1

SCIENCE: CONJECTURES AND REFUTATIONS

> Mr. Turnbull had predicted evil consequences, . . . and was now doing the best in his power to bring about the verification of his own prophecies.
>
> ANTHONY TROLLOPE

I. Creating the creation myth of a great discovery.

I

WHEN I received the list of participants in this course and realized that I had been asked to speak to philosophical colleagues I thought, after some hesitation and consultation, that you would probably prefer me to speak about those problems which interest me most, and about those developments with which I am most intimately acquainted. I therefore decided to do what I have never done before: to give you a report on my own work in the philosophy of science, since the autumn of 1919 when I first began to grapple with the problem, '*When should a theory be ranked as scientific?*' or '*Is there a criterion for the scientific character or status of a theory?*'

Why are dates mentioned so frequently?

The problem which troubled me at the time was neither, 'When is a theory true?' nor, 'When is a theory acceptable?' My problem was different. I *wished to distinguish between science and pseudo-science;* knowing very well that science often errs, and that pseudo-science may happen to stumble on the truth.

I knew, of course, the most widely accepted answer to my problem: that science is distinguished from pseudo-science—or from 'metaphysics'—by its *empirical method*, which is essentially *inductive*, proceeding from observation or experiment. But this did not satisfy me. On the contrary, I often formulated my problem as one of distinguishing between a genuinely empirical method and a non-empirical or even a pseudo-empirical method—that is to say, a method which, although it appeals to observation and experiment, nevertheless

A lecture given at Peterhouse, Cambridge, in Summer 1953, as part of a course on developments and trends in contemporary British philosophy, organized by the British Council; originally published under the title 'Philosophy of Science: a Personal Report' in British Philosophy in Mid-Century, *ed. C. A. Mace*, 1957.

does not come up to scientific standards. The latter method may be exemplified by astrology, with its stupendous mass of empirical evidence based on observation—on horoscopes and on biographies.

But as it was not the example of astrology which led me to my problem I should perhaps briefly describe the atmosphere in which my problem arose and the examples by which it was stimulated. After the collapse of the Austrian Empire there had been a revolution in Austria: the air was full of revolutionary slogans and ideas, and new and often wild theories. Among the theories which interested me Einstein's theory of relativity was no doubt by far the most important. Three others were Marx's theory of history, Freud's psycho-analysis, and Alfred Adler's so-called 'individual psychology'.

There was a lot of popular nonsense talked about these theories, and especially about relativity (as still happens even today), but I was fortunate in those who introduced me to the study of this theory. We all—the small circle of students to which I belonged—were thrilled with the result of Eddington's eclipse observations which in 1919 brought the first important confirmation of Einstein's theory of gravitation. It was a great experience for us, and one which had a lasting influence on my intellectual development.

The three other theories I have mentioned were also widely discussed among students at that time. I myself happened to come into personal contact with Alfred Adler, and even to co-operate with him in his social work among the children and young people in the working-class districts of Vienna where he had established social guidance clinics.

It was during the summer of 1919 that I began to feel more and more dissatisfied with these three theories—the Marxist theory of history, psycho-analysis, and individual psychology; and I began to feel dubious about their claims to scientific status. My problem perhaps first took the simple form, 'What is wrong with Marxism, psycho-analysis, and individual psychology? Why are they so different from physical theories, from Newton's theory, and especially from the theory of relativity?'

To make this contrast clear I should explain that few of us at the time would have said that we believed in the *truth* of Einstein's theory of gravitation. This shows that it was not my doubting the *truth* of those other three theories which bothered me, but something else. Yet neither was it that I merely felt mathematical physics to be more *exact* than the sociological or psychological type of theory. Thus what worried me was neither the problem of truth, at that stage at least, nor the problem of exactness or measurability. It was rather that I felt that these other three theories, though posing as sciences, had in fact more in common with primitive myths than with science; that they resembled astrology rather than astronomy.

