Assigned: 24 Jan '07 Topic: What is science? (part I)

In doing these readings, remember that our ultimate question is whether computer science is a science.

Required:

Martinich, Ch. 3 ("The Structure of a Philosophical Essay") pp. 49-64

1) Kemeny, John G. (1959), A Philosopher Looks at Science [Intro & Chs.5,10 in PDF] (Princeton: D. van Nostrand).

- Introduction (pp. ix-xii)
- Ch.5 ("The [Scientific] Method", pp. 85-105)

* You can just skim Ch.10 ("What Is Science?", pp.174-183), because his answer is just this: A science is any study that follows the scientific method.

2) *Kolak, Daniel; Hirstein, William; Mandik, Peter; & Waskan, Jonathan (2006), Cognitive Science: An Introduction to Mind and Brain (New York: Routledge).*

** "§4.4.2. The Philosophy of Science"*

Strongly Recommended:

1) Quine, Willard van Orman (1951), "Two Dogmas of Empiricism" Philosophical Review 60: 20-43.

2) Popper, Karl R. (1953), "Science: Conjectures and Refutations", from his Conjectures and Refutations: The Growth of Scientific Knowledge (New York: Harper & Row, 1962).

3) Kuhn, Thomas S. (1962) The Structure of Scientific Revolutions (Chicago: University of Chicago Press).

** "Ch. IX: The Nature and Necessity of Scientific Revolutions"*

A PHILOSOPHER LOOKS AT SCIENCE

by

JOHN G. KEMENY

Professor of Philosophy and
Chairman, Department of Mathematics
and Astronomy, Dartmouth College

D. VAN NOSTRAND COMPANY, INC.,
PRINCETON, NEW JERSEY

TORONTO LONDON

NEW YORK

Introduction

ONCE UPON A TIME there was an ugly little caterpillar. All the other small animals strutted around, preening their colorful feathers or showing off their glittering coats, while the little caterpillar hid and felt ashamed. Then one day he made up his mind he would not rest until he changed himself into the most beautiful caterpillar in the world. He struggled, he puffed, he almost burst himself trying, but he did succeed. "Look at me," he shouted, "I am truly a lovely caterpillar." But the other animals snickered and laughed at him behind his back. Finally a wise old owl, who had been watching from above, said to the deflated little caterpillar: "The others are not laughing at you because you are not beautiful. Don't you know there is no such thing as a beautiful caterpillar? You have turned yourself into a butterfly."

Every philosopher must learn the lesson of this fable. No matter how hard the philosopher tries to discover the laws of nature, no philosopher can ever do so for the simple reason that, if he succeeds, people will call him a scientist.

It is often pointed out that all of Science grew out of Philosophy. If you read about ancient Greece, you will find a group of philosophers asking questions, the answers to which form the basis of our Science. For a long time they could do no more than ask questions and indulge in more or less ingenious guesses, but slowly the modern scientific method developed, giving definite and well-founded answers to these questions. Slowly the philosopher found himself in a dilemma. If he asked a question, he was a philosopher; if he answered it, he was a scientist.

This book is not a science book. There are many excellent and easily readable books explaining the results of modern Science.

This is a book on the Philosophy of Science. By this time the question will arise: "Are there any questions left for the philosopher to answer?" There certainly are. Roughly speaking, the philosopher deals with those questions that the scientist either does not answer or cannot answer. Fortunately, some of the most interesting questions fall into these categories. But I have restricted myself further; not only is this a philosophy book, but it is a philosophy of science book. Therefore I must restrict myself to such philosophical questions as arise in connection with Science.

In other words, this is not a science book, but a book about Science. This is not nearly so subtle a distinction as one might suppose. A beautiful painting is a piece of art; yet a book explaining the technique used by the artist is *about* Art, not necessarily a piece of art itself. Similarly, if we perform an experiment and write up the results, we are acting as scientists. But when we discuss the general problems involved in experimentation, we are writing about Science. What are these problems that scientists do not write about (unless, of course, they happen to be acting as philosophers)? They include discussions on the scientific method, on the concepts and basic assumptions of Science, its subject matter and its limitations, what it makes use of and what it is used for in turn. In general, it is an attempt to give a unified picture of the nature of Science.

Let me try to illustrate this distinction in terms of the Theory of Relativity. If we ask what this theory can tell us about the motion of planets, that is a scientific question; if we ask why this answer is accepted by scientists, this is a question about scientific method and hence belongs to the Philosophy of Science. If we want to see how the theory is deduced from a few basic assumptions, we go to a science book (and there is certainly none better than Einstein's own account). But if we want to know how these basic assumptions could possibly be justified, we had better ask a philosopher. If we want to learn about mathematical methods used in Physics, any good science department will provide us with the answer, but it is not their job to explain just what the relation is between Mathematics and Science. We would also do well not to ask the average scientist such questions, because it may very

well be that he is not in a position to answer us. His whole life is devoted to a task that is all-absorbing and may be more than one human being can accomplish. This is perhaps the reason why many of these tremendously important and interesting questions are left for the philosopher to answer. It is questions of this type that we will try to discuss, and whenever possible we will try to answer them in this book.

Since my basic purpose is to present a unified picture of Science, it is difficult to divide the book into parts. However, I can tell you the basic pattern that I have followed. I start out with certain questions that are presupposed by Science. From this I proceed to a discussion of Science itself and I end up with certain problems that arise out of Science. The first chapter deals with the problem of language and its relevance to the various questions discussed in the book. From this we pass to a discussion of Mathematics, the language that has been found most useful for Science. Since it has often been maintained that the usage of this language involves some basic assumptions, these supposed assumptions are discussed in the third chapter. We are then almost ready to go on to a discussion of Science itself. Just one more tool has to be treated, namely, probabilities. The discussion of Science proper starts with a chapter on the scientific method. After that, four basic questions are asked, questions which arise in a discussion of the scientific method, and a chapter is devoted to each one. This leads us to Chapter 10, which is a discussion of what Science is. Then there are the questions that arise out of Science. Do we live in a completely determined universe? What is Life? What is the nature of our minds? What is the status of values? What is the nature of the social sciences? With this we are brought to the final chapter, which attempts to summarize what has gone before.

As the plan of the book might indicate, it is organized to appeal to the interested layman. I do hope, however, that the unified picture here presented, and the approach of the third chapter (on which the remainder of the book is based), will be of interest even to the expert. A unified picture in so small a space must necessarily be sketchy, so I included a few suggestions for additional reading, which will be found at the end of each chapter. I have tried to arrange these so that references can be quickly found and so that it is easy to see which reference will provide the answer to the question in the reader's mind. If used as an introductory text, it should be supplemented with a good popular book on Science. If it is used in a more advanced course, it should be used in conjunction with some of the books mentioned at the ends of the various chapters. This book could provide the continuity, and the references would fill in details on which the book is too sketchy.

Just one more word of warning. We have already noted the fact that philosophers ask many more questions than they can answer. I believe that asking a good clear question is one of the most important things we can do. We find many instances in the history of both Science and Philosophy where a question was unanswered for centuries until some genius came along and rephrased the question, and all of a sudden it was found that the answer was very simple to find as well. For this reason a great deal of time is spent in this book in clarifying issues. Very often this is the best that I can do.

The book is dedicated to the belief that clarification of a difficult problem is a great step forward. It certainly avoids much fruitless and apparently endless debate, and hence clears the air for fruitful work and the solution of the problem. Nevertheless we are often forced to face one unanswerable question. Have we really learned much? So much that we want to know is still left open, and we remain faced with so many uncertainties about what we do know. Perhaps what we have learned so far is really very little. But when we consider that the pursuit of knowledge for its own sake, the attempt to answer these fundamental questions, has been one of the greatest driving forces for all intellectual pursuits, then it is comforting to note that in all probability these questions, or at least many of them, will forever remain unanswered.

PART ONE

What Science Presupposes

5

The Method

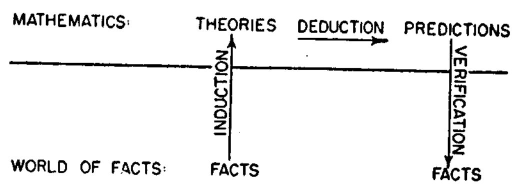
"I should see the garden far better," said Alice to herself, "if I could get to the top of that hill: and here is a path that leads straight to it—at least no, it doesn't do that*—but I suppose it will at last. But how curiously it twists! It's more like a cork-screw than a path! Well* this *turn goes to the hill, I suppose—no it doesn't! This goes straight back to the house! Well then, I'll try it the other way."*

THE FIRST FOUR CHAPTERS covered the preliminaries: what we must know before we can understand Science. Now we are ready to turn to a study of our proper subject matter—the nature of Science. The most characteristic feature of Science is its method, and this is the first thing we want to study. I will maintain the thesis that there is one basic method common to all of Science, and I will try to show just what that method is.

THE CYCLE

As Einstein has repeatedly emphasized, Science must start with facts and end with facts, no matter what theoretical structures it builds in between. First of all the scientist is an observer. Next he tries to describe in complete generality what he saw, and what he expects to see in the future. Next he makes predictions on the basis of his theories, which he checks against facts again.

The most characteristic feature of the method is its cyclic nature. It starts with facts, ends in facts, and the facts ending one cycle are the beginning of the next cycle. A scientist holds his theories tentatively, always prepared to abandon them if the facts do not bear out the predictions. If a series of observations, designed to verify certain predictions, force us to abandon our theory, then we look for a new or improved theory. Thus these facts form the fourth stage for the old theory as well as the first stage of the new theory. Since we expect that Science consists of an endless chain of progress, we may expect this cyclic process to continue indefinitely.



The horizontal line in the diagram separates the world of the experimentalist, the universe of facts, from that of the theoretician, the world of Mathematics. In the world below (the line) we find men peering through microscopes, while above we find an endless string of mathematical formulas. What will interest us most in this chapter is the way we proceed from stage to stage; accordingly we will study three steps. The first step carries us from the original observations to the theories. This is known as "induction," or the formation of theories on the basis of factual knowledge. As we have seen, this means that the scientist finds a mathematical formula which he can interpret to suit the facts that he is trying to incorporate in a theory. Then he asks himself the question: "Is this really what I want?" And he is forced to go back to the world of facts to check his construction. But you cannot check a general law directly; you must first ask what it tells you about particular facts. You cannot observe that the sun rises every day throughout eternity; what you can observe is that it rises today, and that it rises tomorrow, and the next day, etc. Any (finite) number of these can be checked. So the scientist must get from his general laws a prediction as to what will actually happen, say, tomorrow. This step is accomplished by "deduction," by logical analysis of what the general law says about a particular event tomorrow. Then he is ready to return to the facts, and see whether he was right in his prediction. This third and final step, consisting of experiments or observations, is the "verification" of the theory.

As an example of this cycle, I will cite one of the most dramatic chapters from the history of Science.

Our story starts in the year 1820. The French astronomer, Alexis Bouvard, was a little-known scientist whose contribution consisted in a painstaking charting of the paths of the planets. He was especially interested in the three large outer planets, Jupiter, Saturn, and Uranus. Bouvard was performing the very important task of accumulating more factual knowledge, enabling us to check and recheck the accepted theories. Newton's theory was accepted without question as a complete explanation of planetary motion. It was, therefore, a great shock that the observed positions of Uranus did not agree with the predictions. The deviation was small, but it was more than one minute of arc—which could not be put down as an error of observation.

This is the last step in one cycle of the scientific method. Many data were accumulated, primarily by Tycho Brahe. Kepler, Galileo, and Newton succeeded in formulating a good theory. Thousands of predictions were made on the basis of this theory, and until this time all of them were verified. But a single (carefully checked) incorrect prediction is sufficient to force us to modify our theories.

Yet all that we have really shown is that some theory, now accepted, is wrong. We still have a choice as to which theory to abandon. In this we must remember not only general theories, but particular ones; for example, our assumptions as to how many planets there are. Newton's theory was so well established that scientists would rather have abandoned any other part of the accepted body of knowledge. Hence, soon they came to guess that they must have been wrong in assuming that Uranus was the outermost planet.

A new, modified body of theories was formed by assuming that there was a planet beyond Uranus. But this was not enough to explain the observed facts. One had to show that a planet of the right size in the right place would account exactly for the observed deviations. Since the size of Uranus was known, and since Newton's theory as to the strength of attraction between planets was still accepted, it was a problem of pure mathematics to deduce the size and position of the hypothetical planet—or to show that a new planet cannot explain the observations.

The French mathematician Leverrier was the one who succeeded in solving this problem. With the mathematics known in those days this was a very difficult problem requiring considerable originality. Today it would be a routine assignment. Leverrier was able to determine both the size and the position of the unknown planet, which enabled astronomers to look for it.

We might be tempted to say that this was unnecessary. Why couldn't astronomers simply scan that portion of the sky until they found a new planet? The answer is that planets are not at all easy to locate. A planet, unless it is very near us or very large, looks no different from the billions of stars. It can be distinguished only by its path. The stars appear to revolve around the earth as if they were attached to a glass sphere (as the ancients thought they were), while planets move in a more irregular path. We would have to chart the position of all the stars in a region and follow these over a period of weeks until we spotted one whose position relative to the rest is changing. This would be a nearly hopeless task.

But with Leverrier's calculations in hand the Berlin Observatory knew the exact position of the sky and the magnitude of the "star" to look for. This simplified their work sufficiently so that in a brief period of intense observations they verified the existence of the hypothetical planet. The newly discovered planet was named Neptune.

This completed the most recent cycle of scientific method. The previous cycle was finished by Bouvard failing to verify predictions. The inductive step was formulated by several scientists who proposed that we modify our theory as to the number of planets. The deduction was Leverrier's; through a difficult mathematical argument he predicted the size and position of Neptune. The

verification was accomplished at the Berlin Observatory. The cyclic nature of the process is further emphasized by the fact that it was repeated along similar lines in the twentieth century, when Pluto was discovered.

FACTS VS. THEORIES

Before we discuss the three steps of the scientific method, I must say something about the two "worlds" of the scientist. One is the everyday world we are all familiar with, only the scientist's familiarity with this world derives from careful, accurate observations. The other is the mysterious, fascinating world of the theoretician, the world of ideas, the world of mathematical formulas. Establishing a connection between these two worlds is one of the most difficult tasks a scientist must face.

We would like to think of a scientist as starting with "hard facts," and building theories on these. But I doubt that we can state a fact entirely divorced from theoretical interpretations. You might feel that when you see a table, you have a hard fact, but you have actually made use of certain theories you have so thoroughly accepted and assimilated that you use them subconsciously. I certainly do not deny that it is a "hard fact" that you have a sensation which we commonly describe as seeing a table, but this is not all that you mean when you state that you see a table. Suppose I ask you whether you could stick your fist through the object you see; you indignantly reply that the answer is most certainly "No"; after all you just said that it was a table that you saw. But there is nothing in your visual image that makes it logically certain that you see a solid object.

As a matter of fact, under certain circumstances, as in dreams and mirages, you can "put your hand" through a seen "table." It is a theory based on past experience that certain visual images are associated with solid objects. You will also assume that the top of the table looks four-sided from all points of view, but that while it looks like a rectangle from above, the angles will vary as you walk around it; in other words you assume certain primitive optical laws. While there are primitive "hard facts" in your experience, your report of your experience always contains an interpretation of what you think you saw. Sometimes the laws that you assume are far from elementary. When the biologist looks through a microscope and reports seeing a minute living creature, he makes use of advanced laws of optics (in connecting up what he sees through the microscope with what there is) and of laws of Biology (in inferring that the image is that of a living being).

In order to make his facts as reliable as possible, the scientist performs experiments. He arranges a physical situation with specified details, and then he reports nothing but what his instruments "tell" him. It has sometimes been said that all of Science is based on pointer-readings. This is somewhat exaggerated, because the readings tell us nothing unless we know what the meter reads, but it does bring out an important technique. The situation is further complicated by the fact that the scientist himself may serve as the instrument—for example, when he counts the number of albino guinea pigs, or when he "records" the reaction of his subject to a question.

Remembering that this is an oversimplification, let us accept that the scientist starts out with reports of what there really is. (We will return to this problem in the Chapter 7 on concepts.) But what the theoretician gets out of this is a statement like: $x = 3.25$, $y = 2.97$, $z = -4.00$, $t = 49.32$; or he may end up with points of a graph, or a "table" (say of "yes" and "no" against various questions). These are the so-called facts which the theoretician must unite into a theory. His x may be a reading on a ruler, his y the height of a column of mercury on a fixed scale, his z the elevation above sea level measured by a complicated surveying technique, and t the reading on a clock. These are the technical representations of observations, which record one or a series of events. His theories, on the other hand, will be generalized mathematical statements. A theory may be an equation like $xy - z^3 = t - 21.35$, or it may be a complete graph, or a rule of how tables will look in all cases.

