J., 1966, pp. 74ff.

[26] It was for this reason that Leibniz regarded the German of his time and especially the German of the artisans as a perfect observation language, while Latin, for him, was already too much contaminated by theoretical notions. See his 'Unvorgreifliche Gedancken, betreffend die Ausübung und Verbesserung der Teutschen Sprache', published in *Wissenschaftliche Beihefte zur Zeitschrift des allgemeinen deutschen Sprachvereins* IV, 29 (F. Berggold, Berlin, 1907), pp. 292ff.

[27] This objection was raised at a conference by Prof. Roger Buck.

[28] For this point see section I of 'Reply to Criticism', as well as the corresponding sections in 'Problems of Empiricism'.

[29] That the choice between comprehensive theories rests on one's interests entirely and reveals the innermost character of the one who chooses has been emphasized by Fichte in his 'Erste Einleitung in die Wissenschaftslehre'. Fichte discusses the opposition between idealism and materialism which he calls dogmatism. He points out that there are no facts and no considerations of logic which can force us to adopt either the one or the other position. "... we are here faced", he says (*Erste und Zweite Einleitung in die Wissenschaftslehre*, Felix Meiner, Hamburg, 1961, p. 19), "with an absolutely first act that depends on the freedom of thought entirely. It is therefore determined in an arbitrary manner [durch *Willkür*] and, as an arbitrary decision must have a reason nevertheless, by our *inclination* and our *interest*. The final reason for the difference between the idealist and the dogmatist is therefore the difference in their interests".

[30] Here once more the familiar problem of *alienation* arises: what is the result of our own activity becomes separated from it, and assumes an existence of its own. The connection with our intentions and our wishes becomes more and more opaque so that in the end we, instead of leading, follow slavishly the dim outlines of our shadow whether this shadow manifests itself *objectively*, in certain institutions, or *subjectively*, in what some people are pleased to call their 'intellectual honesty', or their 'scientific integrity'. ("... Luther eliminates *external* religiousness and turns religiousness into the *inner* essence of man... he negates the raving parish-priest outside the layman because he puts him into the very heart of the layman". Marx, *Nationaloekonomie und Philosophie*; quoted from Marx, *die Frühschriften*, ed. by Landshut, p. 228.)

In the *economic field* the development is very clear: "In antiquity and in the Middle Ages exploitation was regarded as an obvious, indisputable, and unchangeable fact by both sides, by the free as well as by the slaves, by the feudal lords as well as by their bondsmen. It was precisely because of this knowledge on the part of both parties that the class structure was so transparent; and it was precisely because of the dominance of agriculture that the exploitation of the lower classes *could be seen in the strict sense of the word*. In the Middle Ages the serf worked, say, four days and a half per week on his own plot of land and one day and a half on the land of his master. The place of work for himself was distinctly separated from the place of serfdom... Even the language was clear, it spoke of 'bondsmen' ['Leibeigene', i.e., those whose bodies are owned by someone else]... of 'compulsory service' ['Fronarbeit'] and so on. Thus the class distinctions could not only be *seen*, they could also be *heard*. Language did not conceal the class structure, it expressed it in all desirable clarity. That was true in Egypt, Greece, the European Middle Ages, in Asiatic as well as in European languages. It is no longer true in our present epoch... Workers in early capitalism spent their whole time in the factory. There was neither a spatial nor a temporal separation between the period they worked for their own livelihood and the period they slaved for the capitalist. This led

to the phenomenon I have called ... the 'sociology of repression'. The fact of exploitation was no longer admitted and the repression was facilitated because exploitation could no longer be *seen*". Fritz Sternberg. *Der Dichter und die Ratio; Erinnerungen an Bertolt Brecht* (Sachse und Pohl, Göttingen, 1963), pp. 47ff. *Exactly the same development occurred between Galileo and, say, Laplace.* Science ceased to be a variable human instrument for exploring and changing the world and became a solid block of 'knowledge', impervious to human dreams, wishes, expectations. At the same time the scientists themselves became more and more remote, 'serious', greedy for recognition, and incapable and unwilling to express themselves in a way that could be understood and enjoyed by all. Einstein and Bohr, and Boltzmann before them, were notable exceptions. But they did not change the general trend. There are only a few physicists now who share the humor, the modesty, the sense of perspective, and the philosophical interests of these extraordinary people. All of them have taken over their physics, but they have thoroughly ruined it.

