Popper, Karl (1962), Conjectures and Refutations: The Growth of Scientific Knowledge. New York: Basic Books, pp. 33-65.

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

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?*'

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.

1

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

2

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 star 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

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

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

(4)

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 Protokollsätze*' (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,

IV

I have discussed the problem of demarcation in some detail because I believe that its solution is the key to most of the fundamental problems of the philosophy of science. I am going to give you later a list of some of these other problems, but only one of them—the *problem of induction*—can be discussed here at any length.

I had become interested in the problem of induction in 1923. Although this problem is very closely connected with the problem of demarcation, I did not fully appreciate the connection for about five years.

I approached the problem of induction through Hume. Hume, I felt, was perfectly right in pointing out that induction cannot be logically justified. He held that there can be no valid logical[9] arguments allowing us to establish '*that those instances, of which we have had no experience, resemble those, of which we have had experience*'. Consequently '*even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience*'. For 'shou'd it be said that we have experience' [10]—experience teaching us that objects constantly conjoined with certain other objects continue to be so conjoined—then, Hume says, 'I wou'd renew my question, *why from this experience we form any conclusion beyond those past instances, of which we have had experience*'. In other words, an attempt to justify the practice of induction by an appeal to experience must lead to an *infinite regress*. As a result we can say that theories can never be inferred from observation statements, or rationally justified by them.

I found Hume's refutation of inductive inference clear and conclusive. But I felt completely dissatisfied with his psychological explanation of induction in terms of custom or habit.

It has often been noticed that this explanation of Hume's is philosophically not very satisfactory. It is, however, without doubt intended as a *psychological* rather than a philosophical theory; for it tries to give a causal explanation of a psychological fact—*the fact that we believe in laws*, in statements asserting regularities or constantly conjoined kinds of events—by asserting that this fact is due to (i.e. constantly conjoined with) custom or habit. But even this reformulation of Hume's theory is still unsatisfactory; for what I have just called a 'psychological fact' may itself be described as a custom or habit—

free from the dangers of misinterpretation. (This is one of the reasons why the theory of induction does not work.) The 'empirical basis' consists largely of a mixture of *theories* of lower degree of universality (of 'reproducible effects'). But the fact remains that, relative to whatever basis the investigator may accept (at his peril), he can test his theory only by trying to refute it.

[9] Hume does not say 'logical' but 'demonstrative', a terminology which, I think, is a little misleading. The following two quotations are from the *Treatise of Human Nature*, Book I, Part III, sections vi and xii. (The italics are all Hume's.)

[10] This and the next quotation are from *loc. cit.*, section vi. See also Hume's *Enquiry Concerning Human Understanding*, section IV, Part II, and his *Abstract*, edited 1938 by J. M. Keynes and P. Sraffa, p. 15, and quoted in *L.Sc.D.*, new appendix *VII, text to note 6.

the custom or habit of believing in laws or regularities; and it is neither very surprising nor very enlightening to hear that such a custom or habit must be explained as due to, or conjoined with, a custom or habit (even though a different one). Only when we remember that the words 'custom' and 'habit' are used by Hume, as they are in ordinary language, not merely to *describe* regular behaviour, but rather to *theorize about its origin* (ascribed to frequent repetition), can we reformulate his psychological theory in a more satisfactory way. We can then say that, like other habits, *our habit of believing in laws is the product of frequent repetition*—of the repeated observation that things of a certain kind are constantly conjoined with things of another kind.

This genetico-psychological theory is, as indicated, incorporated in ordinary language, and it is therefore hardly as revolutionary as Hume thought. It is no doubt an extremely popular psychological theory—part of 'common sense', one might say. But in spite of my love of both common sense and Hume, I felt convinced that this psychological theory was mistaken; and that it was in fact refutable on purely logical grounds.

Hume's psychology, which is the popular psychology, was mistaken, I felt, about at least three different things: (*a*) the typical result of repetition; (*b*) the genesis of habits; and especially (*c*) the character of those experiences or modes of behaviour which may be described as 'believing in a law' or 'expecting a law-like succession of events'.

(*a*) The typical result of repetition—say, of repeating a difficult passage on the piano—is that movements which at first needed attention are in the end executed without attention. We might say that the process becomes radically abbreviated, and ceases to be conscious: it becomes 'physiological'. Such a process, far from creating a conscious expectation of law-like succession, or a belief in a law, may on the contrary begin with a conscious belief and destroy it by making it superfluous. In learning to ride a bicycle we may start with the belief that we can avoid falling if we steer in the direction in which we threaten to fall, and this belief may be useful for guiding our movements. After sufficient practice we may forget the rule; in any case, we do not need it any longer. On the other hand, even if it is true that repetition may create unconscious expectations, these become conscious only if something goes wrong (we may not have heard the clock tick, but we may hear that it has stopped).

(*b*) Habits or customs do not, as a rule, *originate* in repetition. Even the habit of walking, or of speaking, or of feeding at certain hours, *begins* before repetition can play any part whatever. We may say, if we like, that they deserve to be called 'habits' or 'customs' only after repetition has played its typical part; but we must not say that the practices in question originated as the result of many repetitions.

(*c*) Belief in a law is not quite the same thing as behaviour which betrays an expectation of a law-like succession of events; but these two are sufficiently closely connected to be treated together. They may, perhaps, in exceptional cases, result from a mere repetition of sense impressions (as in the case of the

stopping clock). I was prepared to concede this, but I contended that normally, and in most cases of any interest, they cannot be so explained. As Hume admits, even a single striking observation may be sufficient to create a belief or an expectation—a fact which he tries to explain as due to an inductive habit, formed as the result of a vast number of long repetitive sequences which had been experienced at an earlier period of life.[11] But this, I contended, was merely his attempt to explain away unfavourable facts which threatened his theory; an unsuccessful attempt, since these unfavourable facts could be observed in very young animals and babies—as early, indeed, as we like. 'A lighted cigarette was held near the noses of the young puppies', reports F. Bäge. 'They sniffed at it once, turned tail, and nothing would induce them to come back to the source of the smell and to sniff again. A few days later, they reacted to the mere sight of a cigarette or even of a rolled piece of white paper, by bounding away, and sneezing.' [12] If we try to explain cases like this by postulating a vast number of long repetitive sequences at a still earlier age we are not only romancing, but forgetting that in the clever puppies' short lives there must be room not only for repetition but also for a great deal of novelty, and consequently of non-repetition.

But it is not only that certain empirical facts do not support Hume; there are decisive arguments of a *purely logical* nature against his psychological theory.

The central idea of Hume's theory is that of *repetition, based upon similarity* (or 'resemblance'). This idea is used in a very uncritical way. We are led to think of the water-drop that hollows the stone: of sequences of unquestionably like events slowly forcing themselves upon us, as does the tick of the clock. But we ought to realize that in a psychological theory such as Hume's, only repetition-for-us, based upon similarity-for-us, can be allowed to have any effect upon us. We must respond to situations as if they were equivalent; *take* them as similar; *interpret* them as repetitions. The clever puppies, we may assume, showed by their response, their way of acting or of reacting, that they recognized or interpreted the second situation as a repetition of the first: that they expected its main element, the objectionable smell, to be present. The situation was a repetition-for-them because they responded to it by *anticipating* its similarity to the previous one.

This apparently psychological criticism has a purely logical basis which may be summed up in the following simple argument. (It happens to be the one from which I originally started my criticism.) The kind of repetition envisaged by Hume can never be perfect; the cases he has in mind cannot be cases of perfect sameness; they can only be cases of similarity. Thus *they are repetitions only from a certain point of view*. (What has the effect upon me of a repetition may not have this effect upon a spider.) But this means that, for logical reasons, there must always be a point of view—such as a system of

[11] *Treatise*, section xiii; section xv, rule 4.