It will be convenient to speak of the mathematical record of a fact (like $x = 3.25$, etc.) as a fact. If we allow this, then the only difference between a fact and a theory is that a fact is something that we already know, while a theory also states things not

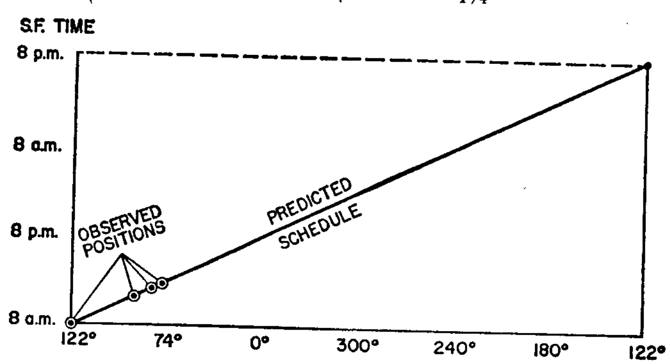
yet observed (and possibly never to be observed). There is a second difference which holds in most, but not all, cases: A fact reports a single event, while a theory reports an unlimited, perhaps infinite, number of events. The reason this does not always hold is that the latter may be false. A fact is always a single thing, like "there is a sun in the sky right now." Although a theory is generally a statement, like "the sun rises every 24 hours," the statement "the sun will rise at 8:00 A.M. tomorrow" also has the status of a theory. The reason for this double possibility is that, while a genuine theory is universal (rather than particular, like a fact), it has logical consequences, which are particular statements. For example, the particular statement "the sun will rise tomorrow" follows from the universal "the sun rises every day."

After making these subtle distinctions, I will ignore them. I will normally use the word "theory" only to apply to universal statements. Hence there will be a twofold difference between facts and theories: Facts are known and particular (refer to a single event), whereas theories are universal and hence can never be known to be entirely true. It is because of this universal nature of theories, because they apply (in principle) to an infinity of events, that Mathematics is almost always indispensable for their statement.

So the scientist makes some observations (perhaps as the result of a planned experiment), and records these in the mathematical language devised by the theoretician. The theoretician tries to formulate a general mathematical proposition, incorporating these facts. Then he develops this theory mathematically, deriving certain predictions of facts. These predictions are, of course, still mathematical propositions, and must be translated back into everyday language before they can be checked.

Let us take an example. We observe a plane that has taken off for a nonstop flight around the world. We note its position at various moments, and record these. What we see is that the plane is over San Francisco at a certain time, over St. Louis at a later time, over Pittsburgh still later, over Newark still later. For simplicity let us record only its longitude and the time (in hours counted from take-off). The table shows that it is traveling at a constant speed (roughly) of about 10 degrees longitude in an hour.

Longitude	*Hours elapsed*
122°	0
90°	3
80°	4
74°	4¾



What we actually have done has been to plot the four reports as points of a graph, and then drawn the simplest smooth curve fitting these points well. It happened to be a straight line. This line is our theory; it is universal since it tells us where the plane will be at every moment from take-off to landing. From it we can read off that the plane will reach the original longitude in 36 hours, and hence we predict that (assuming that the plane left at 8:00 A.M. Monday) it will arrive back in San Francisco at about 8:00 P.M. Tuesday. We started in this by observations (of positions of the plane) and ended with another such observation. In between, these observations were translated into the language of Mathematics (as points on a graph), incorporated into a mathematical theory (a curve that is supposed to hold in general), and the last observation was predicted (from the terminal point on the graph).

This example is typical of the interplay between facts and theories, except that the mathematical propositions are generally

much more complicated, and the connection between theory and facts is quite a bit less direct. With these few remarks as to the two "worlds" of Science, we are prepared to discuss the passage from stage to stage in detail.

INDUCTION

Induction is the process by which the scientist forms a theory to explain the observed facts. Two steps can be distinguished within this procedure: the formation of possible theories and the selection of one of these.

Let us start with the second problem. Given a large number of possible theories, how do we select the one we want? Let me introduce the term "hypothesis" to stand for an interpreted mathematical proposition which we are considering as a possible theory; I will call such a proposition a hypothesis while it is still highly in doubt, and a theory when we have accepted it. Given various hypotheses, we must first of all see whether they explain all the known facts. This is not as simple as it sounds, because our facts are never perfectly accurate, and we must face all the problems of the theory of errors (see the last chapter). But we can select those hypotheses which are reasonably well in agreement with the evidence. Then, of the remaining ones we select the simplest hypothesis.

The question that arises is: Must there be several remaining ones? Isn't it true that if we have enough facts, then there is but one theory that could explain all of these? I would like to convince you that no matter what pains you take to accumulate facts, there will always be many possible hypotheses left; as a matter of fact, there will still be an infinite number of possibilities. The best way to think of this is to think of facts as represented by points on a piece of graph paper, and of hypotheses as curves. For the hypothesis to explain all the facts, the curve must go through all the points (or, since the facts are only approximate, it suffices that the curve should pass very near all the points). Now, no matter how hard you work, you will have only a finite number of points, since within a limited existence you can accumulate only a finite number of facts. Put down a number of points on the graph, and try to draw curves through them. You will soon see that there are an infinite number of possibilities.

Of course, it may happen that of the hypotheses that *you* started with all but one will be eliminated, or even that all of these will be eliminated (since you never really consider *all* mathematical propositions as hypotheses). But there is no reason to expect this to be the normal occurrence. So you must choose from several hypotheses, all of which fit the facts. Why choose the simplest one? For the moment I will just say that it is as good a choice as any, and more convenient than the complex hypotheses. Actually there are better reasons for this, but these will have to wait for the next chapter.

In the example of drawing a curve through given points, this means that we draw as simple or smooth a curve going near the various points as possible. For example, in the airplane example this was a straight line. A straight line is always the simplest, only it is not always a possibility. As a matter of fact, scientists are so fond of straight lines that there have been many examples where a scientist has drawn a straight line through points where these points were nowhere near the line. Of course, this is a violation of the scientific method: first the hypothesis must fit the facts; only then can we worry about its simplicity.

Let us return to the discovery of the planet Neptune. What were the competing hypotheses? First of all, we could have tried some modification of Newton's laws. For example, instead of assuming that the force of gravitation always decreases with the square of the distance, we could have modified this rule. The danger in this is that the rule worked so well for the other planets. Nevertheless, we could have said that this was only because they were pretty near the sun, and hence the deviation did not show up until we got to the outermost planet, Uranus. We could have modified the square of the distance rule by a small term, which was negligible until we reached Uranus. I am quite sure that with sufficient mathematical ingenuity this could have been done, but the resulting rule would have been highly complicated, and it was simpler to formulate the hypothesis that there was an unknown planet. Of course, we run into difficulties trying to say

just when one hypothesis is simpler than another. How complicated must the rule of gravitation become before we decide that it is simpler to look for another planet? I don't know. But fortunately, in most cases, one hypothesis is much simpler than the others. In our present example the rule would have become terribly complicated, so there was no doubt as to which was simplest. The only question was: Can we explain the deviations by stipulating the existence of a new planet (of the right size and in the right place)? When Leverrier showed that we could, this became the simplest hypothesis and was generally accepted even before the planet was actually sighted. We now believe in some sub-atomic particles, not because we have "seen" them (even indirectly), but because assuming their existence is the simplest hypothesis to explain the observed facts.

But how do we form the many different hypotheses from which we are to choose? To this there is no simple answer, since it is essentially a creative process. As soon as someone told us that a new planet could explain why Uranus misbehaved, it seemed most plausible. But how many of us would have thought of this possibility in the first place? How many of us would have thought that the motion of the moon and the falling of the apple are connected? How many of us would have thought that it takes no force to keep things moving, only to start them and stop them? How many of us would have guessed that the blood circulates in our veins? To select one of many hypotheses (once the facts are given) is a mechanical, even if lengthy, procedure; to think these hypotheses up in the first place is the work of genius.

There is one point on which I may be misleading you. When I emphasize the difficulty of thinking up hypotheses, you may get the impression that we must therefore choose from a very small number of possibilities. This is not so. One original idea may give rise to an infinite number of hypotheses. For example, when the idea of a new planet arises, there are the infinite number of different places where it may be (paths that it might follow), and an infinite number of sizes it may have. Let us just consider the distance of the new planet from Uranus, and its size; this already gives a double infinity of possibilities. Of course these are narrowed down by the facts, but there is still some choice: the farther away it is, the larger it must be to account for the deviations. So we start with an infinity of hypotheses, and it takes the most intricate mathematical argument to find just the right distance and just the right size to account for the path of Uranus. You may be interested in knowing that since Leverrier did not know some of the methods now available for this problem, his predictions were actually off by quite a bit.

So we see that the scientist actually thinks up infinitely many hypotheses (thanks to Mathematics), then notes which of these account for all known facts, and finally accepts the simplest remaining hypothesis as his theory.

DEDUCTION

The key to the verification of theories is that you never verify them. What you do verify are logical consequences of the theory. Verification is the process of seeing whether something predicted is really so. Since we can only observe particular facts, we must verify particular consequences of a theory, not the general theory itself.

In the case of Neptune, we could verify that there was a faint "star" at a certain location, and that the same "star" was at a somewhat different location two weeks later. But we could not verify directly how far it was from Uranus, what its path was, nor how large it was. We had to deduce some particular facts from the theory, which could be checked by direct observations.

In the second chapter we noted that logical deduction is no more than the analysis of the meaning of the theory. When we say that these facts follow, we mean that their truth is contained in the truth of the theory, even though we may not have realized this at the time we asserted the theory. When we assert that the sun rises every day, we understand that this implies its rising tomorrow. But few people, if any, would be able to look briefly at the General Theory of Relativity and see that the bending of light rays follows from it. You may feel that this is due to the fact that the theory is so complex. Then let us take Newton's very simple theory and see if it is obvious that planets move in ellipses. Or

is it obvious from the theory (together with some positions of the planet) where a certain planet must be tomorrow? Certainly not, it takes long chains of mathematical deductions to arrive at these conclusions.

I have stated that infinitely many facts are contained in a theory. But it often takes intricate mathematical analysis to bring these out. So the deductive step is designed to derive observable facts from the general theories. The theoretician starts with known facts and with the accepted theories, and finds out just what follows from them. If the theories are true, then every single statement that follows from them must be true! This gives us an unlimited wealth of facts which we can check, no end to the number of verified facts which we can accumulate in support of a given theory.

The only difficulty is that the interesting results are hardly ever the consequences of a single theory, but generally of a large number of theories. So even if the prediction turns out to be false, we are not certain which of the theories is wrong. We are, however, certain that *some* theory is incorrect. It is then again a question of finding the simplest way of improving our body of theories. In the Neptune example, we altered the theory as to the number of planets, but we could have altered Newton's law of gravitation, or even his law of motion, or the laws of optics as they apply to telescopes. We changed that which was simplest to change, but we can never be certain which theory was false. At the price of making the rest of the theories sufficiently complicated, we could rescue any theory. This is the reason why we often hear the claim that each experiment tests our entire body of knowledge.

Consider an everyday example. We have been hearing a great deal about flying saucers. Perhaps by the time you read this book, the mystery will be solved. The hypothesis has been advanced that they are missiles from outer space. What I want to show is that if I am determined to maintain that the saucers have an earthly cause, no amount of evidence can shake my belief. At the moment I can maintain that they are nothing but mass hallucinations. Recently they were spotted on radar. I could try to attribute this to hallucination on the part of the radar operator. If there are too many people who see the image on the screen, I could invent an electronic effect, say, caused by too many television senders, which produce both the "saucers" and the radar images. Of course, this is likely to contradict what we know about electronics, but if I am willing to modify enough theories, I can change the electromagnetic theory, and save my pet hypothesis. If such a missile is actually shot down, I would have to abandon the hypothesis that it is a hallucination, but I could claim that it came from another country on the earth. If it turns out that there is a living being inside the "saucer," different from all we know, I could stipulate that he came from an unexplored island, or even from below the surface of the earth. Of course, to allow for life at the high temperatures below the earth, I would have to modify several theories, but if I am willing to do this, I can still save my pet hypothesis. If one of these "Martians" takes me into the plane and carries me into outer space, I can say that the machine simply took me on a rocket trip and showed me a movie which looked as if I were really looking at the earth disappearing in the distance. Even if we landed on Mars, I could explain this by the great hypnotic power he has over me.

If you got impatient with my skepticism, it was because there comes a point where accepting the fact that we have interplanetary travelers becomes simpler than modifying fundamental theories. But I hope I have convinced you that it is logically possible to save any given theory by giving up others. The reason for this is the following: In checking a theory, we must derive a consequence of the theory, which can be verified by observations. These consequences must make use of several theories, and if they do not check with experience, it is a question whether it is our theory or one of the others that is wrong. We can always suppose the latter, as was shown in the flying-saucer example. This is why the predictions are based on our entire body of knowledge, and why it is best to say that we test this whole body, rather than a particular theory.

VERIFICATION

This third step of the scientific method is similar to the first one: we gather facts. In this case, however, the facts to be observed were predicted, and we "just" see whether they are so or not.

I have tried to show that it is an oversimplification to say that one unfavorable observation can disprove a theory. This is an oversimplification for two reasons: First, the observations and the predictions are only approximate, so that we can make only probability statements (see the last chapter). Secondly, the predictions are based on several theories, and hence there is a choice as to which theory to reject (as was shown in the previous sections). So an unfavorable observation can only make the theory unlikely, or rather it makes the body of theories as a whole unlikely (more precisely, it is unlikely that the whole body is true).

How about favorable observations? First of all, these too are only approximate, so that we are never certain that the prediction was verified. Secondly, the fact that one prediction, or any limited number of predictions, has been verified does not make the theory certain. There always remain an infinity of competing hypotheses, all of which can explain all the known facts. So in this case too we can make only probability statements.

The probabilities that I have discussed here are of the second kind, credibilities. The process of verification consists in checking predictions against observations, and assigning greater or lesser credibility to our body of theories on the basis of the outcome. If the credibility is high, we are satisfied. If it is below a reasonable level, we modify our theories; we must change at least one theory so that our total body will then have a higher credibility. This may mean a relatively minor change, like admitting a new planet, or it may mean replacing the entire structure, as Relativity Theory replaced Newton's System. The decision as to when a change is necessary is complicated, and, in the absence of a good measure of credibility, is highly controversial. It is further confounded by the necessity of taking the simplicity of our theories into account. We consider the credibility of our theories and of competing ones. We abandon our theories either if some other body is much more credible on the given evidence, or if a simpler body can be found which is roughly as credible as ours.

A most interesting illustration of this point is Henri Poincaré's claim that Euclidean Geometry would never be abandoned. Poincaré was an excellent mathematician and philosopher of science. He certainly understood that there was a fundamental difference between pure, abstract Mathematics, and the interpreted version of the same, which is a branch of Science. He also understood the need for selecting the simplest possible theory. But he showed, by an ingenious argument, that one can always salvage Euclidean Geometry (properly interpreted), no matter what facts Physics reveals. He then argued that this Geometry is so much simpler than its competitors that scientists will always stick to this Geometry, and modify their other laws (of Physics) if necessary. I will just give you one example of how this can be done. It seems that there is such a fundamental difference between the finite universe of the Geometry we now favor, and the infinite universe of Euclidean Geometry, that this point should be decidable by experiment. If a ray of light can come back to its origin, then the universe is finite and closed; if it cannot, then it is infinite and open. But we can get around either possibility by modifying our other laws. Suppose a ray of light comes back after a long time (for example, we recognize, somehow, our own gallaxy in the far distance of the universe), this could be explained by stipulating that light travels not in a straight line, but in a very large circle; a circle that is so large that small portions of it seem straight. Then light could come back even in an infinite universe.