I found that those of my friends who were admirers of Marx, Freud, and Adler, were impressed by a number of points common to these theories, and especially by their apparent *explanatory power*. These theories appeared to be able to explain practically everything that happened within the fields to which they referred. The study of any of them seemed to have the effect of an intellectual conversion or revelation, opening your eyes to a new truth hidden from those not yet initiated. Once your eyes were thus opened you saw confirming instances everywhere: the world was full of *verifications* of the theory. Whatever happened always confirmed it. Thus its truth appeared manifest; and unbelievers were clearly people who did not want to see the manifest truth; who refused to see it, either because it was against their class interest, or because of their repressions which were still 'un-analysed' and crying aloud for treatment.

The most characteristic element in this situation seemed to me the incessant stream of confirmations, of observations which 'verified' the theories in question; and this point was constantly emphasized by their adherents. A Marxist could not open a newspaper without finding on every page confirming evidence for his interpretation of history; not only in the news, but also in its presentation—which revealed the class bias of the paper—and especially of course in what the paper did *not* say. The Freudian analysts emphasized that their theories were constantly verified by their 'clinical observations'. As for Adler, I was much impressed by a personal experience. Once, in 1919, I reported to him a case which to me did not seem particularly Adlerian, but which he found no difficulty in analysing in terms of his theory of inferiority feelings, although he had not even seen the child. Slightly shocked, I asked him how he could be so sure. 'Because of my thousandfold experience,' he replied; whereupon I could not help saying: 'And with this new case, I suppose, your experience has become thousand-and-one-fold.'

What I had in mind was that his previous observations may not have been much sounder than this new one; that each in its turn had been interpreted in the light of 'previous experience', and at the same time counted as additional confirmation. What, I asked myself, did it confirm? No more than that a case could be interpreted in the light of the theory. But this meant very little, I reflected, since every conceivable case could be interpreted in the light of Adler's theory, or equally of Freud's. I may illustrate this by two very different examples of human behaviour: that of a man who pushes a child into the water with the intention of drowning it; and that of a man who sacrifices his life in an attempt to save the child. Each of these two cases can be explained with equal ease in Freudian and in Adlerian terms. According to Freud the first man suffered from repression (say, of some component of his Oedipus complex), while the second man had achieved sublimation. According to Adler the first man suffered from feelings of inferiority (producing perhaps the need to prove to himself that he dared to commit some crime), and so did the second man (whose need was to prove to himself that he dared to rescue the child). I could not think of any human behaviour which could not be interpreted in terms of either theory. It was precisely this fact—that they always fitted, that they were always confirmed—which in the eyes of their admirers constituted the strongest argument in favour of these theories. It began to dawn on me that this apparent strength was in fact their weakness.

With Einstein's theory the situation was strikingly different. Take one typical instance—Einstein's prediction, just then confirmed by the findings of Eddington's expedition. Einstein's gravitational theory had led to the result that light must be attracted by heavy bodies (such as the sun), precisely as material bodies were attracted. As a consequence it could be calculated that light from a distant fixed star whose apparent position was close to the sun would reach the earth from such a direction that the star would seem to be slightly shifted away from the sun; or, in other words, that stars close to the sun would look as if they had moved a little away from the sun, and from one another. This is a thing which cannot normally be observed since such stars are rendered invisible in daytime by the sun's overwhelming brightness; but during an eclipse it is possible to take photographs of them. If the same constellation is photographed at night one can measure the distances on the two photographs, and check the predicted effect.

Now the impressive thing about this case is the *risk* involved in a prediction of this kind. If observation shows that the predicted effect is definitely absent, then the theory is simply refuted. The theory is *incompatible with certain possible results of observation*—in fact with results which everybody before Einstein would have expected.[1] This is quite different from the situation I have previously described, when it turned out that the theories in question were compatible with the most divergent human behaviour, so that it was practically impossible to describe any human behaviour that might not be claimed to be a verification of these theories.

These considerations led me in the winter of 1919–20 to conclusions which I may now reformulate as follows.