[12] F. Bäge, 'Zur Entwicklung, etc.', *Zeitschrift f. Hundeforschung*, 1933; cp. D. Katz, *Animals and Men*, ch. vi, footnote.

expectations, anticipations, assumptions, or interests—*before* there can be any repetition; which point of view, consequently, cannot be merely the result of repetition. (See now also appendix *x, (1), to my *L.Sc.D.*)

We must thus replace, for the purposes of a psychological theory of the origin of our beliefs, the naïve idea of events which *are* similar by the idea of events to which we react by *interpreting* them as being similar. But if this is so (and I can see no escape from it) then Hume's psychological theory of induction leads to an infinite regress, precisely analogous to that other infinite regress which was discovered by Hume himself, and used by him to explode the logical theory of induction. For what do we wish to explain? In the example of the puppies we wish to explain behaviour which may be described as *recognizing or interpreting* a situation as a repetition of another. Clearly, we cannot hope to explain this by an appeal to earlier repetitions, once we realize that the earlier repetitions must also have been repetitions-for-them, so that precisely the same problem arises again: that of *recognizing or interpreting* a situation as a repetition of another.

To put it more concisely, similarity-for-us is the product of a response involving interpretations (which may be inadequate) and anticipations or expectations (which may never be fulfilled). It is therefore impossible to explain anticipations, or expectations, as resulting from many repetitions, as suggested by Hume. For even the first repetition-for-us must be based upon similarity-for-us, and therefore upon expectations—precisely the kind of thing we wished to explain.

This shows that there is an infinite regress involved in Hume's psychological theory.

Hume, I felt, had never accepted the full force of his own logical analysis. Having refuted the logical idea of induction he was faced with the following problem: how do we actually obtain our knowledge, as a matter of psychological fact, if induction is a procedure which is logically invalid and rationally unjustifiable? There are two possible answers: (1) We obtain our knowledge by a non-inductive procedure. This answer would have allowed Hume to retain a form of rationalism. (2) We obtain our knowledge by repetition and induction, and therefore by a logically invalid and rationally unjustifiable procedure, so that all apparent knowledge is merely a kind of belief—belief based on habit. This answer would imply that even scientific knowledge is irrational, so that rationalism is absurd, and must be given up. (I shall not discuss here the age-old attempts, now again fashionable, to get out of the difficulty by asserting that though induction is of course logically invalid if we mean by 'logic' the same as 'deductive logic', it is not irrational by its own standards, as may be seen from the fact that every reasonable man applies it *as a matter of fact*: it was Hume's great achievement to break this uncritical identification of the question of fact—*quid facti*—and the question of justification or validity—*quid juris*. (See below, point (13) of the appendix to the present chapter.)

It seems that Hume never seriously considered the first alternative. Having

cast out the logical theory of induction by repetition he struck a bargain with common sense, meekly allowing the re-entry of induction by repetition, in the guise of a psychological theory. I proposed to turn the tables upon this theory of Hume's. Instead of explaining our propensity to expect regularities as the result of repetition, I proposed to explain repetition-for-us as the result of our propensity to expect regularities and to search for them.

Thus I was led by purely logical considerations to replace the psychological theory of induction by the following view. Without waiting, passively, for repetitions to impress or impose regularities upon us, we actively try to impose regularities upon the world. We try to discover similarities in it, and to interpret it in terms of laws invented by us. Without waiting for premises we jump to conclusions. These may have to be discarded later, should observation show that they are wrong.

This was a theory of trial and error—of *conjectures and refutations*. It made it possible to understand why our attempts to force interpretations upon the world were logically prior to the observation of similarities. Since there were logical reasons behind this procedure, I thought that it would apply in the field of science also; that scientific theories were not the digest of observations, but that they were inventions—conjectures boldly put forward for trial, to be eliminated if they clashed with observations; with observations which were rarely accidental but as a rule undertaken with the definite intention of testing a theory by obtaining, if possible, a decisive refutation.

V

The belief that science proceeds from observation to theory is still so widely and so firmly held that my denial of it is often met with incredulity. I have even been suspected of being insincere—of denying what nobody in his senses can doubt.

But in fact the belief that we can start with pure observations alone, without anything in the nature of a theory, is absurd; as may be illustrated by the story of the man who dedicated his life to natural science, wrote down everything he could observe, and bequeathed his priceless collection of observations to the Royal Society to be used as inductive evidence. This story should show us that though beetles may profitably be collected, observations may not.

Twenty-five years ago I tried to bring home the same point to a group of physics students in Vienna by beginning a lecture with the following instructions: 'Take pencil and paper; carefully observe, and write down what you have observed!' They asked, of course, *what* I wanted them to observe. Clearly the instruction, 'Observe!' is absurd.[13] (It is not even idiomatic, unless the object of the transitive verb can be taken as understood.) Observation is always selective. It needs a chosen object, a definite task, an interest, a point of view, a problem. And its description presupposes a descriptive language, with property words; it presupposes similarity and classification, which in its turn presupposes interests, points of view, and problems. 'A hungry animal',

[13] See section 30 of *L.Sc.D.*

writes Katz,[14] 'divides the environment into edible and inedible things. An animal in flight sees roads to escape and hiding places. . . . Generally speaking, objects change . . . according to the needs of the animal.' We may add that objects can be classified, and can become similar or dissimilar, *only* in this way—by being related to needs and interests. This rule applies not only to animals but also to scientists. For the animal a point of view is provided by its needs, the task of the moment, and its expectations; for the scientist by his theoretical interests, the special problem under investigation, his conjectures and anticipations, and the theories which he accepts as a kind of background: his frame of reference, his 'horizon of expectations'.

The problem 'Which comes first, the hypothesis (*H*) or the observation (*O*),' is soluble; as is the problem, 'Which comes first, the hen (*H*) or the egg (*O*)'. The reply to the latter is, 'An earlier kind of egg'; to the former, 'An earlier kind of hypothesis'. It is quite true that any particular hypothesis we choose will have been preceded by observations—the observations, for example, which it is designed to explain. But these observations, in their turn, presupposed the adoption of a frame of reference: a frame of expectations: a frame of theories. If they were significant, if they created a need for explanation and thus gave rise to the invention of a hypothesis, it was because they could not be explained within the old theoretical framework, the old horizon of expectations. There is no danger here of an infinite regress. Going back to more and more primitive theories and myths we shall in the end find unconscious, *inborn* expectations.

The theory of inborn *ideas* is absurd, I think; but every organism has inborn *reactions* or *responses*; and among them, responses adapted to impending events. These responses we may describe as 'expectations' without implying that these 'expectations' are conscious. The new-born baby 'expects', in this sense, to be fed (and, one could even argue, to be protected and loved). In view of the close relation between expectation and knowledge we may even speak in quite a reasonable sense of 'inborn knowledge'. This 'knowledge' is not, however, *valid a priori*; an inborn expectation, no matter how strong and specific, may be mistaken. (The newborn child may be abandoned, and starve.)

Thus we are born with expectations; with 'knowledge' which, although not *valid a priori*, is *psychologically or genetically a priori*, i.e. prior to all observational experience. One of the most important of these expectations is the expectation of finding a regularity. It is connected with an inborn propensity to look out for regularities, or with a *need* to *find* regularities, as we may see from the pleasure of the child who satisfies this need.