If light cannot come back, we can still hang on to our finite, closed universe. We stipulate that the universe expands (as we do stipulate) and that it expands just rapidly enough that light can get closer and closer to our "antipode" (the opposite point of the universe), but can never quite reach it, since it is running away from us with the speed of light. Indeed, some cosmological theories point to this possibility. Then light cannot return even in a finite universe. The latter case is interesting from another standpoint. If in this case we measure lengths as usual, with rulers or their equivalents, the universe is finite. But if we measure distances by

the amount of time it takes light to reach from one end to the other, then the universe becomes infinite—since light can never reach the opposite "pole." Hence we can salvage one or the other Geometry, depending on the way we interpret distances.

Poincaré drew the conclusion that for reasons of simplicity Science will always keep Euclidean Geometry. He published his beliefs in a book that appeared in 1904. In 1905 Einstein's first installment of Relativity Theory appeared, a later installment of which caused us to abandon Euclidean Geometry. The dates here are quite ironic, but they do not mean that so great a thinker as Poincaré made a very bad mistake. What he overlooked was that saving the simplest Geometry might be achievable only at the price of a terrible complication in the other theories. The criterion of simplicity must be applied to our entire body of knowledge. When Einstein found that the theory explaining all the known facts (the Special Theory of Relativity) could be considerably simplified by adopting a non-Euclidean Geometry, he did not hesitate to do so. Thus the General Theory of Relativity was born. Newton's theory was abandoned because of its lack of agreement between predictions and observations. The Special Theory was abandoned because there was a simpler theory explaining the same facts. In one case, the credibility became too low; in the other case, the credibility was high enough, but there was a competing theory with about the same credibility, which was much simpler. This brings out clearly the interplay of credibility (probability) and of simplicity in the verification and consequent acceptance or rejection of theories.

A CASE HISTORY

For the concluding illustration of an application of this method, I turn to one of the greatest masters of the scientific method in history, Mr. Sherlock Holmes.

The one regrettable fact about our record of his activities is that we know him only through the eyes of the somewhat imperfect, if most likable, Dr. Watson. Among other shortcomings we find that the good doctor had his terminology twisted as far as the scientific method is concerned. He has the annoying habit of referring to Mr. Holmes' remarkable inductions (forming of far-reaching theories on scant evidence) as deductions, and of describing the scientific method—so nobly practiced by the immortal master—as the science of deduction. But never mind, let us take one of the fascinating cases and forget the terminology.

Almost any one of the adventures would serve as an illustration. I have, quite arbitrarily, selected the case of "The Red-headed League."

First of all Mr. Holmes collects facts, in this case from the narrative of a Mr. Wilson. It seems that Mr. Wilson's attention was drawn by his assistant to a strange advertisement calling for a red-headed person to collect a fairly nice salary for nominal work. Since Mr. Wilson's pawnshop is not doing well, and since he has a fine head of flaming red hair, he jumps at the opportunity. Although he is one of some thousand applicants, he is fortunate enough to get the job. It turns out that an eccentric English-born American millionaire has left a provision in his will providing for fellow redheads. All that Mr. Wilson has to do is to copy the *Encyclopaedia Brittannica*, only he must do this in an office specifically provided for this purpose. He does this, and collects a handsome fee supplementing his income, until one day (some eight weeks later) he finds the office closed and can find no trace of his employer. During Mr. Holmes' interrogation, Mr. Wilson states that his pawn-shop assistant is a most intelligent young man who has agreed to work for half-pay in order to learn the trade. The only unusual trait of his assistant is his love for photography, which causes him to spend a great deal of time in the cellar developing pictures.

Mr. Holmes has to explain the motivation for the strange employment, and also one or two peculiarities in the assistant. He formulates the hypothesis that the motives are criminal, and that the assistant has a hand in whatever crime is planned. The only purpose the job accomplished was to have Mr. Wilson out of the house; not only was he not defrauded, he actually gained a small sum. Yet there is nothing valuable in the house. The theory formed is that the part of the house which interests the assistant is the cellar (photography being an excuse only) and that the

criminals want to dig a tunnel from the cellar in Mr. Wilson's absence. The fact that he was "fired" suggests that the tunnel is complete, and the fact that the crime has not yet been committed suggests that it will take place in the immediate future.

Mr. Holmes is now in the position of making some of his remarkable predictions. He can predict that there never was a will (which is verified), that there must be an important building easily accessible from the pawnshop's cellar (it turns out to be a bank), and that a robbery is to be attempted in the next few days. He further formulates the theory that it will be Saturday night, to give the robbers extra time to escape before detection. He has all his theories verified by catching the criminals just as they break in through the cellar of the bank.

This is an admirable application of the scientific method to the science of detection. A careful accumulation of facts is followed by the formation of ingenious theories. From these facts logical conclusions are drawn, which are verified one by one, until Mr. Holmes is certain of his theory (or as certain as a human can ever be). Then he can safely predict one more event, this prediction leading to the dramatic climax of the case. To those of you who still feel that there is something miraculous in the scientific method, I will give the master's own answer: "Elementary, dear Watson!"

SUGGESTED READING

Complete references will be found in the Bibliography at the end of the book.

The scientific method.
Cohen and Nagel, Chapter XX.
Lenzen.
Northrop.
Frank [2], Chapters 2, 3.

Scientific theories.
Campbell, Chapters III-V.

Induction.
Cohen and Nagel, Chapter XIV.
Russell, Part Six.

Deduction.
  Black, Chapter II.

Verification.
  Duhem [1].
  Margenau, Chapter VI.

Case histories.
  Conant.
  Doyle.

6

Credibility and Induction

"This conversation is going on a little too fast: let's go back to the last remark but one."

LET US RETURN to the problem raised in the last chapter but one. I have tried to show how fundamental the concept of *credibility* is to the problems of induction and to the whole scientific method. We must now consider this problem in greater detail.

THE PROBLEM OF EXPLICATION

There is no doubt that scientists assign probabilities to theories. They will say that one theory is very probably true, while another is very poorly confirmed and hence not so likely. They will consider two or more alternate hypotheses, and decide which is most likely to be true (on the given evidence). But they have no way of computing these probabilities and there are often considerable arguments as to which of two theories is more probable.

We face the same problem in everyday life. Two racing fans will consult the same dope-sheet, and arrive at different conclusions as to the likelihood of Double Negation winning the fifth race. The first decides that the horse is almost sure to win, and hence bets his shirt, while the latter only assigns an even chance to him, and hence bets on a horse giving longer odds. Double Negation wins by two lengths. Which man was right? The first man has a lot of money to back up his claim, but he has no way of showing that the odds were not even.

10

What Is Science?

"Are we nearly there?" Alice managed to pant out at last.

"Nearly there!" the Queen repeated, "Why we passed it ten minutes ago!"

ONE NATURAL WAY to start a book on the Philosophy of Science is with a definition of Science. However, such a definition would have to be superficial. A much better definition can be given after a great deal of other material has been clarified. By this time the reader is probably ready to ask Alice's question, "Are we nearly there?" We are now in a position to give the same answer that the Queen gave to Alice.

WHAT UNITES SCIENCE

Our alternatives are to define Science by its subject matter, or by its method. But the purpose of Science is to study the whole field of factual knowledge; it has no special topic of its own. Yet we certainly do not classify every study of facts as Science. For example, we refuse to admit Astrology into the family of sciences. Astrology studies facts; it studies the position of stars, and various events in human life, and tries to establish connections between them. The reason that we reject it as a science is not due to its subject matter, but because we consider the methods used by astrologers unscientific. Whenever we find a branch of supposed factual knowledge rejected by Science, it is always on the basis of its method.

Let us look at the other side of the argument. Is every application of the scientific method really a case of Science? The argument to be presented is that every such case can legitimately be called Science, though it is admitted that sometimes this is disputed. There are two types of applications of the method which many scientists would consider outside Science: every-day-life applications and applications by "nonscientists" like criminologists. Custom prevails in these cases: they are not sufficiently "dignified" to be classed as sciences. Perhaps it would be best to state the definition of "Science" to include only important applications of the method, but there is no good way of defining what an important application is. I find it best to consider Science in the broad sense, and call applications to every-day-life Science on an elementary level.

No one can *prove* what the right way of defining "Science" is, but we can argue about the most useful way. Since there is disagreement about the use of the word, our definition cannot agree with all different uses, but we can have several guiding principles: (1) Whenever there is a consensus as to whether some field belongs to Science, our definition must agree with the accepted verdict. (2) In cases where there is considerable disagreement, our definition must settle the dispute. (3) Of the many different ways of settling the disputes, ours must be one that leads to a useful concept. (4) The definition should be as simple as possible. These are the four conditions for the explication of a vague intuitive concept (see Chapter 6). I feel that defining Science by its method is the only definition satisfying these conditions.

The definition of Science by its method certainly agrees with usage whenever there is a clear-cut agreement; it is a simple definition, and it decides disputed cases according to an important principle (whether the procedure used was according to the rules of scientific method), and hence leads to a useful concept of Science. For this reason I shall use the word "Science" to be all knowledge collected by means of the scientific method.

The principal rival to the definition here given would be a definition according to common usage. In this we would poll the leading experts to tell us how they use the word "Science." Such a definition would suffer from all the prejudices of practicing scientists and would be sufficiently vague to destroy its basic fruitfulness. It is a useful precept in explication to err on the side of simplicity rather than on the side of common usage. For this reason I shall reject the definition by common usage and adopt the definition according to method.

There is an apparent circularity in this definition. Doesn't the definition of "Science" use the term "scientific (method)"? It does, but there is no circularity, since the definition of the method was given without using the term "Science." This is a very common procedure; for example, we are likely to define "Mathematics" as the study of mathematical laws, and then we give an independent definition of what such a law is; or we would define "happiness" as the feeling experienced by a happy person. We defined the scientific method by the cycle of induction, deduction, and verification, and by its eternal search for improvement of theories which are only tentatively held. Nowhere in this definition have we used the term "Science," so we can use the definition in turn to define "Science."

Just as a check on the definition we might inquire whether this method is really used in all branches of Science, and the best way of seeing this is through examples chosen from a variety of branches. The example discussed at length in Chapter 5 (discovery of Neptune) was from Physics. From Chemistry we might select the history of the Phlogiston theory. It was held for some time that burning consists in the giving off of a substance, and this substance was called phlogiston. But if this theory is right, then burning should reduce the weight of the burning material. When Lavoisier showed that, on the contrary, it gains in weight, then the old theory was rejected and the new theory (that a substance is taken in during burning) was formed to explain the facts. From this theory many consequences were deduced, and verified. For example, in a closed space there can be only a limited amount of this substance (oxygen) in the air, so when this is used up nothing more can burn in this space until fresh air is let in. We can easily verify this by putting a candle under an inverted glass, and watching it go out long before the candle is

completely burned. This completes one cycle of the scientific method.

From Biology we will choose the discovery of Mendel's laws. Mendel observed over a period of years the variety of plants that he grew in his small garden. From these he formed generalizations concerning the proportion of various traits in the offspring, as determined by the traits of the parents. From these laws we can make sweeping predictions as to the results of breeding experiments, which have been repeatedly confirmed and have led to considerable profit for farmers throughout the world.

It is harder to find good examples of the scientific method in Psychology. However, recent developments in Learning Theory will furnish us with examples. Very interesting theories have been developed by R. R. Bush and F. Mosteller, on the one hand, and by W. K. Estes, on the other hand. These theories start with data collected in simple experiments in which the experimentor attempts to teach a rat, a goldfish, or a human being to learn how to do a task. The two theories provide alternate (and related) models as to how the subject learns to perform these tasks, and from these theories predictions can be made about the outcome of the experiments. Predictions may concern the average time it takes a subject to learn the experiment or the number of errors he will commit before learning how to do the task perfectly, or even whether he will ever learn to do the task perfectly. In the simplest cases these predictions are in excellent agreement with experiments.

The movies provided us with a good case from the Social Sciences. The record of the Kon-Tiki expedition is a gallant testimony to the future the scientific method has in this field. Certain similarities between the ancient traditions of natives in the South Sea islands and of inhabitants of South America led a group of sociologists to form the theory that these natives came from South America, not from the much nearer shores of Asia. This theory was disputed, because it seemed impossible that a thousand years ago these primitive people would have been able to undertake such a journey of several months over the open sea. The scientists deduced from their theories that the type of primitive craft (a loosely constructed raft with poor facilities for steering and only a limited place for storing food) must be capable of completing such a journey. They risked their lives to test this theory, by actually attempting the trip on just such a raft. They found out a great deal not known before. They found a plentiful supply of palatable fish in these unknown waters, they found that the looseness of the raft prevented flooding, and they found that, while steering was impossible, the prevailing currents carried them unerringly to their destination. Over a hundred days later they arrived on these islands and were greeted by the natives who related their old legend according to which the great god Kon-Tiki had brought their ancestors to the islands on just such a craft. Thus they verified their theory (which previously was in general disrepute) and completed a thrilling chapter from the history of the application of the scientific method.

WHAT DIVIDES SCIENCE

We have seen that Science is united not by its subject matter, but by its method. We will now see that it is the subject matter that divides Science into branches.

It is very difficult to explain the reasons for dividing Science the way we do, especially since there is no really good reason. Let us compare the various branches of Science to the various colors. There are five basic colors—red, yellow, green, blue, and violet. Someone else will tell you that there are six or seven, adding orange and/or indigo. That still leaves us with unclassified mixed colors, like pink or brown, not to mention white which is a mixture of all other colors. Even among the basic colors it is difficult to classify all shades. Let us pick a definite shade of blue-green, and we will get quite an argument as to whether it is green or blue. It is even more difficult to distinguish between shades of blue, indigo, and violet.

Quite similarly, we will get an argument as to what the fundamental branches of Science are. Physics, Chemistry, Biology, Psychology, and the Social Sciences form a common list, but others will add Astronomy, and divide the Social Sciences into Economics, Sociology, and Politics. That still leaves us with "mixtures" like Biophysics or like History (insofar as History is made scientific). We have borderline cases too; for example, it is sometimes difficult to say of viruses whether they are alive (and hence belong to the subject-matter of Biology) or whether they are inanimate molecules (and hence belong to Chemistry).

In the case of colors we know that there is a continuous scale, and that explains why there is no natural division into five, seven, or any small number of distinct colors. In addition, we can get an endless variety of secondary shades by mixing the primary ones. We know that it is the difference in wave lengths that differentiates between them, and hence it makes sense to speak of one (pure) shade being closer to a given shade than to another. But any division into colors is highly arbitrary. We do not have as complete a picture of the structure of Science, since our knowledge of theories is incomplete. Scientists will study a group of phenomena which seem related, and try to connect them by means of a theory. Sometimes they fail, and at other times they succeed. In the latter case we have a branch of Science. But there is a great deal of arbitrariness in this procedure. After all, we know that *all* phenomena are connected through The Law of Nature. The laws we are looking for are partial laws, and we are likely to find them where we look for them. Kepler would never have found a connection between the falling of an apple and the motion of the planet Mars, because he never looked for such a connection. It was left to Newton to connect Astronomy with Physics. The ancient Greeks saw a sharp boundary between them; in the heavens circular motion was "the law," while on earth things moved in a straight line. As we learned more, Astronomy became more and more incorporated into Physics. There are still unexplained phenomena, so there is still some excuse for an autonomous Astronomy.

Whenever there is a great deal of arbitrary choice, accidents take a hand in the decision. Whether your university has a separate Astronomy department may depend on whether some rich alumnus has a secret passion for Astrology, but, not being allowed to endow a chair in fortune-telling, he leaves a few millions to finance an Astronomy department. Or, on the other hand, the professor of Astronomy may have the ambition of becoming head of the entire Physics department, and hence there is a merger. I even know of a strange case where Astronomy is part of the Mathematics department.

We witness similar factors in the ordering of shades of color. Suppose you like blue, but dislike green as a rule. If confronted by a shade of green-blue, you are likely to classify it blue or green, according to whether you like the shade or not. Historical factors may have a considerable influence too. If the same person works on two apparently different types of phenomena, these may both be assigned to the same science. Perhaps the fact that Newton studied both Mechanics and Light may account for their becoming branches of Physics, while the fact that he did not contribute to the study of chemical compounds may account for the independent status of Chemistry.

We may summarize the foregoing discussion by stating that Science is divided into branches arbitrarily. If phenomena are connected by known laws, or if some scientist attracts sufficient interest in the study of these phenomena, or for a number of accidental reasons, a group of phenomena is collected into a branch of science. It is dangerous to place too much emphasis on such arbitrary divisions.

REASONS FOR DIVIDING SCIENCE

There are two schools of thought according to which one can motivate the division of Science into branches.