It is even worse in the philosophy of science. For some details, see my papers 'Classical Empiricism' and 'On the Improvement of the Sciences and the Arts, and the Possible Identity of the Two', in *Boston Studies in the Philosophy of Science*, Vol. III, ed. by R. S. Cohen and M. W. Wartofsky (Reidel, Dordrecht, 1968).

[31] Popper has repeately asserted, both in his lectures and in his writings, that while there is progress in the sciences there is no progress in the arts. He bases his assertion on the belief that the content of succeeding theories can be compared and that a judgment of verisimilitude can be made. The refutation of this belief eliminates an important difference, and perhaps the *only* important difference, between science and the arts, and makes it possible to speak of styles and preferences in the first, and of progress in the second.

[32] Cf. B. Brecht, 'Ueber das Zerpfluecken von Gedichten', *Über Lyrik* (Suhrkamp, Frankfurt, 1964). In my lectures on the theory of knowledge I usually present and discuss the thesis that finding a new theory for given facts is exactly like finding a new production for a well-known play. For painting see also E. Gombrich, *Art and Illusion* (Pantheon, New York, 1960).

[33] "The picture of society which we construct for the river-engineers, for the gardeners ... and for the revolutionaries. All of them we invite into our theater. and we ask them not to forget their interest in *entertainment* when they are with us, for we want to turn over the world to their brains and hearts so that they may change it according to their wishes." Brecht, 'Kleines Organon für das Theater', *Schriften zum Theater* (Suhrkamp, Frankfurt, 1964), p. 20; my italics.

INDEX OF NAMES

SYNTHESE LIBRARY

Monographs on Epistemology, Logic, Methodology,
Philosophy of Science, Sociology of Science and of Knowledge, and on the
Mathematical Methods of Social and Behavioral Sciences

1. J. M. BOCHEŃSKI, *A Precis of Mathematical Logic.* 1959, X + 100 pp.
2. P. L. GUIRAUD, *Problèmes et méthodes de la statistique linguistique.* 1960, VI + 146 pp.
3. HANS FREUDENTHAL (ed.), *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, Proceedings of a Colloquium held at Utrecht, The Netherlands, January 1960.* 1961, VI + 194 pp.
4. EVERT W. BETH, *Formal Methods. An Introduction to Symbolic Logic and the Study of Effective Operations in Arithmetic and Logic:* 1962, XIV + 170 pp.
5. B. H. KAZEMIER and D. VUYSJE (eds.), *Logic and Language. Studies dedicated to Professor Rudolf Carnap on the Occasion of his Seventieth Birthday.* 1962, VI + 256 pp.
6. MARX W. WARTOFSKY (ed.), *Proceedings of the Boston Colloquium for the Philosophy of Science, 1961–1962,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume I. 1973, VIII + 212 pp.
7. A. A. ZINOV'EV, *Philosophical Problems of Many-Valued Logic.* 1963, XIV + 155 pp.
8. GEORGES GURVITCH, *The Spectrum of Social Time.* 1964, XXVI + 152 pp.
9. PAUL LORENZEN, *Formal Logic.* 1965, VIII + 123 pp.
10. ROBERT S. COHEN and MARX W. WARTOFSKY (eds.), *In Honor of Philipp Frank,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume II. 1965, XXXIV + 475 pp.
11. EVERT W. BETH, *Mathematical Thought. An Introduction to the Philosophy of Mathematics.* 1965, XII + 208 pp.
12. EVERT W. BETH and JEAN PIAGET, *Mathematical Epistemology and Psychology.* 1966, XII + 326 pp.
13. GUIDO KÜNG, *Ontology and the Logistic Analysis of Language. An Enquiry into the Contemporary Views on Universals.* 1967, XI + 210 pp.
14. ROBERT S. COHEN and MARX W. WARTOFSKY (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science 1964–1966, in Memory of Norwood Russell Hanson,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume III. 1967, XLIX + 489 pp.