This 'instinctive' expectation of finding regularities, which is psychologically *a priori*, corresponds very closely to the 'law of causality' which Kant believed to be part of our mental outfit and to be *a priori* valid. One might thus be inclined to say that Kant failed to distinguish between psychologically *a priori* ways of thinking or responding and *a priori* valid beliefs. But I do

[14] Katz, *loc. cit.*

8

not think that his mistake was quite as crude as that. For the expectation of finding regularities is not only psychologically *a priori*, but also logically *a priori*: it is logically prior to all observational experience, for it is prior to any recognition of similarities, as we have seen; and all observation involves the recognition of similarities (or dissimilarities). But in spite of being logically *a priori* in this sense the expectation is not valid *a priori*. For it may fail: we can easily construct an environment (it would be a lethal one) which, compared with our ordinary environment, is so chaotic that we completely fail to find regularities. (All natural laws could remain valid: environments of this kind have been used in the animal experiments mentioned in the next section.)

Thus Kant's reply to Hume came near to being right; for the distinction between an *a priori* valid expectation and one which is both genetically *and* logically prior to observation, but not *a priori* valid, is really somewhat subtle. But Kant proved too much. In trying to show how knowledge is possible, he proposed a theory which had the unavoidable consequence that our quest for knowledge must necessarily succeed, which is clearly mistaken. When Kant said, 'Our intellect does not draw its laws from nature but imposes its laws upon nature', he was right. But in thinking that these laws are necessarily true, or that we necessarily succeed in imposing them upon nature, he was wrong.[15] Nature very often resists quite successfully, forcing us to discard our laws as refuted; but if we live we may try again.

To sum up this logical criticism of Hume's psychology of induction we may consider the idea of building an induction machine. Placed in a simplified 'world' (for example, one of sequences of coloured counters) such a machine may through repetition 'learn', or even 'formulate', laws of succession which hold in its 'world'. If such a machine can be constructed (and I have no doubt that it can) then, it might be argued, my theory must be wrong; for if a machine is capable of performing inductions on the basis of repetition, there can be no logical reasons preventing us from doing the same.

The argument sounds convincing, but it is mistaken. In constructing an induction machine we, the architects of the machine, must decide *a priori* what constitutes its 'world'; what things are to be taken as similar or equal; and what *kind* of 'laws' we wish the machine to be able to 'discover' in its 'world'. In other words we must build into the machine a framework determining what is relevant or interesting in its world: the machine will have its 'inborn' selection principles. The problems of similarity will have been solved for it by its makers who thus have interpreted the 'world' for the machine.

[15] Kant believed that Newton's dynamics was *a priori* valid. (See his *Metaphysical Foundations of Natural Science*, published between the first and the second editions of the *Critique of Pure Reason*.) But if, as he thought, we can explain the validity of Newton's theory by the fact that our intellect imposes its laws upon nature, it follows, I think, that our intellect *must succeed* in this; which makes it hard to understand why *a priori* knowledge such as Newton's should be so hard to come by. A somewhat fuller statement of this criticism can be found in ch. 2, especially section ix, and chs. 7 and 8 of the present volume.

VI

Our propensity to look out for regularities, and to impose laws upon nature, leads to the psychological phenomenon of *dogmatic thinking* or, more generally, dogmatic behaviour: we expect regularities everywhere and attempt to find them even where there are none; events which do not yield to these attempts we are inclined to treat as a kind of 'background noise'; and we stick to our expectations even when they are inadequate and we ought to accept defeat. This dogmatism is to some extent necessary. It is demanded by a situation which can only be dealt with by forcing our conjectures upon the world. Moreover, this dogmatism allows us to approach a good theory in stages, by way of approximations: if we accept defeat too easily, we may prevent ourselves from finding that we were very nearly right.

It is clear that this *dogmatic attitude*, which makes us stick to our first impressions, is indicative of a strong belief; while a *critical attitude*, which is ready to modify its tenets, which admits doubt and demands tests, is indicative of a weaker belief. Now according to Hume's theory, and to the popular theory, the strength of a belief should be a product of repetition; thus it should always grow with experience, and always be greater in less primitive persons. But dogmatic thinking, an uncontrolled wish to impose regularities, a manifest pleasure in rites and in repetition as such, are characteristic of primitives and children; and increasing experience and maturity sometimes create an attitude of caution and criticism rather than of dogmatism.

I may perhaps mention here a point of agreement with psycho-analysis. Psycho-analysts assert that neurotics and others interpret the world in accordance with a personal set pattern which is not easily given up, and which can often be traced back to early childhood. A pattern or scheme which was adopted very early in life is maintained throughout, and every new experience is interpreted in terms of it; verifying it, as it were, and contributing to its rigidity. This is a description of what I have called the dogmatic attitude, as distinct from the critical attitude, which shares with the dogmatic attitude the quick adoption of a schema of expectations—a myth, perhaps, or a conjecture or hypothesis—but which is ready to modify it, to correct it, and even to give it up. I am inclined to suggest that most neuroses may be due to a partially arrested development of the critical attitude; to an arrested rather than a natural dogmatism; to resistance to demands for the modification and adjustment of certain schematic interpretations and responses. This resistance in its turn may perhaps be explained, in some cases, as due to an injury or shock, resulting in fear and in an increased need for assurance or certainty, analogous to the way in which an injury to a limb makes us afraid to move it, so that it becomes stiff. (It might even be argued that the case of the limb is not merely analogous to the dogmatic response, but an instance of it.) The explanation of any concrete case will have to take into account the weight of the difficulties involved in making the necessary adjustments—difficulties which may be considerable, especially in a complex

9

and changing world: we know from experiments on animals that varying degrees of neurotic behaviour may be produced at will by correspondingly varying difficulties.

I found many other links between the psychology of knowledge and psychological fields which are often considered remote from it—for example the psychology of art and music; in fact, my ideas about induction originated in a conjecture about the evolution of Western polyphony. But you will be spared this story.

VII

My logical criticism of Hume's psychological theory, and the considerations connected with it (most of which I elaborated in 1926–7, in a thesis entitled 'On Habit and Belief in Laws' [16]) may seem a little removed from the field of the philosophy of science. But the distinction between dogmatic and critical thinking, or the dogmatic and the critical attitude, brings us right back to our central problem. For the dogmatic attitude is clearly related to the tendency to *verify* our laws and schemata by seeking to apply them and to confirm them, even to the point of neglecting refutations, whereas the critical attitude is one of readiness to change them—to test them; to refute them; to *falsify* them, if possible. This suggests that we may identify the critical attitude with the scientific attitude, and the dogmatic attitude with the one which we have described as pseudo-scientific.

It further suggests that genetically speaking the pseudo-scientific attitude is more primitive than, and prior to, the scientific attitude: that it is a pre-scientific attitude. And this primitivity or priority also has its logical aspect. For the critical attitude is not so much opposed to the dogmatic attitude as super-imposed upon it: criticism must be directed against existing and influential beliefs in need of critical revision—in other words, dogmatic beliefs. A critical attitude needs for its raw material, as it were, theories or beliefs which are held more or less dogmatically.

Thus science must begin with myths, and with the criticism of myths; neither with the collection of observations, nor with the invention of experiments, but with the critical discussion of myths, and of magical techniques and practices. The scientific tradition is distinguished from the pre-scientific tradition in having two layers. Like the latter, it passes on its theories; but it also passes on a critical attitude towards them. The theories are passed on, not as dogmas, but rather with the challenge to discuss them and improve upon them. This tradition is Hellenic: it may be traced back to Thales, founder of the first *school* (I do not mean 'of the first *philosophical* school', but simply 'of the first school') which was not mainly concerned with the preservation of a dogma.[17]

The critical attitude, the tradition of free discussion of theories with the

[16] A thesis submitted under the title '*Gewohnheit und Gesetzerlebnis*' to the Institute of Education of the City of Vienna in 1927. (Unpublished.)