One school believes that eventually a unified Science will be possible and has definite views as to how it will come about. This school will give a division of Science somewhat as follows: Physics, Chemistry, Biology, Psychology, the Social Sciences. They feel that through the progress of Science the "higher" disciplines will become branches of the lower ones and eventually all of Science will be *reduced* to Physics. The Social Sciences are to be reduced to Psychology by explaining the actions of a group on

the basis of the individual psychology of its members. Psychology is to become a part of Biology and the workings of the human mind explained in terms of the workings of the human body. The human body again is to be considered as made up of certain chemicals and subject to various laws of Chemistry, and hence Biology is to become a branch of Chemistry. Finally Chemistry is to be reduced to Physics, a process that is already fairly well completed. In this manner all of Science will be united as a great, expanded Physics.

It is pointed out that in each case the behavior of the whole is explained in terms of the behavior of its parts. This fact is considered significant by this school of thought. Human groups are to be considered as wholes made up of individual parts. These individuals are in turn made up of cells, the cells are made up of chemicals, and these in turn are made of atoms. Many precedents are cited where the behavior of wholes has been explained by Science in terms of the behavior of their parts. It is further pointed out that the lower sciences have wider applicability than the higher ones. For example, the theories of Physics have universal applicability and are used by all sciences. The Law of Gravitation applies equally well to a stone or to a cat, when one of these objects is thrown out of a window. However, Mendel's laws are applicable only to the cat and not to the stone. Thus the ordering given above leads from more general theories to more specialized ones.

We now have a threefold reason for dividing Science, and for ordering the branches in the given manner. First of all there is the expectation of reducing the later branches to the earlier ones. Secondly, the later branches treat wholes whose parts are treated by earlier branches. Finally, there is successively more specialization as we go through the order of the branches.

Although this kind of division certainly is reasonable for many purposes, it must be pointed out that the reasons given are highly oversimplified. For example, the manner in which Biology makes use of the Laws of Gravitation is entirely different from the manner in which the Social Sciences make use of Psychology. The Law of Gravitation will apply to a biological object as a whole, as well as to any one of its parts. However, psychological laws will apply only to the parts of a social group, certainly not to the whole, unless we decide to use "Psychology" in two different senses. Again there is no good reason why the lower sciences should not utilize theories from the higher disciplines. For example, while the Theory of Evolution makes use of Geology (which we may take to be a mixture of Physics and Chemistry), Geology in turn is helped out by results taken from the Theory of Evolution. One of the most useful tools in modern Geology is the dating of rocks by means of the fossils contained in them. The ordering presented is very neat as long as we do not look at too many examples that fail to fit into this order. For example, we might ask whether Nuclear Physics is a separate branch, and we would have a very difficult time placing Astronomy into the neat order. Even the basic principle that wholes are reduced to their parts must be taken as a great oversimplification. In what sense are the objects of Economics reduced to a study of their parts, and how can the reduction of any objects of the higher disciplines to a field theory (such as the Theory of Relativity) fit into this scheme?

Oddly enough, a second school of thought supports a division of Science precisely because they do *not* believe in the possibility of reduction. Vitalists will separate Biology from the lower sciences because they believe that, in some sense, there is a sharp boundary between them. This will be discussed in Chapter 12. Similarly, dualistic philosophers will maintain a division between Psychology and Biology as we will point out in Chapter 13.

I do not want to enter these disputes at the present moment. Let us be satisfied with stating that, for many purposes, a simple division of Science into branches is very useful, but we have found no sufficient reason for assigning deep significance to this classification. We conclude that Science is an enormous area of human research which is united by a common method. Its divisions are for convenience in describing results and do not represent a fundamental feature of Science.

SUGGESTED READING

Complete references will be found in the Bibliography at the end of the book.

Conant, Chapter 2.
Mises, pp. 205-217.
Castell, pp. 242-252.
[E], Volume 1, Book 1.

BIBLIOGRAPHY

This is in no sense a complete bibliography, but simply a summary of books appearing on the suggested reading lists. The dates of publication given are not necessarily those of the first edition, but rather of the edition relatively most easily available.

The following collections of articles are referred to throughout the bibliography by the indicated abbreviations:

[FS] "Readings in Philosophical Analysis," edited by H. Feigl and W. Sellars, Appleton-Century-Crofts, Inc., New York, 1949.

[FB] "Readings in the Philosophy of Science," edited by H. Feigl and M. Brodbeck, Appleton-Century-Crofts, Inc., New York, 1953.

[SH] "Readings in Ethical Theory," edited by W. Sellars and J. Hospers, Appleton-Century-Crofts, Inc., New York, 1952.

[W] "Readings in the Philosophy of Science," edited by P. P. Wiener, Charles Scribner's Sons, New York, 1953.

[E] "International Encyclopaedia of Unified Science," The University of Chicago Press, Chicago

[LLP] "Albert Einstein Philosopher-Scientist," edited by P. A. Schilpp, Library of Living Philosophers, Tudor Publishing Co., New York, 1951.

Abbott, E. A. "Flatland," Blackwell, Oxford, 1944.

Arley, N. and Buck, K. R. "Introduction to the Theory of Probability and Statistics," John Wiley & Sons, Inc., New York, 1950.

Ayer, A. J. "Critique of Ethics," in [SH].

Beardsley, M. C. "Practical Logic," Prentice-Hall, Inc., New York, 1950.

Benjamin, A. C. "An Introduction to the Philosophy of Science," The Macmillan Co., New York, 1937.

Black, M. "Critical Thinking," Prentice-Hall, Inc., New York, 1952.

Bridgman, P. W. "The Logic of Modern Physics," The Macmillan Co., New York, 1927.

Broad, C. D.

[1] "The Mind and Its Place in Nature," Harcourt, Brace & Co., New York, 1925.

[2] "Review of Julian Huxley's *Evolutionary Ethics*" in [FS].
[3] "Some of the Main Problems of Ethics," in [FS].
Campbell, N. R. "What is Science?" Dover Publications, Inc., New York, 1952.
Carnap, R.
[1] "Foundations of Logic and Mathematics," in [E].
[2] "Logical Foundations of Probability," University of Chicago Press, Chicago, 1950.
[3] "Testability and Meaning," *Philosophy of Science*, October 1936 and January 1937.
[4] "Formal and Factual Science," in [FB].
[5] "The Interpretation of Physics," in [FB].
Cassirer, E. "The Problem of Knowledge," Yale University Press, New Haven, 1950.
Castell, A. "An Introduction to Modern Philosophy," The Macmillan Co., New York, 1943.
Cohen, M. R. and Nagel, E. "An Introduction to Logic and Scientific Method," Harcourt, Brace & Co., New York, 1934.
Conant, J. B. "Science and Common Sense," Yale University Press, New Haven, 1951.
Courant, R. and Robbins, H. "What is Mathematics?" Oxford University Press, New York, 1941.
Darwin, C. R. "The Origin of Species," in [W].
Doyle, Sir A. C. "A Treasury of Sherlock Holmes," Hanover House, Garden City, New York, 1955.
Duhem, P.
[1] "Physical Theory and Experiment," in [FB].
[2] "Representation Versus Explanation," in [W].
Eddington, A. S. "The Nature of the Physical World," The Macmillan Co., New York, 1929.
Einstein, A. "The Laws of Science and the Laws of Ethics," in [FB].
Einstein, A. and Infeld, L. "The Evolution of Physics," Simon and Schuster, Inc., New York, 1942.
Feigl, H.
[1] "Operationism and Scientific Method," in [FS].
[2] "Some Remarks on the Meaning of Scientific Explanation," in [FS].
[3] "The Mind-Body Problem in the Development of Logical Empiricism," in [FB].
Feller, W. "An Introduction to Probability Theory and Its Applications," John Wiley & Sons, Inc., New York, 1958.
Frank, P.
[1] "Einstein, Mach, and Logical Positivism," in [LLP].
[2] "Modern Science and Its Philosophy," Harvard University Press, Cambridge, Mass., 1950.

Frankena, W. K. "The Naturalistic Fallacy," in [SH].
Gamow, G.
[1] "One, Two, Three . . . Infinity," The Viking Press, New York, 1947.
[2] "The Evolutionary Universe," *Scientific American*, September 1956.
Hayakawa, S. I. "Language in Thought and Action," Harcourt, Brace & Co., New York, 1949.
Hempel, C. G.
[1] "Fundamentals of Concept Formation in Empirical Science," in [E].
[2] "Geometry and Empirical Science," in [FS] and in [W].
[3] "On the Nature of Mathematical Truth," in [FS] and in [FB].
Hempel, C. G. and Oppenheim, P. "The Logic of Explanation," in [FB].
Hull, C. L. "Value, Valuation, and Natural Science Methodology," *Philosophy of Science*, July 1944.
Hume, D. "An Enquiry Concerning Human Understanding," Open Court, La Salle, Ill., 1946.
Huxley, J. S. "Evolutionary Ethics," Oxford University Press, New York, 1943.
Kemeny, J. G.
[1] "Man Viewed as a Machine," *Scientific American*, April 1955.
[2] "The Use of Simplicity in Induction," *The Philosophical Review*, July 1953.
Kemeny, J. G. and Oppenheim, P. "On Reduction," *Philosophical Studies*, January-February 1956.
Kemeny, J. G., Snell, J. L. and Thompson, G. L. "Introduction to Finite Mathematics," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1957.
Kemeny, J. G., Mirkil, H., Snell, J. L., and Thompson, G. L., "Finite Mathematical Structures," Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1959.
Laslett, P., editor. "The Physical Basis of Mind," Blackwell, Oxford, 1950.
Lenzen, V. F. "Procedures of Empirical Science," in [E].
Lewis, C. I. "An Analysis of Knowledge and Valuation," Open Court, La Salle, Ill., 1947.
Lynd, R. S. "Knowledge for What?" Princeton University Press, Princeton, N.J., 1948.
Malinowski, B. "Magic, Science, and Religion," Beacon Press, Inc., Boston, 1948.
Margenau, H. "The Nature of Physical Reality," McGraw-Hill Book Co., Inc., New York, 1929.

Mill, J. S.
[1] "System of Logic," Longmans, Green & Co., Inc., New York, 1929.
[2] "On the Logic of the Social Sciences," in [W].
Mises, R. von. "Positivism," Harvard University Press, Cambridge, Mass., 1951.
Moore, G. E. "Ethics," Oxford University Press, New York, 1949.
Nagel, E.
[1] "Mechanistic Explanation and Organismic Biology," *Philosophy and Phenomenalistic Research,* March, 1951.
[2] "Principles of the Theory of Probability," in [E].
[3] "Teleological Explanation and Teleological Systems," in [FB].
Neurath, O. "Foundations of the Social Sciences," in [E].
Northrop, F. S. C. "Einstein's Conception of Science," in [LLP].
Pap, A. "Elements of Analytical Philosophy," The Macmillan Co., New York, 1949.
Passmore, J. A. "Can the Social Sciences Be Value-Free?" in [FB].
Planck, M. "The Universe in the Light of Modern Physics," Allen and Unwin, London, 1937.
Poincaré, H. "Non-Euclidean Geometries and the Non-Euclidean World," in [FB].
Quine, W. V. "Two Dogmas of Empiricism," *Philosophical Review,* January, 1951.
Reichenbach, H.
[1] "The Logical Foundations of the Concept of Probability," in [FS] and in [FB].
[2] "On the Justification of Induction," in [FS].
Richards, I. A.
[1] "Basic English and Its Uses," W. W. Horton, New York, 1943.
[2] "The Philosophy of Rhetoric," Oxford University Press, New York, 1936.
Robertson, H. P. "Geometry as a Branch of Physics," in [LLP].
Russell, B. "Human Knowledge, Its Scope and Limits," Simon and Schuster, Inc., New York, 1948.
Ryle, G. "The Concept of Mind," Hutchinsons University Library, New York, 1949.
Schlick, M. "Description and Explanation," in [W].
Shaw, G. B. "Back to Methuselah, a Metabiological Pentateuch," Oxford University Press, New York, 1947.
Simpson, G. G. "The Meaning of Evolution," Mentor Books, New York, 1951.
Stevenson, C. L. "Ethics and Language," Yale University Press, New Haven, 1944.

Sullivan, J. W. N. "Limitations of Science," Mentor Books, New York, 1949.
University of California Associates, "The Freedom of the Will," in [FS].
Weber, M. "Objectivity in Social Sciences," in [W].
Werkmeister, W. H. "A Philosophy of Science," Harper & Brothers, New York, 1940.
Whitehead, A. N. "The Abstract Nature of Mathematics," in [W].
Wiener, N. "What is Cybernetics?" in [W].
Wilder, R. L. "Introduction to the Foundations of Mathematics," John Wiley & Sons, Inc., New York, 1952.

Cognitive Science

An Introduction to Mind and Brain

Daniel Kolak, William Hirstein, Peter Mandik, Jonathan Waskan

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

© 2006

4.4.2 The philosophy of science

4.4.2.1 The influence of logical positivism

If we were to place our bets on who is most likely to have empirical knowledge, we would be wise to bet on scientists. In particular, the phenomenal successes of physics, chemistry, and biology in the twentieth century have attracted the attention of philosophers trying to explain knowledge.

What is scientific knowledge? It seems to involve the formulation of hypotheses and the testing of those hypotheses by carefully controlled observations. Well-tested hypotheses are those most likely to be regarded as true theories. The continued activity of hypothesis formation and hypothesis testing thus gives rise to knowledge of the world and theories that explain how the world works. Philosophers have been interested in explaining in more detail how the processes of hypothesis testing and theoretical explanations work. Many philosophers of science have tried to explain testing and explanation in terms of logic. Foremost among such philosophical explanations of scientific activity were the logical positivists of the early twentieth century. Here we describe the two main outgrowths of logical positivism: the hypothetico-deductive model of theory development and the deductive-nomological model of scientific explanation.

4.4.2.1.1 The *hypothetico-deductive* model of theory development

Scientific knowledge is codified in the form of scientific theories. Prominent theories include the oxygen theory of combustion, and Einstein's General and Special theories of relativity. Where do theories come from? The basic answer that we inherit from the logical positivists is that theories start as hypotheses – educated guesses. Then these hypotheses are subjected to tests, and if they pass the tests, they rise to the level of theory. The hypotheses are confirmed and the process of confirmation bestows justification, and thus, we start with hypothesis and wind up with knowledge. The logical positivists offered a view of the logical structure of theory development and hypothesis testing known as the hypothetico-deductive (H-D) model of theory development. According to the H-D model, after a scientist generates a hypothesis, he then deduces implications of the hypothesis: statements logically derivable from the hypothesis. Next he performs observations and experiments to see if any of these implications of the hypothesis are true. If the implications of the hypothesis are held to be true, the scientist regards the hypothesis itself to be shown true.

Carl Hempel (1965) illustrates H-D with the example of Ignaz Semmelweis' work during the 1840s on childbed fever. Semmelweis observed cases of childbed fever contracted by women who gave birth in his hospital. He noted that cases were especially frequent among women for whom deliveries were handled by physicians instead of midwives. Semmelweis' key insight into the cause of childbed fever came when he observed that a physician came down with similar symptoms upon injuring himself with an instrument during an autopsy. Semmelweis hypothesized that "cadaveric material" on the instrument caused the disease, and that, similarly, the physicians associated with outbreaks of childbed fever had cadaveric material on their hands prior to delivering babies. Semmelweis tested this hypothesis by examining its implications. One implication of the hypothesis that cadaveric material is the cause of childbed fever is that its removal from the hands of physicians would result in a decrease in cases of childbed fever. Semmelweis tested this implication by requiring that physicians wash their hands in chlorinated lime prior to examining patients (which he assumed would remove the cadaveric matter). He observed that groups of women examined by physicians who washed with chlorinated lime had lower incidence of childbed fever than groups of women examined by physicians who did not.

The example of Semmelweis' hypothesis and test conforms to the H-D model in the following way. His hypothesis took the form of a general statement: "Any woman in the hospital who comes down with childbed fever must have been exposed to cadaveric material." An implication of the hypothesis is the statement "If some woman is not exposed to cadaveric material she will not contract childbed fever." Semmelweis tried to set up conditions in which he could observe women not exposed to cadaveric material by having a group of women examined only by physicians who had washed with chlorinated lime. Such women turned out to have lower incidences of childbed fever, thus confirming the initial hypothesis.