15. C. D. BROAD, *Induction, Probability, and Causation. Selected Papers.* 1968, XI + 296 pp.
16. GÜNTHER PATZIG, *Aristotle's Theory of the Syllogism. A Logical-Philosophical Study of Book A of the Prior Analytics.* 1968, XVII + 215 pp.
17. NICHOLAS RESCHER, *Topics in Philosophical Logic.* 1968, XIV + 347 pp.
18. ROBERT S. COHEN and MARX W. WARTOFSKY (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science 1966–1968*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume IV. 1969, VIII + 537 pp.
19. ROBERT S. COHEN and MARX W. WARTOFSKY (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science 1966–1968*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume V. 1969, VIII + 482 pp.
20. J. W. DAVIS, D. J. HOCKNEY, and W. K. WILSON (eds.), *Philosophical Logic.* 1969, VIII + 277 pp.
21. D. DAVIDSON and J. HINTIKKA (eds.), *Words and Objections: Essays on the Work of W. V. Quine.* 1969, VIII + 366 pp.
22. PATRICK SUPPES, *Studies in the Methodology and Foundations of Science. Selected Papers from 1911 to 1969*, XII + 473 pp.
23. JAAKKO HINTIKKA, *Models for Modalities. Selected Essays.* 1969, IX + 220 pp.
24. NICHOLAS RESCHER *et al.* (eds.). *Essay in Honor of Carl G. Hempel. A Tribute on the Occasion of his Sixty-Fifth Birthday.* 1969, VII + 272 pp.
25. P. V. TAVANEC (ed.), *Problems of the Logic of Scientific Knowledge.* 1969, XII + 429 pp.
26. MARSHALL SWAIN (ed.), *Induction, Acceptance, and Rational Belief.* 1970. VII + 232 pp.
27. ROBERT S. COHEN and RAYMOND J. SEEGER (eds.), *Ernst Mach; Physicist and Philosopher*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume VI. 1970, VIII + 295 pp.
28. JAAKKO HINTIKKA and PATRICK SUPPES, *Information and Inference.* 1970, X + 366 pp.
29. KAREL LAMBERT, *Philosophical Problems in Logic. Some Recent Developments.* 1970, VII + 176 pp.
30. ROLF A. EBERLE, *Nominalistic Systems.* 1970, IX + 217 pp.
31. PAUL WEINGARTNER and GERHARD ZECHA (eds.), *Induction, Physics, and Ethics, Proceedings and Discussions of the 1968 Salzburg Colloquium in the Philosophy of Science.* 1970, X + 382 pp.
32. EVERT W. BETH, *Aspects of Modern Logic.* 1970, XI + 176 pp.
33. RISTO HILPINEN (ed.), *Deontic Logic: Introductory and Systematic Readings.* 1971, VII + 182 pp.
34. JEAN-LOUIS KRIVINE, *Introduction to Axiomatic Set Theory.* 1971, VII + 98 pp.
35. JOSEPH D. SNEED, *The Logical Stricture of Mathematical Physics.* 1971, XV + 311 pp.
36. CARL R. KORDIG, *The Justification of Scientific Change.* 1971, XIV + 119 pp.
37. MILIČ ČAPEK, *Bergson and Modern Physics*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume VII, 1971, XV + 414 pp.
38. NORWOOD RUSSELL HANSON, *What I do not Believe, and other Essays*, ed. by Stephen Toulmin and Harry Woolf, 1971, XII + 390 pp.