[17] Further comments on these developments may be found in chs. 4 and 5, below.

aim of discovering their weak spots so that they may be improved upon, is the attitude of reasonableness, of rationality. It makes far-reaching use of both verbal argument and observation—of observation in the interest of argument, however. The Greeks' discovery of the critical method gave rise at first to the mistaken hope that it would lead to the solution of all the great old problems; that it would establish certainty; that it would help to *prove* our theories, to *justify* them. But this hope was a residue of the dogmatic way of thinking; in fact nothing can be justified or proved (outside of mathematics and logic). The demand for rational proofs in science indicates a failure to keep distinct the broad realm of rationality and the narrow realm of rational certainty: it is an untenable, an unreasonable demand.

Nevertheless, the role of logical argument, of deductive logical reasoning, remains all-important for the critical approach; not because it allows us to prove our theories, or to infer them from observation statements, but because only by purely deductive reasoning is it possible for us to discover what our theories imply, and thus to criticize them effectively. Criticism, I said, is an attempt to find the weak spots in a theory, and these, as a rule, can be found only in the more remote logical consequences which can be derived from it. It is here that purely logical reasoning plays an important part in science.

Hume was right in stressing that our theories cannot be validly inferred from what we can know to be true—neither from observations nor from anything else. He concluded from this that our belief in them was irrational. If 'belief' means here our inability to doubt our natural laws, and the constancy of natural regularities, then Hume is again right: this kind of dogmatic belief has, one might say, a physiological rather than a rational basis. If, however, the term 'belief' is taken to cover our critical acceptance of scientific theories —a *tentative* acceptance combined with an eagerness to revise the theory if we succeed in designing a test which it cannot pass—then Hume was wrong. In such an acceptance of theories there is nothing irrational. There is not even anything irrational in relying for practical purposes upon well-tested theories, for no more rational course of action is open to us.

Assume that we have deliberately made it our task to live in this unknown world of ours; to adjust ourselves to it as well as we can; to take advantage of the opportunities we can find in it; and to explain it, *if* possible (we need not assume that it is), and as far as possible, with the help of laws and explanatory theories. *If we have made this our task, then there is no more rational procedure than the method of trial and error—of conjecture and refutation:* of boldly proposing theories; of trying our best to show that these are erroneous; and of accepting them tentatively if our critical efforts are unsuccessful.

From the point of view here developed all laws, all theories, remain essentially tentative, or conjectural, or hypothetical, even when we feel unable to doubt them any longer. Before a theory has been refuted we can never know in what way it may have to be modified. That the sun will always rise and set within twenty-four hours is still proverbial as a law 'established by induction beyond reasonable doubt'. It is odd that this example is still in use, though it

may have served well enough in the days of Aristotle and Pytheas of Massalia —the great traveller who for centuries was called a liar because of his tales of Thule, the land of the frozen sea and the *midnight sun*.

The method of trial and error is not, of course, simply identical with the scientific or critical approach—with the method of conjecture and refutation. The method of trial and error is applied not only by Einstein but, in a more dogmatic fashion, by the amoeba also. The difference lies not so much in the trials as in a critical and constructive attitude towards errors; errors which the scientist consciously and cautiously tries to uncover in order to refute his theories with searching arguments, including appeals to the most severe experimental tests which his theories and his ingenuity permit him to design.

The critical attitude may be described as the conscious attempt to make our theories, our conjectures, suffer in our stead in the struggle for the survival of the fittest. It gives us a chance to survive the elimination of an inadequate hypothesis—when a more dogmatic attitude would eliminate it by eliminating us. (There is a touching story of an Indian community which disappeared because of its belief in the holiness of life, including that of tigers.) We thus obtain the fittest theory within our reach by the elimination of those which are less fit. (By 'fitness' I do not mean merely 'usefulness' but truth; see chapters 3 and 10, below.) I do not think that this procedure is irrational or in need of any further rational justification.

VIII

Let us now turn from our logical criticism of the *psychology of experience* to our real problem—the problem of *the logic of science*. Although some of the things I have said may help us here, in so far as they may have eliminated certain psychological prejudices in favour of induction, my treatment of the *logical problem of induction* is completely independent of this criticism, and of all psychological considerations. Provided you do not dogmatically believe in the alleged psychological fact that we make inductions, you may now forget my whole story with the exception of two logical points: my logical remarks on testability or falsifiability as the criterion of demarcation; and Hume's logical criticism of induction.

From what I have said it is obvious that there was a close link between the two problems which interested me at that time: demarcation, and induction or scientific method. It was easy to see that the method of science is criticism, i.e. attempted falsifications. Yet it took me a few years to notice that the two problems—of demarcation and of induction—were in a sense one.

Why, I asked, do so many scientists believe in induction? I found they did so because they believed natural science to be characterized by the inductive method—by a method starting from, and relying upon, long sequences of observations and experiments. They believed that the difference between genuine science and metaphysical or pseudo-scientific speculation depended solely upon whether or not the inductive method was employed. They

believed (to put it in my own terminology) that only the inductive method could provide a satisfactory *criterion of demarcation*.

I recently came across an interesting formulation of this belief in a remarkable philosophical book by a great physicist—Max Born's *Natural Philosophy of Cause and Chance*.[18] He writes: 'Induction allows us to generalize a number of observations into a general rule: that night follows day and day follows night . . . But while everyday life has no definite criterion for the validity of an induction, . . . science has worked out a code, or rule of craft, for its application.' Born nowhere reveals the contents of this inductive code (which, as his wording shows, contains a 'definite criterion for the validity of an induction'); but he stresses that 'there is no logical argument' for its acceptance: 'it is a question of faith'; and he is therefore 'willing to call induction a metaphysical principle'. But why does he believe that such a code of valid inductive rules must exist? This becomes clear when he speaks of the 'vast communities of people ignorant of, or rejecting, the rule of science, among them the members of anti-vaccination societies and believers in astrology. It is useless to argue with them; I cannot compel them to accept the same criteria of valid induction in which I believe: the code of scientific rules.' This makes it quite clear that *'valid induction' was here meant to serve as a criterion of demarcation between science and pseudo-science.*

But it is obvious that this rule or craft of 'valid induction' is not even metaphysical: it simply does not exist. No rule can ever guarantee that a generalization inferred from true observations, however often repeated, is true. (Born himself does not believe in the truth of Newtonian physics, in spite of its success, although he believes that it is based on induction.) And the success of science is not based upon rules of induction, but depends upon luck, ingenuity, and the purely deductive rules of critical argument.

I may summarize some of my conclusions as follows:

(1) Induction, i.e. inference based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor one of scientific procedure.

(2) The actual procedure of science is to operate with conjectures: to jump to conclusions—often after one single observation (as noticed for example by Hume and Born).

(3) Repeated observations and experiments function in science as *tests* of our conjectures or hypotheses, i.e. as attempted refutations.

(4) The mistaken belief in induction is fortified by the need for a criterion of demarcation which, it is traditionally but wrongly believed, only the inductive method can provide.

(5) The conception of such an inductive method, like the criterion of verifiability, implies a faulty demarcation.

(6) None of this is altered in the least if we say that induction makes theories only probable rather than certain. (See especially chapter 10, below.)

[18] Max Born, *Natural Philosophy of Cause and Chance*, Oxford, 1949, p. 7.

11.