Another way in which hypotheses are thought to be confirmed is by way of induction. Semmelweis' observations could be formulated as a series of observation statements, statements of particular states of affairs such as "Jane Doe was exposed to cadaveric material and contracted childbed fever," "Mary Smith was exposed to cadaveric material and contracted childbed fever," and so on. His hypothesis took the form of a law-like general statement: "Any woman exposed to cadaveric material will contract childbed fever," which is a general statement inductively supported by a collection of particular observations.

4.4.2.1.2 *The D-N model of explanation*

One of the main goals of science is to explain things. An idea that goes back as far as Aristotle is that events are explained by showing that they conform to general laws. This idea was developed by the logical positivists into the deductive-nomological (D-N) model of explanation. The gist of the D-N model is that a phenomenon is explained by showing that a description of that phenomenon is deducible from one or more statements of law: statements that take the form of universal generalizations. One example of a law is Newton's law that force equals mass times acceleration. This law may be expressed as a universal generalization of the form, for any quantity of force exerted by an object, that force is the product of that object's mass and that object's acceleration. The explanation of why some particular object exerts the force that it does will involve showing that the relations of its force, mass, and acceleration are derivable from Newton's law. (Note the similarity to foundationalist accounts of justification discussed above.)

Combining the H-D model with the D-N model yields the picture of science depicted in Figure 4.1. According to D-N, particular phenomena are explained by showing how they are deductively derivable from general statements of law. According to H-D, general statements of law are supported by inductive arguments, the premises of which are observation statements: descriptions of observed particular phenomena.

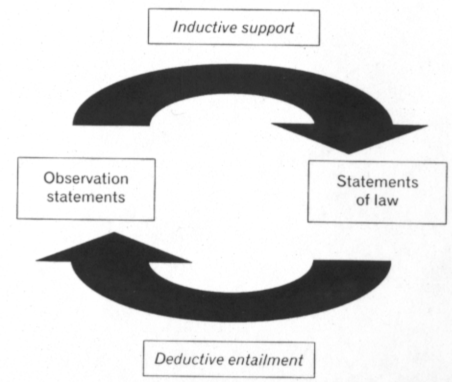


FIGURE 4.1 According to the H-D model, observation statements offer inductive support for statements of law, and according to the D-N model, phenomena are explained by showing that observation statements describing the phenomena are deductively inferred from the statements of law

POPPER'S CRITIQUE OF CONFIRMATION

Karl Popper (Popper 1958 [1935]) attacked the positivists' view that hypotheses could be confirmed. Popper argued that hypotheses could only be falsified.

According to the H-D model, the logic of the relation of hypotheses and tests had the following form:

Premise 1: Hypothesis H entails prediction P.
Premise 2: Prediction P is true.
Conclusion: Therefore hypothesis H is true.

Popper pointed out that any argument of this form embodies fallacious reasoning. The key to seeing this is to realize that statements of the form "If P then Q" are logically

continued

equivalent to "P is sufficient for Q" and "Q is necessary for P." In the above schematic argument, the second premise involves the truth of only one of the necessary conditions of the hypothesis, which is thus insufficient for the truth of the hypothesis. This fallacious form of reasoning is known as the fallacy of affirming the consequent. Here is a more obvious example of the fallacy:

Premise 1: If my car starts then the battery works.
Premise 2: My battery works.
Conclusion: My car will start.

Popper argued that instead of attempting to verify, the best that scientists could do to test for hypotheses was to try to see if they were false. Consider arguments of the following form:

Premise 1: If Hypothesis H is true, then Prediction P would be true.
Premise 2: Prediction P is false.
Conclusion: Therefore hypothesis H is false.

Here we have a valid form of reasoning known as modus tollens. This reasoning is valid because the second premise shows the failure of one of the necessary conditions on the truth of the hypothesis. Thus, Popper argued, scientists should try as hard as possible to devise tests that could falsify hypotheses. The hypothesis that has survived more attempted falsifications is the better hypothesis.

4.4.2.2 *Kuhn: revolutions and paradigms*

The logical positivists had a somewhat ahistorical view of science: they sought to uncover the timeless logical structure to which science should conform. Kuhn (1996), in contrast, saw science as a historically grounded phenomenon. Further, Kuhn saw science as something that changed radically over time. According to Kuhn, scientific theories vary so significantly over time that the findings and theories at one time cannot be meaningfully related to the findings and theories of other times: different theories are thus incommensurable. They constitute different languages that cannot be intertranslated. Instead of viewing the historical progression of science as the progressive accumulation of truths, Kuhn argued for a non-cumulative shift from one paradigm to the next. These changes over time conform to a cyclic pattern that can be broken down into five stages of

(1) immature science
(2) normal mature science
(3) crisis science

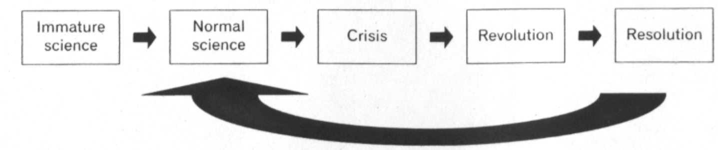


FIGURE 4.2 The cyclic structure of the historical development of science according to Kuhn

(4) revolutionary science
(5) resolutions, which is followed by a return to normal science (see Figure 4.2).

The key notion in understanding Kuhnian philosophy of science is the notion of a *paradigm*. The key stage is normal science and normal science is paradigm-based science. The remaining four of the stages are understood by way of contrast with normal science. What, precisely, a Kuhnian paradigm is supposed to be has been a matter of debate, but the following sketch will suffice.

For Kuhn (1996, p. 20), paradigms are "Universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners." Further, paradigms define what problems and methods are to count as legitimate for succeeding generations of practitioners. Paradigms accomplish these feats in virtue of two essential characteristics (p. 10): first, "Their achievement was sufficiently unprecedented to attract an enduring group of adherents away from competing modes of scientific activity." Second, their achievement "was sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve." Examples of paradigms include, according to Kuhn, Ptolemaic astronomy, Copernican astronomy, Aristotelian dynamics, Newtonian dynamics, corpuscular optics, and wave optics. Arguably, behaviorism constituted a paradigm in psychology that was superseded by cognitive psychology.

Normal science is science that takes place under the guidance of a paradigm: normal science is paradigm-based science. Prior to the arrival of a paradigm, science is immature, according to Kuhn. Immature science is science studying a domain recognizably the same as that studied by paradigm-based successors, but without the utilization of any paradigms. Examples include the cases of optics prior to Newton and electrical research in the first half of the eighteenth century (Kuhn 1996, pp. 12–14). Once a paradigm takes hold, its influence is not exerted forever. A paradigm exerts its influence only as long as a relative consensus as to its applicability exists. When the consensus begins to unravel, a stage of crisis emerges. After a period of crisis, novel approaches of problem solving emerge, thus constituting a scientific revolution. The fruits of revolution are a new paradigm, returning the cycle to a stage of normal science.

According to Kuhn different paradigms are incommensurable and thus choice of one over another cannot be subject to rational procedures. We can understand the incommensurability of paradigms by analogy to different languages, the terms of which cannot be translated into each other. For instance, the term "space" as used in Newtonian physics cannot be translated as the term "space" used in Einsteinian physics. Einsteinians mean different things by "space" than do Newtonians: they use the term in different ways. Unlike Newtonians, Eisteinians hold that space is curved by mass. Kuhn buys into an account of the meaning of theoretical terms whereby the meaning of a term depends on the theory it is embedded in. Where theories diverge, so do the meaning of their terms, regardless of superficial similarities like spelling and pronunciation.

Among Kuhn's arguments that paradigms are not open to rational choice are those that concern the theory-ladenness of perception and observation (recall our discussion from Chapter 3). According to Kuhn, observation statements cannot serve as neutral points of arbitration. There is no theory-neutral observation language because how one perceives the world depends on the theory with which one conceives the world.

Kuhn argues that since paradigms are incommensurable, the history of a scientific discipline is non-cumulative. None of the discoveries and theories of an earlier paradigm can be retained by later paradigms. Thus the progress of science is not the accumulation of scientific truths. Scientists are merely changing their minds over time, not adding to an ever-increasing store of knowledge. Non-cumulativity follows from incommensurability. Since the language of one paradigm cannot be translated into the language of another, the statements held to be true with one paradigm cannot be expressed, let alone judged to be true within another. Adding the theory-ladenness of perception to the equation means that just as theories are not accumulated, neither are observations, since observations depend on theories. Kuhn's thesis of non-cumulativity challenges the traditional view of science a source of progress. Instead, science seems more analogous to changes in clothing fashion: what is valued by one generation is no better or worse than any other. People are merely changing their minds about what they like. The history of science, as viewed through the Kuhnian lens, is of a series of paradigms and revolutions, none bearing any rational relation to any other.

Philosophers and scientists have reacted strongly against many of Kuhn's claims. For instance, Kuhn's hypothesis of incommensurability has been challenged. Some have argued against the view that the meanings of theoretical terms are determined wholly by factors internal to a paradigm, but instead may be determined, at least in part, by causal relations between the term and items in the external world. Putnam (1975) suggests that the meaning of certain scientific terms involves causal relations between the terms and things in the world that they denote. For instance, part of the meaning of water is the substance H_2O that was present when the term *water* was first brought into use to denote that substance. A causal chain leads from current uses of *water* to the initial dubbing of H_2O as water. These causal chains remain constant regardless of a scientist's theory. Thus, water discourse need not be incommensurable between adherents of divergent theories about water. The debate about meaning reflected here is a conflict between internalists and externalists about representational content as discussed in the box on Cartesian skepticism and in Chapter 1 in the discussion of theories of mental representation. We will discuss this further in Chapter 6 in the discussion of the philosophy of language.

Another challenge to Kuhnian incommensurability arises from theorists who propose that the mind is modular. Recall from Chapter 3 that Fodor (1983), for example, argues that many perceptual processes are modular in the sense of being "informationally encapsulated" so that

their outputs are immune to influence by theoretical and other acquired beliefs. Fodor, therefore, contends that observational reports can be treated as univocal even when theorists hold different theories. Though, as discussed in Chapter 3, it is not clear that this solves the sorts of problems raised by Kuhn.

Two Dogmas of Empiricism

Willard Van Orman Quine

Originally published in *The Philosophical Review* 60 (1951): 20-43. Reprinted in W.V.O. Quine, *From a Logical Point of View* (Harvard University Press, 1953; second, revised, edition 1961), with the following alterations: "The version printed here diverges from the original in footnotes and in other minor respects: __1 and 6 have been abridged where they encroach on the preceding essay ["On What There Is"], and __3-4 have been expanded at points."

Except for minor changes, additions and deletions are indicated in interspersed tables. I wish to thank Torstein Lindaas for bringing to my attention the need to distinguish more carefully the 1951 and the 1961 versions. Endnotes ending with an "a" are in the 1951 version; "b" in the 1961 version. (Andrew Chrucky, Feb. 15, 2000)

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

1. BACKGROUND FOR ANALYTICITY

Kant's cleavage between analytic and synthetic truths was foreshadowed in Hume's distinction between relations of ideas and matters of fact, and in Leibniz's distinction between truths of reason and truths of fact. Leibniz spoke of the truths of reason as true in all possible worlds. Picturesqueness aside, this is to say that the truths of reason are those which could not possibly be false. In the same vein we hear analytic statements defined as statements whose denials are self-contradictory. But this definition has small explanatory value; for the notion of self-contradictoriness, in the quite broad sense needed for this definition of analyticity, stands in exactly the same need of clarification as does the notion of analyticity itself.[1a] The two notions are the two sides of a single dubious coin.

Kant conceived of an analytic statement as one that attributes to its subject no more than is already conceptually contained in the subject. This formulation has two shortcomings: it limits itself to statements of subject-predicate form, and it appeals to a notion of containment which is left at a metaphorical level. But Kant's intent, evident more from the use he makes of the notion of analyticity than from his definition of it, can be restated thus: a statement is analytic when it is true by virtue of meanings and independently of fact. Pursuing this line, let us examine the concept of *meaning* which is presupposed.

(1951)	(1961)
We must observe to begin with that meaning is not to be identified with naming or reference. Consider Frege's example of 'Evening Star' and 'Morning Star.' Understood not merely as a recurrent evening apparition but as a body, the Evening Star is the planet Venus, and the Morning Star is the same. The two singular terms name the same thing. But the meanings must be treated as distinct, since the identity 'Evening Star = Morning Star' is a statement of fact established by astronomical observation. If 'Evening Star' and 'Morning Star' were alike in meaning, the identity 'Evening Star = Morning Star' would be analytic. Again there is Russell's example of 'Scott' and 'the author of Waverly.' Analysis of the meanings of words was by no means sufficient to reveal to George IV that the person named by these two singular terms was one and the same. The distinction between meaning and naming is no less important at the level of abstract terms. The terms '9' and 'the number of planets' name one and the same abstract entity but presumably must be regarded as unlike in meaning; for astronomical observation was needed, and not mere reflection on meanings, to determine the sameness of the entity in question. Thus far we have been considering singular terms.	Meaning, let us remember, is not to be identified with naming.[1b] Frege's example of 'Evening Star' and 'Morning Star' and Russell's of 'Scott' and 'the author of *Waverly*', illustrate that terms can name the same thing but differ in meaning. The distinction between meaning and naming is no less important at the level of abstract terms. The terms '9' and 'the number of the planets' name one and the same abstract entity but presumably must be regarded as unlike in meaning; for astronomical observation was needed, and not mere reflection on meanings, to determine the sameness of the entity in question. The above examples consist of singular terms, concrete and abstract.

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

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

The Aristotelian notion of essence was the forerunner, no doubt, of the modern notion of intension or meaning. For Aristotle it was essential in men to be rational, accidental to be two-legged. But there is an important difference between this attitude and the doctrine of meaning. From the latter point of view it may indeed be conceded (if only for the sake of argument) that rationality is involved in the meaning of the word 'man' while two-leggedness is not; but two-leggedness may at the same time be viewed as involved in the meaning of 'biped' while rationality is not. Thus from the point of view of the doctrine of meaning it makes no sense to say of the actual individual, who is at once a man and a biped, that his rationality is essential and his two-leggedness accidental or vice versa. Things had essences, for Aristotle, but only linguistic forms have meanings. Meaning is what essence becomes when it is divorced from the object of reference and wedded to the word.

For the theory of meaning the most conspicuous question is as to the nature of its objects: what sort of things are meanings?

(1951)	
They are evidently intended to be ideas, somehow -- mental ideas for some semanticists, Platonic ideas for others. Objects of either sort are so elusive, not to say debatable, that there seems little hope of erecting a fruitful science about them. It is not even clear, granted meanings, when we have two and when we have one; it is not clear when linguistic forms should be regarded as *synonymous*, or alike in meaning, and when they should not. If a standard of synonymy should be arrived at, we may reasonably expect that the appeal to meanings as entities will not have played a very useful part in the enterprise.	

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

(1951)	(1961)
The description of analyticity as truth by virtue of meanings started us off in pursuit	The problem of analyticity confronts us anew.

of a concept of meaning. But now we have abandoned the thought of any special realm of entities called meanings. So the problem of analyticity confronts us anew.	

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

(1) No unmarried man is married.

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

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

(2) No bachelor is married.

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

In recent years Carnap has tended to explain analyticity by appeal to what he calls state-descriptions.[2a] [4b] A state-description is any exhaustive assignment of truth values to the atomic, or noncompound, statements of the language. All other statements of the language are, Carnap assumes, built up of their component clauses by means of the familiar logical devices, in such a way that the truth value of any complex statement is fixed for each state-description by specifiable logical laws. A statement is then explained as analytic when it comes out true under every state-description. This account is an adaptation of Leibniz's "true in all possible worlds." But note that this version of analyticity serves its purpose only if the atomic statements of the language are, unlike 'John is a bachelor' and 'John is married,' mutually independent. Otherwise there would be a state-description which assigned truth to 'John is a bachelor' and falsity to 'John is married,' and consequently 'All bachelors are married' would turn out synthetic rather than analytic under the proposed criterion. Thus the criterion of analyticity in terms of state-descriptions serves only for languages devoid of extralogical synonym-pairs, such as 'bachelor' and 'unmarried man': synonym-pairs of the type which give rise to the "second class" of analytic statements. The criterion in terms of state-descriptions is a reconstruction at best of logical truth.