39. ROGER C. BUCK and ROBERT S. COHEN (eds.), *PSA 1970. In Memory of Rudolf Carnap*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky, Volume VIII. 1971, LXVI + 615 pp. Also available as a paperback.
40. DONALD DAVIDSON and GILBERT HARMAN (eds.), *Semantics of Natural Language*. 1972, X + 769 pp. Also available as a paperback.
41. YEHOSHUA BAR-HILLEL (ed)., *Pragmatics of Natural Languages*. 1971, VII + 231 pp.
42. SÖREN STENLUND, *Combinators, λ-Terms and Proof Theory*. 1972, 184 pp.
43. MARTIN STRAUSS, *Modern Physics and Its Philosophy. Selected Papers in the Logic, History, and Philosophy of Science*. 1972, X + 297 pp.
44. MARIO BUNGE, *Method, Model and Matter*. 1973, VII + 196 pp.
45. MARIO BUNGE, *Philosophy of Physics*. 1973, IX + 248 pp.
46. A. A. ZINOV'EV, *Foundations of the Logical Theory of Scientific Knowledge (Complex Logic)*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume IX. Revised and enlarged English edition with an appendix, by G. A. Smirnov, E. A. Sidorenko, A. M. Fedina, and L. A. Bobrova 1973, XXII + 301 pp. Also available as a paperback.
47. LADISLOV TONDL, *Scientific Procedures*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume X. 1973, XII + 268 pp. Also available as a paperback.
48. NORWOOD RUSSELL HANSON, *Constellations and Conjectures*, ed. by Willard C. Humphreys, Jr. 1973, X + 282 pp.
49. K. J. J. HINTIKKA, J. M. E. MORAVCSIK, and P. SUPPES (eds.), *Approaches to Natural Language. Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*. 1973, VIII + 526 pp. Also available as a paperback.
50. MARIO BUNGE (ed.), *Exact Philosophy – Problems, Tools, and Goals*. 1973, X + 214 pp.
51. RADU J. BOGDAN and ILKKA NIINILUOTO (eds.), *Logic, Language, and Probability*. A selection of papers contributed to Sections IV, VI, and XI of the Fourth International Congress for Logic, Methodology, and Philosophy of Science, Bucharest, September 1971. 1973, X + 323 pp.
52. GLENN PEARCE and PATRICK MAYNARD (eds.), *Conceptual Chance*. 1973, XII + 282 pp.
53. ILKKA NIINILUOTO and RAIMO TUOMELA, *Theoretical Concepts and Hypothetico-Inductive Inference*. 1973, VII + 264 pp.
54. ROLAND FRAÏSSÉ, *Course of Mathematical Logic* – Volume I: *Relation and Logical Formula*. 1973, XVI + 186 pp. Also available as a paperback.
55. ADOLF GRÜNBAUM, *Philosophical Problems of Space and Time*. Second, enlarged edition, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XII. 1973, XXIII + 884 pp. Also available as a paperback.
56. PATRICK SUPPES (ed.), *Space, Time, and Geometry*. 1973, XI + 424 pp.
57. HANS KELSEN, *Essays in Legal and Moral Philosophy*, selected and introduced by Ota Weinberger. 1973, XXVIII + 300 pp.
58. R. J. SEEGER and ROBERT S. COHEN (eds.), *Philosophical Foundations of Science. Proceedings of an AAAS Program, 1969*. Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XI. 1974, X + 545 pp. Also available as paperback.
59. ROBERT S. COHEN and MARX W. WARTOFSKY (eds.), *Logical and Epistemological*

Studies in Contemporary Physics, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XIII. 1973, VIII + 462 pp. Also available as paperback.