IX

If, as I have suggested, the problem of induction is only an instance or facet of the problem of demarcation, then the solution to the problem of demarcation must provide us with a solution to the problem of induction. This is indeed the case, I believe, although it is perhaps not immediately obvious.

For a brief formulation of the problem of induction we can turn again to Born, who writes: '. . . no observation or experiment, however extended, can give more than a finite number of repetitions'; therefore, 'the statement of a law—B depends on A—always transcends experience. Yet this kind of statement is made everywhere and all the time, and sometimes from scanty material.' [19]

In other words, the logical problem of induction arises from (*a*) Hume's discovery (so well expressed by Born) that it is impossible to justify a law by observation or experiment, since it 'transcends experience'; (*b*) the fact that science proposes and uses laws 'everywhere and all the time'. (Like Hume, Born is struck by the 'scanty material', i.e. the few observed instances upon which the law may be based.) To this we have to add (*c*) *the principle of empiricism* which asserts that in science, only observation and experiment may decide upon the *acceptance or rejection* of scientific statements, including laws and theories.

These three principles, (*a*), (*b*), and (*c*), appear at first sight to clash; and this apparent clash constitutes the *logical problem of induction.*

Faced with this clash, Born gives up (*c*), the principle of empiricism (as Kant and many others, including Bertrand Russell, have done before him), in favour of what he calls a 'metaphysical principle'; a metaphysical principle which he does not even attempt to formulate; which he vaguely describes as a 'code or rule of craft'; and of which I have never seen any formulation which even looked promising and was not clearly untenable.

But in fact the principles (*a*) to (*c*) do not clash. We can see this the moment we realize that the acceptance by science of a law or of a theory is *tentative only*; which is to say that all laws and theories are conjectures, or tentative *hypotheses* (a position which I have sometimes called 'hypotheticism'); and that we may reject a law or theory on the basis of new evidence, without necessarily discarding the old evidence which originally led us to accept it.[20]

The principle of empiricism (*c*) can be fully preserved, since the fate of a theory, its acceptance or rejection, is decided by observation and experiment —by the result of tests. So long as a theory stands up to the severest tests we can design, it is accepted; if it does not, it is rejected. But it is never inferred, in any sense, from the empirical evidence. There is neither a psychological nor

[19] *Natural Philosophy of Cause and Chance*, p. 6.

[20] I do not doubt that Born and many others would agree that theories are accepted only tentatively. But the widespread belief in induction shows that the far-reaching implications of this view are rarely seen.

a logical induction. *Only the falsity of the theory can be inferred from empirical evidence, and this inference is a purely deductive one.*

Hume showed that it is not possible to infer a theory from observation statements; but this does not affect the possibility of refuting a theory by observation statements. The full appreciation of this possibility makes the relation between theories and observations perfectly clear.

This solves the problem of the alleged clash between the principles (*a*), (*b*), and (*c*), and with it Hume's problem of induction.

X

Thus the problem of induction is solved. But nothing seems less wanted than a simple solution to an age-old philosophical problem. Wittgenstein and his school hold that genuine philosophical problems do not exist;[21] from which it clearly follows that they cannot be solved. Others among my contemporaries do believe that there are philosophical problems, and respect them; but they seem to respect them too much; they seem to believe that they are insoluble, if not taboo; and they are shocked and horrified by the claim that there is a simple, neat, and lucid, solution to any of them. If there is a solution it must be deep, they feel, or at least complicated.

However this may be, I am still waiting for a simple, neat and lucid criticism of the solution which I published first in 1933 in my letter to the Editor of *Erkenntnis*,[22] and later in *The Logic of Scientific Discovery*.

Of course, one can invent new problems of induction, different from the one I have formulated and solved. (Its formulation was half its solution.) But I have yet to see any reformulation of the problem whose solution cannot be easily obtained from my old solution. I am now going to discuss some of these re-formulations.

One question which may be asked is this: how do we really jump from an observation statement to a theory?

Although this question appears to be psychological rather than philosophical, one can say something positive about it without invoking psychology. One can say first that the jump is not from an observation statement, but from a problem-situation, and that the theory must allow us *to explain* the observations which created the problem (that is, *to deduce* them from the theory strengthened by other accepted theories and by other observation statements, the so-called initial conditions). This leaves, of course, an immense number of possible theories, good and bad; and it thus appears that our question has not been answered.

But this makes it fairly clear that when we asked our question we had more in mind than, 'How do we jump from an observation statement to a theory?' The question we had in mind was, it now appears, 'How do we jump from an observation statement to a *good* theory?' But to this the answer is: by jumping first to *any* theory and then testing it, to find whether it is good or not; i.e.

[21] Wittgenstein still held this belief in 1946; see note 8 to ch. 2, below.

[22] See Note 5 above.

by repeatedly applying the critical method, eliminating many bad theories, and inventing many new ones. Not everybody is able to do this; but there is no other way.

Other questions have sometimes been asked. The original problem of induction, it was said, is the problem of *justifying* induction, i.e. of justifying inductive inference. If you answer this problem by saying that what is called an 'inductive inference' is always invalid and therefore clearly not justifiable, the following new problem must arise: how do you justify your method of trial and error? Reply: the method of trial and error is a *method of eliminating false theories* by observation statements; and the justification for this is the purely logical relationship of deducibility which allows us to assert the falsity of universal statements if we accept the truth of singular ones.

Another question sometimes asked is this: why is it reasonable to prefer non-falsified statements to falsified ones? To this question some involved answers have been produced, for example pragmatic answers. But from a pragmatic point of view the question does not arise, since false theories often serve well enough: most formulae used in engineering or navigation are known to be false, although they may be excellent approximations and easy to handle; and they are used with confidence by people who know them to be false.

The only correct answer is the straightforward one: because we search for truth (even though we can never be sure we have found it), and because the falsified theories are known or believed to be false, while the non-falsified theories may still be true. Besides, we do not prefer *every* non-falsified theory —only one which, in the light of criticism, appears to be better than its competitors: which solves our problems, which is well tested, and of which we think, or rather conjecture or hope (considering other provisionally accepted theories), that it will stand up to further tests.

It has also been said that the problem of induction is, 'Why is it *reasonable* to believe that the future will be like the past?', and that a satisfactory answer to this question should make it plain that such a belief is, in fact, reasonable. My reply is that it is reasonable to believe that the future will be very different from the past in many vitally important respects. Admittedly it is perfectly reasonable to *act* on the assumption that it will, in many respects, be like the past, and that well-tested laws will continue to hold (since we can have no better assumption to act upon); but it is also reasonable to believe that such a course of action will lead us at times into severe trouble, since some of the laws upon which we now heavily rely may easily prove unreliable. (Remember the midnight sun!) One might even say that to judge from past experience, and from our general scientific knowledge, the future will *not* be like the past, in perhaps most of the ways which those have in mind who say that it will. Water will sometimes not quench thirst, and air will choke those who breathe it. An apparent way out is to say that the future will be like the past *in the sense that the laws of nature will not change*, but this is begging the question. We speak of a 'law of nature' only if we think that we have before us a regularity which does not change; and if we find that it changes then we shall not

continue to call it a 'law of nature'. Of course our search for natural laws indicates that we hope to find them, and that we believe that there are natural laws; but our belief in any particular natural law cannot have a safer basis than our unsuccessful critical attempts to refute it.