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

II. DEFINITION

There are those who find it soothing to say that the analytic statements of the second class reduce to those of the first class, the logical truths, by *definition*; 'bachelor,' for example, is *defined* as 'unmarried man.' But how do we find that 'bachelor' is defined as 'unmarried man'? Who defined it thus, and when? Are we to appeal to the nearest dictionary, and accept the lexicographer's formulation as law? Clearly this would be to put the cart before the horse. The lexicographer is an empirical scientist, whose business is the recording of antecedent facts; and if he glosses 'bachelor' as 'unmarried man' it is because of his belief that there is a relation of synonymy between these forms, implicit in general or preferred usage prior to his own work. The notion of synonymy presupposed here has still to be clarified, presumably in terms relating to linguistic behavior. Certainly the "definition" which is the lexicographer's report of an observed synonymy cannot be taken as the ground of the synonymy.

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

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

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

Two alternative definientia may be equally appropriate for the purposes of a given task of explication and yet not be synonymous with each other; for they may serve interchangeably within the favored contexts but diverge elsewhere. By cleaving to one of these definientia rather than the other, a definition of explicative kind generates, by fiat, a relationship of synonymy between definiendum and definiens which did not hold before. But such a definition still owes its explicative function, as seen, to pre-existing synonymies.

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

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

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

Both sorts of economy, though prima facie incompatible, are valuable in their separate ways. The custom has consequently arisen of combining both sorts of economy by forging in effect two languages, the one a part of the other. The inclusive language, though redundant in grammar and vocabulary, is economical in message lengths, while the part, called *primitive notation*, is economical in grammar and vocabulary. Whole and part are correlated by rules of translation whereby each idiom not in primitive notation is equated to some complex built up of primitive notation. These rules of translation are the so-called *definitions* which appear in formalized systems. They are best viewed not as adjuncts to one language but as correlations between two languages, the one a part of the other.

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

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

III. INTERCHANGEABILITY

A natural suggestion, deserving close examination, is that the synonymy of two linguistic forms consists simply in their interchangeability in all contexts without change of truth value; interchangeability, in Leibniz's phrase, *salva veritate*.[5] [6b] Note that synonyms so conceived need not even be free from vagueness, as long as the vaguenesses match.

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

> 'Bachelor' has less than ten letters.

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

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

(3) All and only bachelors are unmarried men

is analytic.[3a] [7b]

What we need is an account of cognitive synonymy not presupposing analyticity -- if we are to explain analyticity conversely with help of cognitive synonymy as undertaken in Section 1. And indeed such an independent account of cognitive synonymy is at present up for consideration, namely, interchangeability *salva veritate* everywhere except within words. The question before us, to resume the thread at last, is whether such interchangeability is a sufficient condition for cognitive synonymy. We can quickly assure ourselves that it is, by examples of the following sort. The statement:

(4) Necessarily all and only bachelors are bachelors

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

(5) Necessarily, all and only bachelors are unmarried men

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

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

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

Interchangeability *salva veritate* is meaningless until relativized to a language whose extent is specified in relevant respects. Suppose now we consider a language containing just the following materials. There is an indefinitely large stock of one- and many-place predicates,

(1951)	(1961)
There is an indefinitely large stock of one- and many-place predicates,	There is an indefinitely large stock of one-place predicates, (for example, 'F' where 'Fx' means that x is a man) and many-placed predicates (for example, 'G' where 'Gxy' means that x loves y,

mostly having to do with extralogical subject matter. The rest of the language is logical. The atomic sentences consist each of a predicate followed by one or more variables 'x', 'y', etc.; and the complex sentences are built up of atomic ones by truth functions ('not', 'and', 'or', etc.) and quantification.[8b] In effect such a language enjoys the benefits also of descriptions and class names and indeed singular terms generally, these being contextually definable in known ways.[4a] [9b]

	(1961)
	Even abstract singular terms naming classes, classes of classes, etc., are contextually definable in case the assumed stock of predicates includes the two-place predicate of class membership.[10b]

(1951)	(1961)
Such a language can be adequate to classical mathematics and indeed to scientific discourse generally, except in so far as the latter involves debatable devices such as modal adverbs and contrary-to-fact conditionals.	Such a language can be adequate to classical mathematics and indeed to scientific discourse generally, except in so far as the latter involves debatable devices such as contrary-to-fact conditionals or modal adverbs like 'necessarily'.[11b]

Now a language of this type is *extensional*, in this sense: any two predicates which *agree extensionally* (i.e., are true of the same objects) are interchangeable *salva veritate*.[12b]

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

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

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

The effort to explain cognitive synonymy first, for the sake of deriving analyticity from it afterward as in Section I, is perhaps the wrong approach. Instead we might try explaining analyticity somehow without appeal to cognitive synonymy. Afterward we could doubtless derive cognitive synonymy from analyticity satisfactorily enough if desired. We have seen that cognitive synonymy of 'bachelor' and 'unmarried man' can be explained as analyticity of (3). The same explanation works for any pair of one-place predicates, of course, and it can be extended in obvious fashion to many-place predicates. Other syntactical categories can also he accommodated in fairly parallel fashion. Singular terms may be said to be cognitively synonymous when the statement of identity formed by putting '=' between them is analytic. Statements may be said simply to be cognitively synonymous when their biconditional (the result of joining them by 'if and only if') is analytic.[5a] [13b] If we care to lump all categories into a single formulation, at the expense of assuming again the notion of "word" which was appealed to early in this section, we can describe any two linguistic forms as cognitively synonymous when the two forms are interchangeable (apart from occurrences within "words") *salva* (no longer *veritate* but) *analyticitate*. Certain technical questions arise, indeed, over cases of ambiguity or homonymy; let us not pause for them, however, for we are already digressing. Let us rather turn our backs on the problem of synonymy and address ourselves anew to that of analyticity.

IV. SEMANTICAL RULES

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

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

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

The notion of analyticity about which we are worrying is a purported relation between statements and languages: a statement S is said to be *analytic for* a language L, and the problem is to make sense of this relation generally, for example, for variable 'S' and 'L.' The point that I want to make is that the gravity of this problem is not perceptibly less for artificial languages than for natural ones. The problem of making sense of the idiom 'S *is analytic for* L,' with variable 'S' and 'L,' retains its stubbornness even if we limit the range of the variable 'L' to artificial languages. Let me now try to make this point evident.

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

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

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

Actually we do know enough about the intended significance of 'analytic' to know that analytic statements are supposed to be true. Let us then turn to a second form of semantical rule, which says not that such and such statements are analytic but simply that such and such statements are included among the truths. Such a rule is not subject to the criticism of containing the un-understood word 'analytic'; and we may grant for the sake of argument that there is no difficulty over the broader term 'true.' A semantical rule of this second type, a rule of truth, is not supposed to specify all the truths of the language; it merely stipulates, recursively or otherwise, a certain multitude of statements which, along with others unspecified, are to count as true. Such a rule may be conceded to be quite clear. Derivatively, afterward, analyticity can be demarcated thus: a statement is analytic if it is (not merely true but) true according to the semantical rule.

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

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

	(1961) It may be instructive to compare the notion of semantical rule with that of postulate. Relative to the given set of postulates, it is easy to say that what a postulate is: it is a member of the set. Relative to a given set of semantical rules, it is equally easy to say what a semantical rule is. But given simply a notation, mathematical or otherwise, and indeed as thoroughly understood a notation as you please in point of the translation or truth conditions of its statements, who can say which of its true statements rank as postulates? Obviously the question is meaningless -- as meaningless as asking which points in Ohio are starting points. Any finite (or effectively specifiable infinite) selection of statements (preferably true ones, perhaps) is as much *a* set of postulates as any other. The word 'postulate' is significant only relative to an act of inquiry; we apply the word to a set of statements just in so far as we happen, for the year or the argument, to be thinking of those statements which can be reached from them by some set of trasformations to

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

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

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

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

Appeal to hypothetical languages of an artificially simple kind could conceivably bc useful in clarifying analyticity, if the mental or behavioral or cultural factors relevant to analyticity -- whatever they may be -- were somehow sketched into the simplified model. But a model which takes analyticity merely as an irreducible character is unlikely to throw light on the problem of explicating analyticity.

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

V. THE VERIFICATION THEORY AND REDUCTIONISM

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

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

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

This is an account of cognitive synonymy not of linguistic forms generally, but of statements.[6a] [15b] However, from the concept of synonymy of statements we could derive the concept of synonymy for other linguistic forms, by considerations somewhat similar to those at the end of Section III. Assuming the notion of "word," indeed, we could explain any two forms as synonymous when the putting of the one form for an occurrence of the other in any statement (apart from occurrences within "words") yields a synonymous statement. Finally, given the concept of synonymy thus for linguistic forms generally, we could define analyticity in terms of synonymy and logical truth as in Section I. For that matter, we could define analyticity more simply in terms of just synonymy of statements together with logical truth; it is not necessary to appeal to synonymy of linguistic forms other than statements. For a statement may be described as analytic simply when it is synonymous with a logically true statement.

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

The most naive view of the relationship is that it is one of direct report. This is *radical reductionism*. Every meaningful statement is held to be translatable into a statement (true or false) about immediate experience. Radical reductionism, in one form or another, well antedates the verification theory of meaning explicitly so called. Thus Locke and Hume held that every idea must either originate directly in sense experience or else be compounded of ideas thus originating; and taking a hint from Tooke[7a] we might rephrase this doctrine in semantical jargon by saying that a term, to be significant at all, must be either a name of a sense datum or a compound of such names or an abbreviation of such a compound. So stated, the doctrine remains ambiguous as between sense data as sensory events and sense data as sensory qualities; and it remains vague as to the admissible ways of compounding. Moreover, the doctrine is unnecessarily and intolerably restrictive in the term-by-term critique which it imposes. More reasonably, and without yet exceeding the limits of what I have called radical reductionism, we may take full statements as our significant units -- thus demanding that our statements as wholes be translatable into sense-datum language, but not that they be translatable term by term.

(1951)	(1961)
This emendation would unquestionably have been welcome to Locke and Hume and Tooke, but historically it had to await two intermediate developments. One of these developments was the increasing emphasis on verification or confirmation, which came with the explicitly so-called verification theory of meaning. The objects of verification or confirmation being statements, this emphasis gave the statement an ascendancy over the word or term as unit of significant discourse. The other development, consequent upon the first, was Russell's discovery of the concept of incomplete symbols defined in use.	This emendation would unquestionably have been welcome to Locke and Hume and Tooke, but historically it had to await an important reorientation in semantics -- the reorientation whereby the primary vehicle of meaning came to be seen no longer in the term but in the statement. This reorientation, explicit in Frege (Gottlieb Frege, Foundations of Arithmetic (New York: Philosophical Library, 1950). Reprinted in Grundlagen der Arithmetik (Breslau, 1884) with English translations in parallel. Section 60), underlies Russell'a concept of incomplete symbols defined in use;[16b] also it is implicit in the verification theory of meaning, since the objects of verification are statements.

Radical reductionism, conceived now with statements as units, sets itself the task of specifying a sense-datum language and showing how to translate the rest of significant discourse, statement by statement, into it. Carnap embarked on this project in the *Aufbau*.[8a]

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

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

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

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

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

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

The dogma of reductionism, even in its attenuated form, is intimately connected with the other dogma: that there is a cleavage between the analytic and the synthetic. We have found ourselves led, indeed, from the latter problem to the former through the verification theory of meaning. More directly, the one dogma clearly supports the other in this way: as long as it is taken to be significant in general to speak of the confirmation and infirmation of a statement, it seems significant to speak also of a limiting kind of statement which is vacuously confirmed, *ipso facto*, come what may; and such a statement is analytic.

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

(1951)	(1961)
Russell's concept of definition in use was, as remarked, an advance over the impossible term-by-term empiricism of Locke and Hume. The statement, rather than the term, came with Russell to be recognized as the unit accountable to an empiricist critique.	The idea of defining a symbol in use was, as remarked, an advance over the impossible term-by-term empiricism of Locke and Hume. The statement, rather than the term, came with Frege to be recognized as the unit accountable to an empiricist critique.

But what I am now urging is that even in taking the statement as unit we have drawn our grid too finely. The unit of empirical significance is the whole of science.

VI. EMPIRICISM WITHOUT THE DOGMAS

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Re-evaluation of some statements entails re-evaluation of others, because of their logical interconnections -- the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having re-evaluated one statement we must re-evaluate some others, whether they be statements logically connected with the first or whether they be the statements of logical connections themselves. But the total field is so undetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.

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

For vividness I have been speaking in terms of varying distances from a sensory periphery. Let me try now to clarify this notion without metaphor. Certain statements, though about physical objects and not sense experience, seem peculiarly germane to sense experience -- and in a selective way: some statements to some experiences, others to others. Such statements, especially germane to particular experiences, I picture as near the periphery. But in this relation of "germaneness" I envisage nothing more than a loose association reflecting the relative likelihood, in practice, of our choosing one statement rather than another for revision in the event of recalcitrant experience. For example, we can imagine recalcitrant experiences to which we would surely be inclined to accommodate our system by re-evaluating just the statement that there are brick houses on Elm Street, together with related statements on the same topic. We can imagine other recalcitrant experiences to which we would be inclined to accommodate our system by re-evaluating just the statement that there are no centaurs, along with kindred statements. A recalcitrant experience can, I have already urged, bc accommodated by any of various alternative re-evaluations in various alternative quarters of the total system; but, in the cases which we are now imagining, our natural tendency to disturb the total system as little as possible would lead us to focus our revisions upon these specific statements concerning brick houses or centaurs. These statements are felt, therefore, to have a sharper empirical reference than highly theoretical statements of physics or logic or ontology. The latter statements may be thought of as relatively centrally located within the total network, meaning merely that little preferential connection with any particular sense data obtrudes itself.

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

(1951)

Imagine, for the sake of analogy, that we are given the rational numbers. We develop an algebraic theory for reasoning about them, but we find it inconveniently complex, because certain functions such as square root lack values for some arguments. Then it is discovered that the rules of our algebra can be much simplified by conceptually augmenting our ontology with some mythical entities, to be called irrational numbers. All we continue to be really interested in, first and last, are rational numbers; but we find that we can commonly get from one law about rational numbers to another much more quickly and simply by pretending that the irrational numbers are there too.

I think this a fair account of the introduction of irrational numbers and other extensions of the number system. The fact that the mythical status of irrational numbers eventually gave way to the Dedekind- Russell version of them as certain infinite classes of ratios is irrelevant to my analogy. That version is impossible anyway as long as reality is limited to the rational numbers and not extended to classes of them.

Now I suggest that experience is analogous to the rational numbers and that the physical objects, in analogy to the irrational numbers, are posits which serve merely to simplify our treatment of experience. The physical objects are no more reducible to experience than the irrational numbers to rational numbers, but their incorporation into the theory enables us to get more easily from one statement about experience to another.

The salient differences between the positing of physical objects and the positing of irrational numbers are, I think, just two. First, the factor of simplication is more overwhelming in the case of physical objects than in the numerical case. Second, the positing of physical objects is far more archaic, being indeed coeval, I expect, with language itself. For language is social and so depends for its development upon intersubjective reference.

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

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

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

Ontological questions, under this view, are on a par with questions of natural science.[20b] Consider the question whether to countenance classes as entities. This, as I have argued elsewhere,[9a21b] is the question whether to quantify with respect to variables which take classes as values. Now Carnap ["Empiricism, semantics, and ontology," *Revue internationale de philosophie* 4 (1950), 20-40.] has maintained[10a] that this is a question not of matters of fact but of choosing a convenient language form, a convenient conceptual scheme or framework for science. With this I agree, but only on the proviso that the same be conceded regarding scientific hypotheses generally. Carnap has recognized[11a] that he is able to preserve a double standard for ontological questions and scientific hypotheses only by assuming an absolute distinction between the analytic and the synthetic; and I need not say again that this is a distinction which I reject. [22b]

(1951) Some issues do, I grant, seem more a question of convenient conceptual scheme and others more a question of brute fact.	