60. ROBERT S. COHEN and MARX W. WARTOFSKY (eds.), *Methodological and Historical Essays in the Natural and Social Sciences. Proceedings of the Boston Colloquium for the Philosophy of Science, 1969–1972*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XIV. 1974, VIII + 405 pp. Also available as paperback.
61. ROBERT S. COHEN, J. J. STACHEL, and MARX W. WARTOFSKY (eds.), *For Dirk Struik. Scientific, Historical and Political Essays in Honor of Dirk J. Struik*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XV. 1974, XXVII + 652 pp. Also available as paperback.
62. KAZIMIERZ AJDUKIEWICZ, *Pragmatic Logic*, transl. from the Polish by Olgierd Wojtasiewicz. 1974, XV + 460 pp.
63. SÖREN STENLUND (ed.), *Logical Theory and Semantic Analysis. Essays Dedicated to Stig Kanger on His Fiftieth Birthday*. 1974, V + 217 pp.
64. KENNETH F. SCHAFFNER and ROBERT S. COHEN (eds.), *Proceedings of the 1972 Biennial Meeting, Philosophy of Science Association*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XX. 1974, IX + 444 pp. Also available as paperback.
65. HENRY E. KYBURG, JR., *The Logical Foundations of Statistical Inference*. 1974, IX + 421 pp.
66. MARJORIE GRENE, *The Understanding of Nature: Essays in the Philosophy of Biology*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXIII. 1974, XII + 360 pp. Also available as paperback.
67. JAN M. BROEKMAN, *Structuralism: Moscow, Prague, Paris*. 1974, IX + 117 pp.
68. NORMAN GESCHWIND, *Selected Papers on Language and the Brain*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XVI. 1974, XII + 549 pp. Also available as paperback.
69. ROLAND FRAÏSSÉ. *Course of Mathematical Logic* – Volume II: *Model Theory*. 1974, XIX + 192 pp.
70. ANDRZEJ GRZEGORCZYK, *An Outline of Mathematical Logic. Fundamental Results and Notions Explained with all Details*. 1974, X + 596 pp.
71. FRANZ VON KUTSCHERA, *Philosophy of Language*. 1975, VII + 305 pp.
75. JAAKKO HINTIKKA and UNTO REMES, *The Method of Analysis. Its Geometrical Origin and Its General Significance*. 1974, XVIII + 144 pp.
76. JOHN EMERY MURDOCH and EDITH DUDLEY SYLLA, *The Cultural Context of Medieval Learning. Proceedings of the First International Colloquium on Philosophy, Science, and Theology in the Middle Ages – September 1973*. Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx. W. Wartofsky), Volume XXVI. 1975, X + 566 pp. Also available as paperback.
77. STEFAN AMSTERDAMSKI, *Between Experience and Metaphysics. Philosophical Problems of the Evolution of Science*. Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXXV. 1975, XVIII + 193 pp. Also available as paperback.

SYNTHESE HISTORICAL LIBRARY

Texts and Studies
in the History of Logic and Philosophy

Editors:

N. KRETZMANN (Cornell University)
G. NUCHELMANS (University of Leyden)
L. M. DE RIJK (University of Leyden)

1. M. T. BEONIO-BROCCHIERI FUMAGALLI, *The Logic of Abelard.* Translated from the Italian. 1969, IX + 101 pp.
2. GOTTFRIED WILHELM LEIBNITZ, *Philosophical Papers and Letters.* A selection translated and edited with an introduction, by Leroy E. Loemker. 1969, XII + 736 pp.
3. ERNST MALLY, *Logische Schriften*, ed. by Karl Wolf and Paul Weingartner. 1971, X + 340 pp.
4. LEWIS WHITE BECK (ed.), *Proceedings of the Third International Kant Congress.* 1972, XI + 718 pp.
5. BERNARD BOLZANO, *Theory of Science*, ed. by Jan Berg. 1973, XV + 398 pp.
6. J. M. E. MORAVCSIK (ed.), *Patterns in Plato's Thought. Papers arising out of the 1971 West Coast Greek Philosophy Conference.* 1973, VIII + 212 pp.
7. NABIL SHEHABY, *The Propositional Logic of Avicenna: A Translation from al-Shifā': al-Qiyās*, with Introduction, Commentary and Glossary. 1973, XIII + 296 pp.
8. DESMOND PAUL HENRY, *Commentary on De Grammatico: The Historical-Logical Dimensions of a Dialogue of St. Anselm's.* 1974, IX + 345 pp.
9. JOHN CORCORAN, *Ancient Logic and Its Modern Interpretations.* 1974. X + 208 pp.
10. E. M. BARTH, *The Logic of the Articles in Traditional Philosophy.* 1974, XXVII + 533 pp.
11. JAAKKO HINTIKKA, *Knowledge and the Known. Historical Perspectives in Epistemology.* 1974, XII + 243 pp.
12. E. J. ASHWORTH, *Language and Logic in the Post-Medieval Period.* 1974, XIII + 304 pp.
13. ARISTOTLE, *The Nicomachean Ethics.* Translated with Commentaries and Glossary by Hyppocrates G. Apostle. 1975, XXI+372 pp.
14. R. M. DANCY, *Sense and Contradiction: A Study in Aristotle.* 1975, XII+184 pp.
15. WILBUR RICHARD KNORR, *The Evolution of the Euclidean Elements. A Study of the Theory of Incommensurable Magnitudes and Its Significance for Early Greek Geometry.* 1975, IX+374 pp.
16. AUGUSTINE, *De Dialectica.* Translated with the Introduction and Notes by B. Darrell Jackson. 1975, XI+151 pp.