I think that those who put the problem of induction in terms of the *reasonableness* of our beliefs are perfectly right if they are dissatisfied with a Humean, or post-Humean, sceptical despair of reason. We must indeed reject the view that a belief in science is as irrational as a belief in primitive magical practices —that both are a matter of accepting a 'total ideology', a convention or a tradition based on faith. But we must be cautious if we formulate our problem, with Hume, as one of the reasonableness of our *beliefs*. We should split this problem into three—our old problem of demarcation, or of how to *distinguish* between science and primitive magic; the problem of the rationality of the scientific or critical *procedure*, and of the role of observation within it; and lastly the problem of the rationality of our *acceptance* of theories for scientific and for practical purposes. To all these three problems solutions have been offered here.

One should also be careful not to confuse the problem of the reasonableness of the scientific procedure and the (tentative) acceptance of the results of this procedure—i.e. the scientific theories—with the problem of the rationality or otherwise *of the belief that this procedure will succeed*. In practice, in practical scientific research, this belief is no doubt unavoidable and reasonable, there being no better alternative. But the belief is certainly unjustifiable in a theoretical sense, as I have argued (in section v). Moreover, if we could show, on general logical grounds, that the scientific quest is likely to succeed, one could not understand why anything like success has been so rare in the long history of human endeavours to know more about our world.

Yet another way of putting the problem of induction is in terms of probability. Let t be the theory and e the evidence: we can ask for $P(t,e)$, that is to say, the probability of t, given e. The problem of induction, it is often believed, can then be put thus: construct a *calculus of probability* which allows us to work out for any theory t what its probability is, relative to any given empirical evidence e; and show that $P(t,e)$ increases with the accumulation of supporting evidence, and reaches high values—at any rate values greater than $\frac{1}{2}$.

In *The Logic of Scientific Discovery* I explained why I think that this approach to the problem is fundamentally mistaken.[23] To make this clear, I introduced there the distinction between *probability* and *degree of corroboration or confirmation*. (The term 'confirmation' has lately been so much used and misused that I have decided to surrender it to the verificationists and to use for my own purposes 'corroboration' only. The term 'probability' is best

[23] *L.Sc.D.* (see note 5 above), ch. x, especially sections 80 to 83, also section 34 ff. See also my note 'A Set of Independent Axioms for Probability', *Mind*, N.S. 47, 1938, p. 275. (This note has since been reprinted, with corrections, in the new appendix *ii of *L.Sc.D.* See also the next note but one to the present chapter.)

13

used in some of the many senses which satisfy the well-known calculus of probability, axiomatized, for example, by Keynes, Jeffreys, and myself; but nothing of course depends on the choice of words, as long as we do not *assume*, uncritically, that degree of corroboration must also be a probability —that is to say, that it must satisfy the calculus of probability.)

I explained in my book why we are interested in theories with a *high degree of corroboration*. And I explained why it is a mistake to conclude from this that we are interested in *highly probable* theories. I pointed out that the probability of a statement (or set of statements) is always the greater the less the statement says: it is inverse to the content or the deductive power of the statement, and thus to its explanatory power. Accordingly every interesting and powerful statement must have a low probability; and *vice versa*: a statement with a high probability will be scientifically uninteresting, because it says little and has no explanatory power. Although we seek theories with a high degree of corroboration, *as scientists we do not seek highly probable theories but explanations; that is to say, powerful and improbable theories.*[24] The opposite view—that science aims at high probability—is a characteristic development of verificationism: if you find that you cannot verify a theory, or make it certain by induction, you may turn to probability as a kind of '*Ersatz*' for certainty, in the hope that induction may yield at least that much.

I have discussed the two problems of demarcation and induction at some length. Yet since I set out to give you in this lecture a kind of report on the work I have done in this field I shall have to add, in the form of an Appendix, a few words about some other problems on which I have been working, between 1934 and 1953. I was led to most of these problems by trying to think out the consequences of the solutions to the two problems of demarcation and induction. But time does not allow me to continue my narrative, and to tell you how my new problems arose out of my old ones. Since I cannot even start a discussion of these further problems now, I shall have to confine myself to giving you a bare list of them, with a few explanatory words here and there. But even a bare list may be useful, I think. It may serve to give an idea of the fertility of the approach. It may help to illustrate what our problems look like; and it may show how many there are, and so convince you that there is no need whatever to worry over the question whether philosophical problems exist, or what philosophy is really about. So this list contains, by implication, an apology for my unwillingness to break with the old tradition of trying to solve problems with the help of rational argument, and thus for my unwillingness to participate wholeheartedly in the developments, trends, and drifts, of contemporary philosophy.

[24] A definition, in terms of probabilities (see the next note), of $C(t,e)$, i.e. of the degree of corroboration (of a theory t relative to the evidence e) satisfying the demands indicated in my *L.Sc.D.*, sections 82 to 83, is the following:

$$C(t, e) = E(t,e)\,(1 + P(t)P(t,e)),$$

where $E(t,e) = (P(e,t) - P(e))/(P(e,t) + P(e))$ is a (non-additive) measure of the explanatory power of t with respect to e. Note that $C(t,e)$ is not a probability: it may have values between -1 (refutation of t by e) and $C(t, t) \leqslant +1$. Statements t which are lawlike and thus non-verifiable cannot even reach $C(t,e) = C(t,t)$ upon empirical evidence e. $C(t,t)$ is the *degree of corroborability* of t, and is equal to the *degree of testability* of t, or to the *content* of t. Because of the demands implied in point (6) at the end of section I above, I do not think, however, that it is possible to give a complete formalization of the idea of corroboration (or, as I previously used to say, of confirmation).

(Added 1955 to the first proofs of this paper:)

See also my note 'Degree of Confirmation', *British Journal for the Philosophy of Science*, 5, 1954, pp. 143 ff. (See also 5, pp. 334.) I have since simplified this definition as follows (*B.J.P.S.*, 1955, 5, p. 359:)

$$C(t,e) = (P(e,t) - P(e))/(P(e,t) - P(et) + P(e))$$

For a further improvement, see *B.J.P.S.* 6, 1955, p. 56.

APPENDIX: SOME PROBLEMS IN THE PHILOSOPHY OF SCIENCE

My first three items in this list of additional problems are connected with the calculus of probabilities.

(1) The frequency theory of probability. In *The Logic of Scientific Discovery* I was interested in developing a consistent theory of probability as it is used in science; which means, a statistical or frequency theory of probability. But I also operated there with another concept which I called 'logical probability'. I therefore felt the need for a generalization—for a formal theory of probability which allows different *interpretations*: (*a*) as a theory of the logical probability of a statement relative to any given evidence; including a theory of absolute logical probability, i.e. of the measure of the probability of a statement relative to zero evidence; (*b*) as a theory of the probability of an event relative to any given *ensemble* (or 'collective') of events. In solving this problem I obtained a simple theory which allows a number of further interpretations: it may be interpreted as a calculus of contents, or of deductive systems, or as a class calculus (Boolean algebra) or as propositional calculus; and also as a calculus of *propensities*.[25]

[25] See my note in *Mind*, *loc. cit.* The axiom system given there for elementary (i.e. non-continuous) probability can be simplified as follows ('$\bar{x}$' denotes the complement of x; 'xy' the intersection or conjunction of x and y):

(A1)	$P(xy) \geqslant P(yx)$	(Commutation)
(A2)	$P(x(yz)) \geqslant P((xy)z)$	(Association)
(A3)	$P(xx) \geqslant P(x)$	(Tautology)
(B1)	$P(x) \geqslant P(xy)$	(Monotony)
(B2)	$P(xy) + P(x\bar{y}) = P(x)$	(Addition)
(B3)	$(x)\,(Ey)\,(P(y) \neq O$ *and* $P(xy) = P(x)P(y))$	(Multiplication)
(C1)	*If* $P(y) \neq O$, *then* $P(x,y) = P(xy)/P(y)$	(Definition of relative
(C2)	*If* $P(y) = O$, *then* $P(x,y) = P(x,x) = P(y,y)$	Probability)