(1) It is easy to obtain confirmations, or verifications, for nearly every theory—if we look for confirmations.

(2) Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory—an event which would have refuted the theory.

(3) Every 'good' scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is.

(4) A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.

(5) Every genuine *test* of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability; but there are degrees of testability: some theories are more testable, more exposed to refutation, than others; they take, as it were, greater risks.

(6) Confirming evidence should not count *except when it is the result of a genuine test of the theory*; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory. (I now speak in such cases of 'corroborating evidence'.)

[1] This is a slight oversimplification, for about half of the Einstein effect may be derived from the classical theory, provided we assume a ballistic theory of light.

(7) Some genuinely testable theories, when found to be false, are still upheld by their admirers—for example by introducing *ad hoc* some auxiliary assumption, or by re-interpreting the theory *ad hoc* in such a way that it escapes refutation. Such a procedure is always possible, but it rescues the theory from refutation only at the price of destroying, or at least lowering, its scientific status. (I later described such a rescuing operation as a '*conventionalist twist*' or a '*conventionalist stratagem*'.)

One can sum up all this by saying that *the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.*

II. Falsificationism restated

II

I may perhaps exemplify this with the help of the various theories so far mentioned. Einstein's theory of gravitation clearly satisfied the criterion of falsifiability. Even if our measuring instruments at the time did not allow us to pronounce on the results of the tests with complete assurance, there was clearly a possibility of refuting the theory.

Astrology did not pass the test. Astrologers were greatly impressed, and misled, by what they believed to be confirming evidence—so much so that they were quite unimpressed by any unfavourable evidence. Moreover, by making their interpretations and prophecies sufficiently vague they were able to explain away anything that might have been a refutation of the theory had the theory and the prophecies been more precise. In order to escape falsification they destroyed the testability of their theory. It is a typical soothsayer's trick to predict things so vaguely that the predictions can hardly fail: that they become irrefutable.

The Marxist theory of history, in spite of the serious efforts of some of its founders and followers, ultimately adopted this soothsaying practice. In some of its earlier formulations (for example in Marx's analysis of the character of the 'coming social revolution') their predictions were testable, and in fact falsified.[2] Yet instead of accepting the refutations the followers of Marx re-interpreted both the theory and the evidence in order to make them agree. In this way they rescued the theory from refutation; but they did so at the price of adopting a device which made it irrefutable. They thus gave a 'conventionalist twist' to the theory; and by this stratagem they destroyed its much advertised claim to scientific status.

The two psycho-analytic theories were in a different class. They were simply non-testable, irrefutable. There was no conceivable human behaviour which could contradict them. This does not mean that Freud and Adler were not seeing certain things correctly: I personally do not doubt that much of what they say is of considerable importance, and may well play its part one day in a psychological science which is testable. But it does mean that those 'clinical observations' which analysts naïvely believe confirm their theory cannot do this any more than the daily confirmations which astrologers find

[2] See, for example, my *Open Society and Its Enemies*, ch. 15, section iii, and notes 13–14.

in their practice.[3] And as for Freud's epic of the Ego, the Super-ego, and the Id, no substantially stronger claim to scientific status can be made for it than for Homer's collected stories from Olympus. These theories describe some facts, but in the manner of myths. They contain most interesting psychological suggestions, but not in a testable form.

At the same time I realized that such myths may be developed, and become testable; that historically speaking all—or very nearly all—scientific theories originate from myths, and that a myth may contain important anticipations of scientific theories. Examples are Empedocles' theory of evolution by trial and error, or Parmenides' myth of the unchanging block universe in which nothing ever happens and which, if we add another dimension, becomes Einstein's block universe (in which, too, nothing ever happens, since everything is, four-dimensionally speaking, determined and laid down from the beginning). I thus felt that if a theory is found to be non-scientific, or 'metaphysical' (as we might say), it is not thereby found to be unimportant, or insignificant, or 'meaningless', or 'nonsensical'.[4] But it cannot claim to be backed by empirical evidence in the scientific sense—although it may easily be, in some genetic sense, the 'result of observation'.