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

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

Notes

1a. See White, "The Analytic and the Synthetic: An Untenable Dualism," *John Dewey: Philosopher of Science and Freedom* (New York: 1950), p. 324. [Back]

1b. See "On What There Is", p. 9. [Back]

2a. R. Carnap, *Meaning and Necessity* (Chicago, 1947), pp. 9 ff.; *Logical Foundations of Probability* (Chicago, 1950), pp. 70 ff. [Back]

2b. See "On What There Is", p. 10. [Back]

3a. This is cognitive synonymy in a primary, broad sense. Carnap (*Meaning and Necessity*, pp. 56 ff.) and Lewis (*Analysis of Knowledge and Valuation* [La Salle, Ill., 1946], pp. 83 ff.) have suggested how, once this notion is at hand, a narrower sense of cognitive synonymy which is preferable for some purposes can in turn be derived. But this special ramification of concept-building lies aside from the present purposes and must not be confused with the broad sort of cognitive synonymy here concerned. [Back]

3b. See "On What There Is", p. 11f, and "The Problem of Meaning in Linguistics," p. 48f.[Back]

4a. See, for example my *Mathematical Logic* (New York, 1949; Cambridge, Mass., 1947), sec. 24, 26, 27; or *Methods of Logic* (New York, 1950), sec. 37 ff. [Back]

4b. Rudolf Carnap, *Meaning and Necessity* (Chicago: University of Chicago Press, 1947), pp. 9ff; *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950).[Back]

5a. The 'if and only if' itself is intended in the truth functional sense. See Carnap, *Meaning and Necessity*, p. 14. [Back]

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

6a. The doctrine can indeed be formulated with terms rather than statements as the units. Thus C. I. Lewis describes the meaning of a term as "a criterion in mind, by reference to which one is able to apply or refuse to apply the expression in question in the case of presented, or imagined things or situations" (Carnap, *Meaning and Necessity*, p. 133.). [Back]

6b. Cf. C.I. Lewis, A Survey of Symbolic Logic (Berkeley, 1918), p. 373.[Back]

7a. John Horne Tooke, *The Diversions of Purely* (London, 1776; Boston, 1806), I, ch. ii. [Back]

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

8a. R. Carnap, *Der logische Aufbau der Welt* (Berlin, 1928). [Back]

8b. Pp. 81ff, "New Foundations for Mathematical Logic," contains a description of just such a language, except that there happens to be just one predicate, the two-place predicate 'ε'. [Back]

9a. For example, in "Notes on Existence and Necessity," *Journal of Philosophy*, 11 (1943), 113-127. [Back]

9b. See "On What There Is," pp. 5-8; see also "New Foundations for Mathematical Logic," p. 85f; "Meaning and Existential Inference," p. 166f. [Back]

10a. Carnap, "Empiricism, Semantics, and Ontology," *Revue internationale de philosophie*, 4 (1950), 20-40. [Back]

10b. See "New Foundations for Mathematical Logic," p. 87.[Back]

11a. Carnap, "Empiricism, Semantics, and Ontology," p. 32. [Back]

11b. On such devices see also "Reference and Modality."[Back]

12b. This is the substance of Quine, Mathematical Logic (1940; rev. ed., 1951).[Back]

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

14b. The foregoing paragraph was not part of the present essay as originally published. It was prompted by Martin, (R. M. Martin, "On 'analytic'," *Philosophical Studies* 3 (1952), 42-47. [Back]

There could be no fairer destiny for any . . . theory than that it should point the way to a more comprehensive theory in which it lives on, as a limiting case.

ALBERT EINSTEIN

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.

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

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

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

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

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

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

I found that those of my friends who were admirers of Marx, Freud, and Adler, were impressed by a number of points common to these theories, and especially by their apparent *explanatory power*. These theories appeared to be able to explain practically everything that happened within the fields to which they referred. The study of any of them seemed to have the effect of an

intellectual conversion or revelation, opening your eyes to a new truth hidden from those not yet initiated. Once your eyes were thus opened you saw confirming instances everywhere: the world was full of *verifications* of the theory. Whatever happened always confirmed it. Thus its truth appeared manifest; and unbelievers were clearly people who did not want to see the manifest truth; who refused to see it, either because it was against their class interest, or because of their repressions which were still 'un-analysed' and crying aloud for treatment.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

II

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 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 which was first published in *Erkenntnis*, 3, 1933, pp. 426 f.

Concerning my never published book mentioned here in the text, see R. Carnap's paper '*Ueber Protokollstäze*' (On Protocol-Sentences), *Erkenntnis*, 3, 1932, pp. 215–28 where he gives an outline of my theory on pp. 223–8, and accepts it. He calls my theory 'procedure B', and says (p. 224, top): 'Starting from a point of view different from Neurath's' (who developed what Carnap calls on p. 223 'procedure A'), 'Popper developed procedure B as

and elsewhere I described the problem of meaning as a pseudo-problem, in contrast to the problem of demarcation. But my contribution was classified by members of the Circle as a proposal to replace the verifiability criterion of *meaning* by a falsifiability criterion of *meaning*—which effectively made nonsense of my views.[6] My protests that I was trying to solve, not their pseudo-problem of meaning, but the problem of demarcation, were of no avail.

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

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

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

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

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

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

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—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.

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.

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.*

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 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.

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.)

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 my-

[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 $C(t, t) \leq +1$. 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.

self 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.

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)	$P(xy) > P(yx)$	(Commutation)
(A2)	$P(x(yz)) > P((xy)z)$	(Association)
(A3)	$P(xx) > P(x)$	(Tautology)
(B1)	$P(x) > P(xy)$	(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) $P(x,z) \geqslant P(xy,z)$
(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, 1957, 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 and Addendum to vol. ii (Fourth Edition). 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.

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]

[31] (13) was added in 1961. Since 1953, when this lecture was delivered, and 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.

CONJECTURES AND REFUTATIONS

The Growth of Scientific Knowledge

by

KARL R. POPPER

HARPER TORCHBOOKS
Harper & Row, Publishers
New York and Evanston

Thomas Kuhn (1962)

The Structure of Scientific Revolutions

Source: *The Structure of Scientific Revolutions* (1962) publ. University of Chicago Press, 1962. One chapter plus one postscript reproduced here;
Transcribed: by Andy Blunden in 1998; proofed and corrected March 2005.

IX. The Nature and Necessity of Scientific Revolutions

These remarks permit us at last to consider the problems that provide this essay with its title. What are scientific revolutions, and what is their function in scientific development? Much of the answer to these questions has been anticipated in earlier sections. In particular, the preceding discussion has indicated that scientific revolutions are here taken to be those non-cumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one. There is more to be said, however, and an essential part of it can be introduced by asking one further question. Why should a change of paradigm be called a revolution? In the face of the vast and essential differences between political and scientific development, what parallelism can justify the metaphor that finds revolutions in both?

One aspect of the parallelism must already be apparent. Political revolutions are inaugurated by a growing sense, often restricted to a segment of the political community, that existing institutions have ceased adequately to meet the problems posed by an environment that they have in part created. In much the same way, scientific revolutions are inaugurated by a growing sense, again often restricted to a narrow subdivision of the scientific community, that an existing paradigm has ceased to function adequately in the exploration of an aspect of nature to which that paradigm itself had previously led the way. In both political and scientific development the sense of malfunction that can lead to crisis is prerequisite to revolution. Furthermore, though it admittedly strains the metaphor, that parallelism holds not only for the major paradigm changes, like those attributable to Copernicus and Lavoisier, but also for the far smaller ones associated with the assimilation of a new sort of phenomenon, like oxygen or X-rays. Scientific revolutions, as we noted at the end of Section V, need seem revolutionary only to those whose paradigms are affected by them. To outsiders they may, like the Balkan revolutions of the early twentieth century, seem normal parts of the developmental process. Astronomers, for example, could accept X-rays as a mere addition to knowledge, for their paradigms were unaffected by the existence of the new radiation. But for men like Kelvin, Crookes, and Roentgen, whose research dealt with radiation theory or with cathode ray tubes, the emergence of X-rays necessarily violated one paradigm as it created another. That is why these rays could be discovered only through something's first going wrong with normal research.

This genetic aspect of the parallel between political and scientific development should no longer be open to doubt. The parallel has, however, a second and more profound aspect upon which the significance of the first depends. Political revolutions aim to change political institutions in ways that those institutions themselves prohibit. Their success therefore necessitates the partial relinquishment of one set of institutions in favour of another, and in the interim, society is not fully governed by institutions at all. Initially it is crisis alone that attenuates the role of political institutions as we have already seen it attenuate the role of paradigms. In increasing numbers individuals become increasingly estranged from political life and behave more and more eccentrically within it. Then, as the crisis deepens, many of these individuals commit themselves to some concrete proposal for the reconstruction of society in a new institutional framework. At that point the society is divided into competing camps or parties, one seeking to defend the old institutional constellation, the others seeking to institute some new one. And, once that polarisation has occurred, **political recourse fails**. Because they differ about the institutional matrix within which political change is to be achieved and evaluated, because they acknowledge no supra-institutional framework for the adjudication of revolutionary difference, the parties to a revolutionary conflict must finally resort to the techniques of mass persuasion, often including force. Though revolutions have had a vital role in the evolution of political institutions, that role depends upon their being partially extrapolitical or extrainstitutional events. The remainder of this essay aims to demonstrate that the historical study of paradigm change reveals very similar characteristics in the evolution of the sciences. Like the choice between competing political institutions, that between competing paradigms proves to be a choice between incompatible modes of community life. Because it has that character, the choice is not and cannot be determined merely by the evaluative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue. When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defence.

The resulting circularity does not, of course, make the arguments wrong or even ineffectual. The man who premises a paradigm when arguing in its defence can nonetheless provide a clear exhibit of what scientific practice will be like for those who adopt the new view of nature. That exhibit can be immensely persuasive, often compellingly so. Yet, whatever its force, the status of the circular argument is only that of persuasion. It cannot be made logically or even probabilistically compelling for those who refuse to step into the circle. The premises and values shared by the two parties to a debate over paradigms are not sufficiently extensive for that. As in political revolutions, so in paradigm choice – there is no standard higher than the assent of the relevant community. To discover how scientific revolutions are effected, we shall therefore have to examine not only the impact of nature and of logic, but also the techniques of persuasive argumentation effective within the quite special groups that constitute the community of scientists.

To discover why this issue of paradigm choice can never be unequivocally settled by logic and experiment alone, we must shortly examine the nature of the differences that separate the proponents of a traditional paradigm from their revolutionary successors. That examination is the principal object of this section and the next. We have, however, already noted numerous examples of such differences, and no one will doubt that history can supply many others. What is more likely to be doubted than their existence – and what must therefore be considered first – is that such examples provide essential information about the nature of science. Granting that paradigm rejection has been a historic fact, does it illuminate more than human credulity and confusion? Are there intrinsic reasons why the assimilation of either a new sort of phenomenon or a new scientific theory must demand the rejection of an older paradigm?

First notice that if there are such reasons, they do not derive from the logical structure of scientific knowledge. In principle, a new phenomenon might emerge without reflecting destructively upon any part of past scientific practice. Though discovering life on the moon would today be destructive of existing paradigms (these tell us things about the moon that seem incompatible with life's existence there), discovering life in some less well-known part of the galaxy would not. By the same token, a new theory does not have to conflict with any of its predecessors. It might deal exclusively with phenomena not previously known, as the quantum theory deals (but, significantly, not exclusively) with subatomic phenomena unknown before the twentieth century. Or again, the new theory might be simply a higher level theory than those known before, one that linked together a whole group of lower level theories without substantially changing any. Today, the theory of energy conservation provides just such links between dynamics, chemistry, electricity, optics, thermal theory, and so on. Still other compatible relationships between old and new theories can be conceived. Any and all of them might be exemplified by the historical process through which science has developed. If they were, scientific development would be genuinely cumulative. New sorts of phenomena would simply disclose order in an aspect of nature where none had been seen before. In the evolution of science new knowledge would replace ignorance rather than replace knowledge of another and incompatible sort.

Of course, science (or some other enterprise, perhaps less effective) might have developed in that fully cumulative manner. Many people have believed that it did so, and most still seem to suppose that cumulation is at least the ideal that historical development would display if only it had not so often been distorted by human idiosyncrasy. There are important reasons for that belief. In Section X we shall discover how closely the view of science-as-cumulation is entangled with a dominant epistemology that takes knowledge to be a construction placed directly upon raw sense data by the mind. And in Section XI we shall examine the strong support provided to the same historiographic schema by the techniques of effective science pedagogy. Nevertheless, despite the immense plausibility of that ideal image, there is increasing reason to wonder whether it can possibly be an image of science. After the pre-paradigm period the assimilation of all new theories and of almost all new sorts of phenomena has in fact demanded the destruction of a prior paradigm and a consequent conflict between competing schools of scientific thought. Cumulative acquisition of unanticipated novelties proves to be an almost non-existent exception to the rule of scientific development. The man who takes historic fact seriously must suspect that science does not tend toward the ideal that our image of its cumulativeness has suggested. Perhaps it is another sort of enterprise.

If, however, resistant facts can carry us that far, then a second look at the ground we have already covered may suggest that cumulative acquisition of novelty is not only rare in fact but improbable in principle. Normal research, which is cumulative, owes its success to the ability of scientists regularly to select problems that can be solved with conceptual and instrumental techniques close to those already in existence. (That is why an excessive concern with useful problems, regardless of their relation to existing knowledge and technique, can so easily inhibit scientific development.) The man who is striving to solve a problem defined by existing knowledge and technique is not, however, just looking around. He knows what he wants to achieve, and he designs his instruments and directs his thoughts accordingly. Unanticipated novelty, the new discovery, can emerge only to the extent that his anticipations about nature and his instruments prove wrong. Often the importance of the resulting discovery will itself be proportional to the extent and stubbornness of the anomaly that foreshadowed it. Obviously, then, there must be a conflict between the paradigm that discloses anomaly and the one that later renders the anomaly law-like. The examples of discovery through paradigm destruction examined in Section VI did not confront us with mere historical accident. There is no other effective way in which discoveries might be generated.

The same argument applies even more clearly to the invention of new theories. There are, in principle, only three types of phenomena about which a new theory might be developed. The first consists of phenomena already well explained by existing paradigms, and these seldom provide either motive or point of departure for theory construction. When they do, as with the three famous anticipations discussed at the end of Section VII, the theories that result are seldom accepted, because nature provides no ground for discrimination. A second class of phenomena consists of those whose nature is indicated by existing paradigms but whose details can be understood only through further theory articulation. These are the phenomena to which scientists direct their research much of the time, but that research aims at the articulation of existing paradigms rather than at the invention of new ones. Only when these attempts at articulation fail do scientists encounter the third type of phenomena, the recognised anomalies whose characteristic feature is their stubborn refusal to be assimilated to existing paradigms. This type alone gives rise to new theories. Paradigms provide all phenomena except anomalies with a theory-determined place in the scientist’s field of vision.

But if new theories are called forth to resolve anomalies in the relation of an existing theory to nature, then the successful new theory must somewhere permit predictions that are different from those derived from its predecessor. That difference could not occur if the two were logically compatible. In the process of being assimilated, the second must displace the first. Even a theory like energy conservation, which today seems a logical superstructure that relates to nature only through independently established theories, did not develop historically without paradigm destruction. Instead, it emerged from a crisis in which an essential ingredient was the incompatibility between Newtonian dynamics and some recently formulated consequences of the caloric theory of heat. Only after the caloric theory had been rejected could energy conservation become part of science. And only after it had been part of science for some time could it come to seem a theory of a logically higher type, one not in conflict with its predecessors. It is hard to see how new theories could arise without these destructive changes in beliefs about nature. Though logical inclusiveness remains a permissible view of the relation between successive scientific theories, it is a historical implausibility.

Logical Positivism

A century ago it would, I think, have been possible to let the case for the necessity of revolutions rest at this point. But today, unfortunately, that cannot be done because the view of the subject developed above cannot be maintained if the most prevalent contemporary interpretation of the nature and function of scientific theory is accepted. That interpretation, closely associated with early logical positivism and not categorically rejected by its successors, would restrict the range and meaning of an accepted theory so that it could not possibly conflict with any later theory that made predictions about some of the same natural phenomena. The best-known and the strongest case for this restricted conception of a scientific theory emerges in discussions of the relation between contemporary Einsteinian dynamics and the older dynamical equations that descend from Newton's *Principia*. From the viewpoint of this essay these two theories are fundamentally incompatible in the sense illustrated by the relation of Copernican to Ptolemaic astronomy: Einstein's theory can be accepted only with the recognition that Newton's was wrong. Today this remains a minority view. We must therefore examine the most prevalent objections to it.

The gist of these objections can be developed as follows. Relativistic dynamics cannot have shown Newtonian dynamics to be wrong, for Newtonian dynamics is still used with great success by most engineers and, in selected applications, by many physicists. Furthermore, the propriety of this use of the older theory can be proved from the very theory that has, in other applications, replaced it. Einstein's theory can be used to show that predictions from Newton's equations will be as good as our measuring instruments in all applications that satisfy a small number of restrictive conditions. For example, if Newtonian theory is to provide a good approximate solution, the relative velocities of the bodies considered must be small compared with the velocity of light. Subject to this condition and a few others, Newtonian theory seems to be derivable from Einsteinian, of which it is therefore a special case.