Axiom (C2) holds, in this form, for the finitist theory only; it may be omitted if we are prepared to put up with a condition such as $P(y) \neq O$ in most of the theorems on relative probability. For relative probability, (A1) — (B2) and (C1) — (C2), is sufficient; (B3) is not needed. For absolute probability, (A1) — (B3) is necessary and sufficient: without (B3)

(2) This problem of a *propensity interpretation of probability* arose out of my interest in Quantum Theory. It is usually believed that Quantum Theory has to be interpreted statistically, and no doubt statistics is essential for its empirical tests. But this is a point where, I believe, the dangers of the testability theory of meaning become clear. Although the tests of the theory are statistical, and although the theory (say, Schrödinger's equation) may imply statistical consequences, it need not have a statistical meaning: and one can give examples of objective propensities (which are something like generalized forces) and of fields of propensities, which can be measured by statistical methods without being themselves statistical. (See also the last paragraph of chapter 3, below, with note 35.)

(3) The use of statistics in such cases is, in the main, to provide *empirical tests* of theories which need not be purely statistical; and this raises the question of the *refutability of statistical statements*—a problem treated, but not to my full satisfaction, in the 1934 edition of my *The Logic of Scientific Discovery*. I later found, however, that all the elements for constructing a satisfactory solution lay ready for use in that book; certain examples I had given allow a mathematical characterization of a class of infinite chance-like

---

we cannot, for example, derive the definition of absolute in terms of relative probability,

$$P(x) = P(x,\overline{x\bar{x}})$$

nor its weakened corollary

$$(x)(Ey)\,(P(y) \neq O \text{ and } P(x) = P(x,y))$$

from which (B3) results immediately (by substituting for '$P(x,y)$' its definiens). Thus (B3), like all other axioms with the possible exception of (C2), expresses part of the intended meaning of the concepts involved, and we must not look upon $1 > P(x)$ or $1 > P(x,y)$, which are derivable from (B1), with (B3) or with (C1) and (C2), as 'inessential conventions' (as Carnap and others have suggested).

(Added 1955 to the first proofs of this paper; see also note 31, below.)

I have since developed an axiom system for *relative probability* which holds for finite *and* infinite systems (and in which absolute probability can be defined as in the penultimate formula above). Its axioms are:

(B1) $P(x,z) > P(xy,z)$
(B2) If $P(y,y) \neq P(u,y)$ then $P(x,y) + P(\bar{x},y) = P(y,y)$
(B3) $P(xy,z) = P(x,yz)P(y,z)$
(C1) $P(x,x) = P(y,y)$
(D1) If $((u)P(x,u) = P(y,u))$ then $P(w,x) = P(w,y)$
(E1) $(Ex)\,(Ey)\,(Eu)\,(Ew)\,P(x,y) \neq P(u,w)$

This is a slight improvement on a system which I published in *B.J.P.S.*, 6, 1955, pp. 56 f.; 'Postulate 3' is here called 'D1'. (See also *vol. cit.*, bottom of p. 176. Moreover, in line 3 of the last paragraph on p. 57, the words 'and that the limit exists' should be inserted, between brackets, before the word 'all'.)

(Added 1961 to the proofs of the present volume.)

A fairly full treatment of all these questions will now be found in the new addenda to *L.Sc.D.*

I have left this note as in the first publication because I have referred to it in various places. The problems dealt with in this and the preceding note have since been more fully treated in the new appendices to *L.Sc.D.* (To its 1961 American Edition I have added a system of only three axioms; see also section 2 of the *Addenda* to the present volume.)

sequences which are, in a certain sense, the *shortest sequences* of their kind.[26] A statistical statement may now be said to be testable by comparison with these 'shortest sequences'; it is refuted if the statistical properties of the tested *ensembles* differ from the statistical properties of the initial sections of these 'shortest sequences'.

(4) There are a number of further problems connected with the interpretation of the formalism of a quantum theory. In a chapter of *The Logic of Scientific Discovery* I criticized the 'official' interpretation, and I still think that my criticism is valid in all points but one: one example which I used (in section 77) is mistaken. But since I wrote that section, Einstein, Podolski, and Rosen have published a thought-experiment which can be substituted for my example, although their tendency (which is deterministic) is quite different from mine. Einstein's belief in determinism (which I had occasion to discuss with him) is, I believe, unfounded, and also unfortunate: it robs his criticism of much of its force, and it must be emphasized that much of his criticism is quite independent of his determinism.

(5) As to the problem of determinism itself, I have tried to show that even classical physics, which is deterministic in a certain *prima facie* sense, is misinterpreted if used to support a deterministic view of the physical world in Laplace's sense.

(6) In this connection, I may also mention the *problem of simplicity*—of the simplicity of a theory, which I have been able to connect with the content of a theory. It can be shown that what is usually called the simplicity of a theory is associated with its logical improbability, and not with its probability, as has often been supposed. This, indeed, allows us to deduce, from the theory of science outlined above, why it is always advantageous to try the simplest theories first. They are those which offer us the best chance to submit them to severe tests: the simpler theory has always a higher degree of testability than the more complicated one.[27] (Yet I do not think that this settles all problems about simplicity. See also chapter 10, section xviii, below.)

(7) Closely related to this problem is the problem of the *ad hoc* character of a hypothesis, and of degrees of this *ad hoc* character (of '*ad hocness*', if I may so call it). One can show that the methodology of science (and the history of science also) becomes understandable in its details if we assume that the aim of science is to get explanatory theories which are as little *ad hoc* as possible: a 'good' theory is not *ad hoc*, while a 'bad' theory is. On the other hand one can show that the probability theories of induction imply, inadvertently but necessarily, the unacceptable rule: always use the theory which is the most *ad hoc*, i.e. which transcends the available evidence as little as possible. (See also my paper 'The Aim of Science', mentioned in note 28 below.)

(8) An important problem is the problem of the *layers of explanatory hypotheses* which we find in the more developed theoretical sciences, and of

[26] See *L.Sc.D.*, p. 163 (section 55); see especially the new appendix *xvi.
[27] Ibid., sections 41 to 46. But see now also ch. 10, section xviii.

the relations between these layers. It is often asserted that Newton's theory can be induced or even deduced from Kepler's and Galileo's laws. But it can be shown that Newton's theory (including his theory of absolute space) strictly speaking contradicts Kepler's (even if we confine ourselves to the two-body problem[28] and neglect the mutual attraction between the planets) and also Galileo's; although approximations to these two theories can, of course, be deduced from Newton's. But it is clear that neither a deductive nor an inductive inference can lead, from consistent premises, to a conclusion which contradicts them. These considerations allow us to analyse the logical relations between 'layers' of theories, and also the idea of an *approximation*, in the two senses of (*a*) The theory x is an approximation to the theory y; and (*b*) The theory x is 'a good approximation to the facts'. (See also chapter 10, below.)

(9) A host of interesting problems is raised by *operationalism*, the doctrine that theoretical concepts have to be defined in terms of measuring operations. Against this view, it can be shown that *measurements presuppose theories*. There is no measurement without a theory and no operation which can be satisfactorily described in non-theoretical terms. The attempts to do so are always circular; for example, the description of the measurement of length needs a (rudimentary) theory of heat and temperature-measurement; but these, in turn, involve measurements of length.