(There were a great many other theories of this pre-scientific or pseudo-

[3] 'Clinical observations', like all other observations, are *interpretations in the light of theories* (see below, sections iv ff.); and for this reason alone they are apt to seem to support those theories in the light of which they were interpreted. But real support can be obtained only from observations undertaken as tests (by 'attempted refutations'); and for this purpose *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. But what kind of clinical responses would refute to the satisfaction of the analyst not merely a particular analytic diagnosis but psycho-analysis itself? And have such criteria ever been discussed or agreed upon by analysts? Is there not, on the contrary, a whole family of analytic concepts, such as 'ambivalence' (I do not suggest that there is no such thing as ambivalence), which would make it difficult, if not impossible, to agree upon such criteria? Moreover, how much headway has been made in investigating the question of the extent to which the (conscious or unconscious) expectations and theories held by the analyst influence the 'clinical responses' of the patient? (To say nothing about the conscious attempts to influence the patient by proposing interpretations to him, etc.) Years ago I introduced the term '*Oedipus effect*' to describe the influence of a theory or expectation or prediction *upon the event which it predicts* or describes: it will be remembered that the causal chain leading to Oedipus' parricide was started by the oracle's prediction of this event. This is a characteristic and recurrent theme of such myths, but one which seems to have failed to attract the interest of the analysts, perhaps not accidentally. (The problem of confirmatory dreams suggested by the analyst is discussed by Freud, for example in *Gesammelte Schriften*, III, 1925, where he says on p. 314: 'If anybody asserts that most of the dreams which can be utilized in an analysis . . . owe their origin to [the analyst's] suggestion, then no objection can be made from the point of view of analytic theory. Yet there is nothing in this fact', he surprisingly adds, 'which would detract from the reliability of our results.')

[4] The case of astrology, nowadays a typical pseudo-science, may illustrate this point. It was attacked, by Aristotelians and other rationalists, down to Newton's day, for the wrong reason—for its now accepted assertion that the planets had an 'influence' upon terrestrial ('sublunar') events. In fact Newton's theory of gravity, and especially the lunar theory of the tides, was historically speaking an offspring of astrological lore. Newton, it seems, was most reluctant to adopt a theory which came from the same stable as for example the theory that 'influenza' epidemics are due to an astral 'influence'. And Galileo, no doubt for the same reason, actually rejected the lunar theory of the tides; and his misgivings about Kepler may easily be explained by his misgivings about astrology.

scientific character, some of them, unfortunately, as influential as the Marxist interpretation of history; for example, the racialist interpretation of history—another of those impressive and all-explanatory theories which act upon weak minds like revelations.)

Thus the problem which I tried to solve by proposing the criterion of falsifiability was neither a problem of meaningfulness or significance, nor a problem of truth or acceptability. It was the problem of drawing a line (as well as this can be done) 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. Years later—it must have been in 1928 or 1929—I called this first problem of mine the '*problem of demarcation*'. The criterion of falsifiability is a solution to this problem of demarcation, for it says that statements or systems of statements, in order to be ranked as scientific, must be capable of conflicting with possible, or conceivable, observations.

III. The Battle with the Positivists

III

Today I know, of course, that this *criterion of demarcation*—the criterion of testability, or falsifiability, or refutability—is far from obvious; for even now its significance is seldom realized. At that time, in 1920, it seemed to me almost trivial, although it solved for me an intellectual problem which had worried me deeply, and one which also had obvious practical consequences (for example, political ones). But I did not yet realize its full implications, or its philosophical significance. When I explained it to a fellow student of the Mathematics Department (now a distinguished mathematician in Great Britain), he suggested that I should publish it. At the time I thought this absurd; for I was convinced that my problem, since it was so important for me, must have agitated many scientists and philosophers who would surely have reached my rather obvious solution. That this was not the case I learnt from Wittgenstein's work, and from its reception; and so I published my results thirteen years later in the form of a criticism of Wittgenstein's *criterion of meaningfulness*.

Wittgenstein, as you all know, tried to show in the *Tractatus* (see for example his propositions 6.53; 6.54; and 5) that all so-called philosophical or metaphysical propositions were actually non-propositions or pseudo-propositions: that they were senseless or meaningless. All genuine (or meaningful) propositions were truth functions of the elementary or atomic propositions which described 'atomic facts', i.e.—facts which can in principle be ascertained by observation. In other words, meaningful propositions were fully reducible to elementary or atomic propositions which were simple statements describing possible states of affairs, and which could in principle be established or rejected by observation. If we call a statement an 'observation statement' not only if it states an actual observation but also if it states anything that *may* be observed, we shall have to say (according to the *Tractatus*, 5 and 4.52) that every genuine proposition must be a truth-function of, and therefore deducible from, observation statements. All other apparent propositions will be meaningless pseudo-propositions; in fact they will be nothing but nonsensical gibberish.

This idea was used by Wittgenstein for a characterization of science, as opposed to philosophy. We read (for example in 4.11, where natural science is taken to stand in opposition to philosophy): 'The totality of true propositions is the total natural science (or the totality of the natural sciences).' This means that the propositions which belong to science are those deducible from *true* observation statements; they are those propositions which can be *verified* by true observation statements. Could we know all true observation statements, we should also know all that may be asserted by natural science.

This amounts to a crude verifiability criterion of demarcation. To make it slightly less crude, it could be amended thus: 'The statements which may possibly fall within the province of science are those which may possibly be verified by observation statements; and these statements, again, coincide with the class of *all* genuine or meaningful statements.' For this approach, then, *verifiability, meaningfulness, and scientific character all coincide*.

I personally was never interested in the so-called problem of meaning; on the contrary, it appeared to me a verbal problem, a typical pseudo-problem. I was interested only in the problem of demarcation, i.e. in finding a criterion of the scientific character of theories. It was just this interest which made me see at once that Wittgenstein's verifiability criterion of meaning was intended to play the part of a criterion of demarcation as well; and which made me see that, as such, it was totally inadequate, even if all misgivings about the dubious concept of meaning were set aside. For Wittgenstein's criterion of demarcation—to use my own terminology in this context—is verifiability, or deducibility from observation statements. But this criterion is too narrow (*and* too wide): it excludes from science practically everything that is, in fact, characteristic of it (while failing in effect to exclude astrology). No scientific theory can ever be deduced from observation statements, or be described as a truth-function of observation statements.

All this I pointed out on various occasions to Wittgensteinians and members of the Vienna Circle. In 1931–2 I summarized my ideas in a largish book (read by several members of the Circle but never published; although part of it was incorporated in my *Logic of Scientific Discovery*); and in 1933 I published a letter to the Editor of *Erkenntnis* in which I tried to compress into two pages my ideas on the problems of demarcation and induction.[5] In this letter

5 My *Logic of Scientific Discovery* (1959, 1960, 1961), here usually referred to as *L.Sc.D.*, is the translation of *Logik der Forschung* (1934), with a number of additional notes and appendices, including (on pp. 312–14) the letter to the Editor of *Erkenntnis* mentioned here in the text; it was first published in *Erkenntnis*, **3**, 1933, pp. 426 f.

Concerning my never published book mentioned here in the text, see R. Carnap's paper '*Ueber Protokollstäze*' (On Protocol-Sentences), *Erkenntnis*, **3**, 1932, pp. 215–28 where he gives an outline of my theory on pp. 223–8, and accepts it. He calls my theory 'procedure B', and says (p. 224, top): 'Starting from a point of view different from Neurath's' (who developed what Carnap calls on p. 223 'procedure A'), 'Popper developed procedure B as and elsewhere I described the problem of meaning as a pseudo-problem, in contrast to the problem of demarcation. But my contribution was classified by members of the Circle as a proposal to replace the verifiability criterion of *meaning* by a falsifiability criterion of *meaning*—which effectively made nonsense of my views.[6] My protests that I was trying to solve, not their pseudo-problem of meaning, but the problem of demarcation, were of no avail.