The analysis of operationalism shows the need for a *general theory of measurement*; a theory which does not, naïvely, take the practice of measuring as 'given', but explains it by analysing its function in the testing of scientific hypotheses. This can be done with the help of the doctrine of degrees of testability.

Connected with, and closely parallel to, operationalism is the doctrine of *behaviourism*, i.e. the doctrine that, since all test-statements describe behaviour, our theories too must be stated in terms of possible behaviour. But the inference is as invalid as the phenomenalist doctrine which asserts that since all test-statements are observational, theories too must be stated in terms of possible observations. All these doctrines are forms of the verifiability theory of meaning; that is to say, of inductivism.

Closely related to operationalism is *instrumentalism*, i.e. the interpretation of scientific theories as practical instruments or tools for such purposes as the

[28] The contradictions mentioned in this sentence of the text were pointed out, for the case of the many-body problem, by P. Duhem, *The Aim and Structure of Physical Theory* (1905; trans. by P. P. Wiener, 1954). In the case of the two-body problem, the contradictions arise in connection with Kepler's third law, which may be reformulated for the two-body problem as follows. 'Let S be any *set of pairs* of bodies such that *one* body of each pair is of the mass of our sun; then $a^3/T^2 = constant$, for any set S.' Clearly this contradicts Newton's theory, which yields for appropriately chosen units $a^3/T^2 = m_0 + m_1$ (where m_0 = mass of the sun = *constant*, and m_1 = mass of the second body, which varies with this body). But '$a^3/T^2 = constant$' is, of course, an excellent approximation, *provided* the varying masses of the second bodies are all negligible compared with that of our sun. (See also my paper 'The Aim of Science', *Ratio*, 1, pp. 24 ff., and section 15 of the *Postscript* to my *Logic of Scientific Discovery*.)

prediction of impending events. That theories may be used in this way cannot be doubted; but instrumentalism asserts that they can be best understood as instruments; and that this is mistaken, I have tried to show by a comparison of the *different functions* of the formulae of applied and pure science. In this context the problem of the *theoretical* (i.e. non-practical) function of predictions can also be solved. (See chapter 3, section 5, below.)

It is interesting to analyse from the same point of view the function of language—as an instrument. One immediate finding of this analysis is that we use descriptive language in order to talk *about the world*. This provides new arguments in favour of *realism*.

Operationalism and instrumentalism must, I believe, be replaced by 'theoreticism', if I may call it so: by the recognition of the fact that we are always operating within a complex framework of theories, and that we do not aim simply at correlations, but at explanations.

(10) The problem of *explanation* itself. It has often been said that scientific explanation is reduction of the unknown to the known. If pure science is meant, nothing could be further from the truth. It can be said without paradox that scientific explanation is, on the contrary, the reduction of the known to the unknown. In pure science, as opposed to an applied science which takes pure science as 'given' or 'known', explanation is always the logical reduction of hypotheses to others which are of a higher level of universality; of 'known' facts and 'known' theories to assumptions of which we know very little as yet, and which have still to be tested. The analysis of degrees of explanatory power, and of the relationship between genuine and sham explanation and between explanation and prediction, are examples of problems which are of great interest in this context.

(11) This brings me to the problem of the relationship between explanation in the natural sciences and historical explanation (which, strangely enough, is logically somewhat analogous to the problem of explanation in the pure and applied sciences); and to the vast field of problems in the methodology of the social sciences, especially the problems of *historical prediction*; *historicism* and *historical determinism*; and *historical relativism*. These problems are linked, again, with the more general problems of determinism and relativism, including the problems of linguistic relativism.[29]

(12) A further problem of interest is the analysis of what is called 'scientific objectivity'. I have treated this problem in several places, especially in connection with a criticism of the so-called 'sociology of knowledge'.[30]

(13) One type of solution of the problem of induction should be mentioned here again (see section iv, above), in order to warn against it. (Solutions of this kind are, as a rule, put forth without a clear formulation of the problem which they are supposed to solve.) The view I have in mind may be described

[29] See my *Poverty of Historicism*, 1957, sections 28 and note 30 to 32; also the Addendum to vol. ii of my *Open Society* (added to the 4th edition 1962).

[30] *Poverty of Historicism*, section 32; *L.Sc.D.*, section 8; *Open Society*, ch. 23. The passages are complementary.

as follows. It is first taken for granted that nobody seriously doubts that we do, *in fact*, make inductions, and successful ones. (My suggestion that this is a myth, and that the apparent cases of induction turn out, if analysed more carefully, to be cases of the method of trial and error, is treated with the contempt which an utterly unreasonable suggestion of this kind deserves.) It is then said that the task of a theory of induction is to describe and classify our inductive policies or procedures, and perhaps to point out which of them are the most successful and reliable ones and which are less successful or reliable; and that any further question of justification is misplaced. Thus the view I have in mind is characterized by the contention that the distinction between the factual problem of describing how we argue inductively (*quid facti?*), and the problem of the justification of our inductive arguments (*quid juris?*) is a misplaced distinction. It is also said that the justification required is unreasonable, since we cannot expect inductive arguments to be 'valid' in the same sense in which deductive ones may be 'valid': induction simply is not deduction, and it is unreasonable to demand from it that it should conform to the standards of logical—that is, deductive—validity. We must therefore judge it by its own standards—by inductive standards—of reasonableness.

I think that this defence of induction is mistaken. It not only takes a myth for a fact, and the alleged fact for a standard of rationality, with the result that a myth becomes a standard of rationality; but it also propagates, in this way, a principle which may be used to defend *any* dogma against *any* criticism. Moreover, it mistakes the status of formal or 'deductive' logic. (It mistakes it just as much as those who saw it as the systematization of our factual, that is, psychological, 'laws of thought'.) For deduction, I contend, is not valid because we choose or decide to adopt its rules as a standard, or decree that they shall be accepted; rather, it is valid because it adopts, and incorporates, the rules by which truth is transmitted from (logically stronger) premises to (logically weaker) conclusions, and by which falsity is re-transmitted from conclusions to premises. (This re-transmission of falsity makes formal logic the *Organon of rational criticism*—that is, of refutation.)

One point that may be conceded to those who hold the view I am criticizing here is this. In arguing from premises to the conclusion (or in what may be called the 'deductive direction'), we argue from the truth or the certainty or the probability of the premises to the corresponding property of the conclusion; while if we argue from the conclusion to the premises (and thus in what we have called the 'inductive direction'), we argue from the falsity or the uncertainty or the impossibility or the improbability of the conclusion to the corresponding property of the premises; accordingly, we must indeed concede that standards such as, more especially, *certainty*, which apply to arguments in the deductive direction, do not also apply to arguments in the inductive direction. Yet even this concession of mine turns in the end against those who hold the view which I am criticizing here; for they assume, wrongly, that we may argue in the inductive direction, though not to the certainty, yet to the *probability* of our 'generalizations'. But this assumption

is mistaken, for all the intuitive ideas of probability which have ever been suggested.[31]

This is a list of just a few of the problems of the philosophy of science to which I was led in my pursuit of the two fertile and fundamental problems whose story I have tried to tell you.

[31] (Added 1961.) Since 1953, when this lecture was delivered, and since 1955, when I read the proofs, the list given in this appendix has grown considerably, and some more recent contributions which deal with problems not listed here will be found in this volume (see especially ch. 10, below) and in my other books (see especially the new appendices to my *L.Sc.D.*, and the new *Addendum* to vol. ii of my *Open Society* which I have added to the fourth edition, 1962). See especially also my paper 'Probability Magic, or Knowledge out of Ignorance', *Dialectica*, 11, 1957, pp. 354–374.