My attacks upon verification had some effect, however. They soon led to complete confusion in the camp of the verificationist philosophers of sense and nonsense. The original proposal of verifiability as the criterion of meaning was at least clear, simple, and forceful. The modifications and shifts which were now introduced were the very opposite.[7] This, I should say, is now seen even by the participants. But since I am usually quoted as one of them I wish to repeat that although I created this confusion I never participated in it. Neither falsifiability nor testability were proposed by me as criteria of meaning; and although I may plead guilty to having introduced both terms into the discussion, it was not I who introduced them into the theory of meaning.

Criticism of my alleged views was widespread and highly successful. I have yet to meet a criticism of my views.[8] Meanwhile, testability is being widely accepted as a criterion of demarcation.

---

part of his system.' And after describing in detail my theory of tests, Carnap sums up his views as follows (p. 228): 'After weighing the various arguments here discussed, it appears to me that the second language form with procedure B—that is in the form here described—is the most adequate among the forms of scientific language at present advocated . . . in the . . . theory of knowledge.' This paper of Carnap's contained the first published report of my theory of critical testing. (See also my critical remarks in *L.Sc.D.*, note 1 to section 29, p. 104, where the date '1933' should read '1932'; and ch. 11, below, text to note 39.)

[6] Wittgenstein's example of a nonsensical pseudo-proposition is: 'Socrates is identical'. Obviously, 'Socrates is not identical' must also be nonsense. Thus the negation of any nonsense will be nonsense, and that of a meaningful statement will be meaningful. *But the negation of a testable (or falsifiable) statement need not be testable*, as was pointed out, first in my *L.Sc.D.*, (e.g. pp. 38 f.) and later by my critics. The confusion caused by taking testability as a criterion of *meaning* rather than of *demarcation* can easily be imagined.

[7] The most recent example of the way in which the history of this problem is misunderstood is A. R. White's 'Note on Meaning and Verification', *Mind*, **63**, 1954, pp. 66 ff. J. L. Evans's article, *Mind*, **62**, 1953, pp. 1 ff., which Mr. White criticizes, is excellent in my opinion, and unusually perceptive. Understandably enough, neither of the authors can quite reconstruct the story. (Some hints may be found in my *Open Society*, notes 46, 51 and 52 to ch. 11; and a fuller analysis in ch. 11 of the present volume.)

[8] In *L.Sc.D.* I discussed, and replied to, some likely objections which afterwards were indeed raised, without reference to my replies. One of them is the contention that the falsification of a natural law is just as impossible as its verification. The answer is that this objection mixes two entirely different levels of analysis (like the objection that mathematical demonstrations are impossible since checking, no matter how often repeated, can never make it quite certain that we have not overlooked a mistake). On the first level, there is a logical asymmetry: one singular statement—say about the perihelion of Mercury—can formally falsify Kepler's laws; but these cannot be formally verified by any number of singular statements. The attempt to minimize this asymmetry can only lead to confusion. On another level, we may hesitate to accept any statement, even the simplest observation statement; and we may point out that every statement involves *interpretation in the light of theories*, and that it is therefore uncertain. This does not affect the fundamental asymmetry, but it is important: most dissectors of the heart before Harvey observed the wrong things—those, which they expected to see. There can never be anything like a completely safe observation,

CONJECTURES AND REFUTATIONS

The Growth of Scientific Knowledge

by

KARL R. POPPER

LONDON

ROUTLEDGE AND KEGAN PAUL

1963

Popper, “Science Conjectures and Refutations”

- The rhetorical power of a strong, simply stated, easily communicable idea.

- The rhetorical power of a human narrative. “How I discovered falsificationism.”

- New light on Popper’s work from more recent Einstein scholarship.