But, the objection continues, no theory can possibly conflict with one of its special cases. If Einsteinian science seems to make Newtonian dynamics wrong, that is only because some Newtonians were so incautious as to claim that Newtonian theory yielded entirely precise results or that it was valid at very high relative velocities. Since they could not have had any evidence for such claims, they betrayed the standards of science when they made them. In so far as Newtonian theory was ever a truly scientific theory supported by valid evidence, it still is. Only extravagant claims for the theory – claims that were never properly parts of science can have been shown by Einstein to be wrong. Purged of these merely human extravagances, Newtonian theory has never been challenged and cannot be.

Some variant of this argument is quite sufficient to make any theory ever used by a significant group of competent scientists immune to attack. The much-maligned phlogiston theory, for example, gave order to a large number of physical and chemical phenomena. It explained why bodies burned – they were rich in phlogiston – and why metals had so many more properties in common than did their ores. The metals were all compounded from different elementary earths combined with phlogiston, and the latter, common to all metals, produced common properties. In addition, the phlogiston theory accounted for a number of reactions in which acids were formed by the combustion of substances like carbon and sulphur. Also, it explained the decrease of volume when combustion occurs in a confined volume of air the phlogiston released by combustion "spoils" the elasticity of the air that absorbed it, just as fire "spoils" the elasticity of a steel spring. If these were the only phenomena that the phlogiston theorists had claimed for their theory, that theory could never have been challenged. A similar argument will suffice for any theory that has ever been successfully applied to any range of phenomena at all.

But to save theories in this way, their range of application must be restricted to those phenomena and to that precision of observation with which the experimental evidence in hand already deals. Carried just a step further (and the step can scarcely be avoided once the first is taken), such a limitation prohibits the scientist from claiming to speak "scientifically" about any phenomenon not already observed. Even in its present form the restriction forbids the scientist to rely upon a theory in his own research whenever that research enters an area or seeks a degree of precision for which past practice with the theory offers no precedent. These prohibitions are logically unexceptionable. But the result of accepting them would be the end of the research through which science may develop further.

By now that point too is virtually a tautology. Without commitment to a paradigm there could be no normal science. Furthermore, that commitment must extend to areas and to degrees of precision for which there is no full precedent. If it did not, the paradigm could provide no puzzles that had not already been solved. Besides, it is not only normal science that depends upon commitment to a paradigm. If existing theory binds the scientist only with respect to existing applications, then there can be no surprises, anomalies, or crises. But these are just the signposts that point the way to extraordinary science. If positivistic restrictions on the range of a theory's legitimate applicability are taken literally, the mechanism that tells the scientific community what problems may lead to fundamental change must cease to function. And when that occurs, the community will inevitably return to something much like its pre-paradigm state a condition in which all members practice science but in which their gross product scarcely resembles science at all. Is it really any wonder that the price of significant scientific advance is a commitment that runs the risk of being wrong?

More important, there is a revealing logical lacuna in the positivist's argument, one that will reintroduce us immediately to the nature of revolutionary change. Can Newtonian dynamics really be derived from relativistic dynamics? What would such a derivation look like? Imagine a set of statements, E_1, E_2, ... E_n which together embody the laws of relativity theory. These statements contain variables and parameters representing spatial position, time, rest mass, etc. From them, together with the apparatus of logic and mathematics, is deducible a whole set of further statements including some that can be checked by observation. To prove the adequacy of Newtonian dynamics as a special case, we must add to the E_i's additional statements, like $(v/c)^2 << 1$, restricting the range of the parameters and variables. This enlarged set of statements is then manipulated to yield a new set, N_1, N_2, ..., N_m, which is identical in form with Newton's laws of motion, the law of gravity, and so on. Apparently Newtonian dynamics has been derived from Einsteinian, subject to a few limiting conditions.

Yet the derivation is spurious, at least to this point. Though the N_i's are a special case of the laws of relativistic mechanics, they are not Newton's Laws. Or at least they are not unless those laws are reinterpreted in a way that would have been impossible until after Einstein's work. The variables and parameters that in the Einsteinian E_i's represented spatial position, time, mass, etc., still occur in the N_i's; and they there still represent Einsteinian space, time, and mass. But the physical referents of these Einsteinian concepts are by no means identical with those of the Newtonian concepts that bear the same name. (Newtonian mass is conserved; Einsteinian is convertible with energy. Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same.) Unless we change the definitions of the variables in the N_i's, the statements we have derived are not Newtonian. If we do change them, we cannot properly be said to have derived Newton's Laws, at least not in any sense of "derive" now generally recognised. Our argument has, of course, explained why Newton's Laws ever seemed to work. In doing so it has justified, say, an automobile driver in acting as though he lived in a Newtonian universe. An argument of the same type is used to justify teaching earth-centred astronomy to surveyors. But the argument has still not done what it purported to do. It has not, that is, shown Newton's Laws to be a limiting case of Einstein's. For in the passage to the limit it is not only the forms of the laws that have changed. Simultaneously we have had to alter the fundamental structural elements of which the universe to which they apply is composed.

This need to change the meaning of established and familiar concepts is central to the revolutionary impact of Einstein's theory. Though subtler than the changes from geocentrism to heliocentrism, from phlogiston to oxygen, or from corpuscles to waves, the resulting conceptual transformation is no less decisively destructive of a previously established paradigm. We may even come to see it as a prototype for revolutionary reorientations in the sciences. Just because it did not involve the introduction of additional objects or concepts, the transition from Newtonian to Einsteinian mechanics illustrates with particular clarity the scientific revolution as a displacement of the conceptual network through which scientists view the world.

These remarks should suffice to show what might, in another philosophical climate, have been taken for granted. At least for scientists, most of the apparent differences between a discarded scientific theory and its successor are real. Though an out-of-date theory can always be viewed as a special case of its up-to-date successor, it must be transformed for the purpose. And the transformation is one that can be undertaken only with the advantages of hindsight, the explicit guidance of the more recent theory. Furthermore, even if that transformation were a legitimate device to employ in interpreting the older theory, the result of its application would be a theory so restricted that it could only restate what was already known. Because of its economy, that restatement would have utility, but it could not suffice for the guidance of research.

Let us, therefore, now take it for granted that the differences between successive paradigms are both necessary and irreconcilable. Can we then say more explicitly what sorts of differences these are? The most apparent type has already been illustrated repeatedly. Successive paradigms tell us different things about the population of the universe and about that population's behaviour. They differ, that is, about such questions as the existence of subatomic particles, the materiality of light, and the conservation of heat or of energy. These are the substantive differences between successive paradigms, and they require no further illustration. But paradigms differ in more than substance, for they are directed not only to nature but also back upon the science that produced them. They are the source of the methods, problem-field, and standards of solution accepted by any mature scientific community at any given time. As a result, the reception of a new paradigm often necessitates a redefinition of the corresponding science. Some old problems may be relegated to another science or declared entirely “unscientific.” Others that were previously non-existent or trivial may, with a new paradigm, become the very archetypes of significant scientific achievement. And as the problems change, so, often, does the standard that distinguishes a real scientific solution from a mere metaphysical speculation, word game, or mathematical play. The normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually incommensurable with that which has gone before.

The impact of Newton’s work upon the normal seventeenth century tradition of scientific practice provides a striking example of these subtler effects of paradigm shift. Before Newton was born the “new science” of the century had at last succeeded in rejecting Aristotelian and scholastic explanations expressed in terms of the essences of material bodies. To say that a stone fell because its “nature” drove it toward the center of the universe had been made to look a mere tautological word-play, something it had not previously been. Henceforth the entire flux of sensory appearances, including colour, taste, and even weight, was to be explained in terms of the size, shape, position, and motion of the elementary corpuscles of base matter. The attribution of other qualities to the elementary atoms was a resort to the occult and therefore out of bounds for science. Molière caught the new spirit precisely when he ridiculed the doctor who explained opium’s efficacy as a soporific by attributing to it a dormitive potency. During the last half of the seventeenth century many scientists preferred to say that the round shape of the opium particles enabled them to sooth the nerves about which they moved.

In an earlier period explanations in terms of occult qualities had been an integral part of productive scientific work. Nevertheless, the seventeenth century’s new commitment to mechanico-corpuscular explanation proved immensely fruitful for a number of sciences, ridding them of problems that had defied generally accepted solution and suggesting others to replace them. In dynamics, for example, Newton’s three laws of motion are less a product of novel experiments than of the attempt to reinterpret well-known observations in terms of the motions and interactions of primary neutral corpuscles. Consider just one concrete illustration. Since neutral corpuscles could act on each other only by contact, the mechanico-corpuscular view of nature directed scientific attention to a brand-new subject of study, the alteration of particulate motions by collisions. Descartes announced the problem and provided its first putative solution. Huygens, Wren, and Wallis carried it still further, partly by experimenting with colliding pendulum bobs, but mostly by applying previously well-known characteristics of motion to the new problem. And Newton embedded their results in his laws of motion. The equal "action" and "reaction" of the third law are the changes in quantity of motion experienced by the two parties to a collision. The same change of motion supplies the definition of dynamical force implicit in the second law. In this case, as in many others during the seventeenth century, the corpuscular paradigm bred both a new problem and a large part of that problem's solution.

Yet, though much of Newton's work was directed to problems and embodied standards derived from the mechanico-corpuscular world view, the effect of the paradigm that resulted from his work was a further and partially destructive change in the problems and standards legitimate for science. Gravity, interpreted as an innate attraction between every pair of particles of matter, was an occult quality in the same sense as the scholastics' "tendency to fall" had been. Therefore, while the standards of corpuscularism remained in effect, the search for a mechanical explanation of gravity was one of the most challenging problems for those who accepted the *Principia* as paradigm. Newton devoted much attention to it and so did many of his eighteenth-century successors. The only apparent option was to reject Newton's theory for its failure to explain gravity, and that alternative, too, was widely adopted. Yet neither of these views ultimately triumphed. Unable either to practice science without the *Principia* or to make that work conform to the corpuscular standards of the seventeenth century, scientists gradually accepted the view that gravity was indeed innate. By the mid-eighteenth century that interpretation had been almost universally accepted, and the result was a genuine reversion (which is not the same as a retrogression) to a scholastic standard. Innate attractions and repulsions joined size, shape, position, and motion as physically irreducible primary properties of matter.

The resulting change in the standards and problem-field of physical science was once again consequential. By the 1740's, for example, electricians could speak of the attractive "virtue" of the electric fluid without thereby inviting the ridicule that had greeted Molière's doctor a century before. As they did so, electrical phenomena increasingly displayed an order different from the one they had shown when viewed as the effects of a mechanical effluvium that could act only by contact. In particular, when electrical action-at-a-distance became a subject for study in its own right, the phenomenon we now call charging by induction could be recognised as one of its effects. Previously, when seen at all, it had been attributed to the direct action of electrical "atmospheres" or to the leakages inevitable in any electrical laboratory. The new view of inductive effects was, in turn, the key to Franklin's analysis of the Leyden jar and thus to the emergence of a new and Newtonian paradigm for electricity. Nor were dynamics and electricity the only scientific fields affected by the legitimisation of the search for forces innate to matter. The large body of eighteenth-century literature on chemical affinities and replacement series also derives from this supramechanical aspect of Newtonianism. Chemists who believed in these differential attractions between the various chemical species set up previously unimagined experiments and searched for new sorts of reactions. Without the data and the chemical concepts developed in that process, the later work of Lavoisier and, more particularly, of Dalton would be incomprehensible. Changes in the standards governing permissible problems, concepts, and explanations can transform a science. In the next section I shall even suggest a sense in which they transform the world.

Other examples of these non-substantive differences between successive paradigms can be retrieved from the history of any science in almost any period of its development. For the moment let us be content with just two other and far briefer illustrations. Before the chemical revolution, one of the acknowledged tasks of chemistry was to account for the qualities of chemical substances and for the changes these qualities underwent during chemical reactions. With the aid of a small number of elementary "principles" – of which phlogiston was one – the chemist was to explain why some substances are acidic, others metalline, combustible, and so forth. Some success in this direction had been achieved. We have already noted that phlogiston explained why the metals were so much alike, and we could have developed a similar argument for the acids. Lavoisier's reform, however, ultimately did away with chemical "principles," and thus ended by depriving chemistry of some actual and much potential explanatory power. To compensate for this loss, a change in standards was required. During much of the nineteenth century failure to explain the qualities of compounds was no indictment of a chemical theory.

Or again, Clerk Maxwell shared with other nineteenth-century proponents of the wave theory of light the conviction that light waves must be propagated through a material ether. Designing a mechanical medium to support such waves was a standard problem for many of his ablest contemporaries. His own theory, however, the electromagnetic theory of light, gave no account at all of a medium able to support light waves, and it clearly made such an account harder to provide than it had seemed before. Initially, Maxwell's theory was widely rejected for those reasons. But, like Newton's theory, Maxwell's proved difficult to dispense with, and as it achieved the status of a paradigm the community's attitude toward it changed. In the early decades of the twentieth century Maxwell's insistence upon the existence of a mechanical ether looked more and more like lip service, which it emphatically had not been, and the attempts to design such an ethereal medium were abandoned. Scientists no longer thought it unscientific to speak of an electrical "displacement" without specifying what was being displaced. The result, again, was a new set of problems and standards, one which, in the event, had much to do with the emergence of relativity theory.

These characteristic shifts in the scientific community's conception of its legitimate problems and standards would have less significance to this essay's thesis if one could suppose that they always occurred from some methodologically lower to some higher type. In that case their effects, too, would seem cumulative. No wonder that some historians have argued that the history of science records a continuing increase in the maturity and refinement of man's conception of the nature of science. Yet the case for cumulative development of science's problems and standards is even harder to make than the case for cumulation of theories. The attempt to explain gravity, though fruitfully abandoned by most eighteenth-century scientists, was not directed to an intrinsically illegitimate problem; the objections to innate forces were neither inherently unscientific nor metaphysical in some pejorative sense. There are no external standards to permit a judgment of that sort. What occurred was neither a decline nor a raising of standards, but simply a change demanded by the adoption of a new paradigm. Furthermore, that change has since been reversed and could be again. In the twentieth century Einstein succeeded in explaining gravitational attractions, and that explanation has returned science to a set of canons and problems that are, in this particular respect, more like those of Newton's predecessors than of his successors. Or again, the development of quantum mechanics has reversed the methodological prohibition that originated in the chemical revolution. Chemists now attempt, and with great success, to explain the colour, state of aggregation, and other qualities of the substances used and produced in their laboratories. A similar reversal may even be underway in electromagnetic theory. Space, in contemporary physics, is not the inert and homogenous substratum employed in both Newton's and Maxwell's theories; some of its new properties are not unlike those once attributed to the ether; we may some day come to know what an electric displacement is.

By shifting emphasis from the cognitive to the normative functions of paradigms, the preceding examples enlarge our understanding of the ways in which paradigms give form to the scientific life. Previously, we had principally examined the paradigm's role as a vehicle for scientific theory. In that role it functions by telling the scientist about the entities that nature does and does not contain and about the ways in which those entities behave. That information provides a map whose details are elucidated by mature scientific research. And since nature is too complex and varied to be explored at random, that map is as essential as observation and experiment to science's continuing development. Through the theories they embody, paradigms prove to be constitutive of the research activity. They are also, however, constitutive of science in other respects, and that is now the point. In particular, our most recent examples show that paradigms provide scientists not only with a map but also with some of the directions essential for map-making. In learning a paradigm the scientist acquires theory, methods, and standards together, usually in an inextricable mixture. Therefore, when paradigms change, there are usually significant shifts in the criteria determining the legitimacy both of problems and of proposed solutions.

That observation returns us to the point from which this section began, for it provides our first explicit indication of why the choice between competing paradigms regularly raises questions that cannot be resolved by the criteria of normal science. To the extent, as significant as it is incomplete, that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk through each other when debating the relative merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent. There are other reasons, too, for the incompleteness of logical contact that consistently characterises paradigm debates. For example, since no paradigm ever solves all the problems it defines and since no two paradigms leave all the same problems unsolved, paradigm debates always involve the question: Which problems is it more significant to have solved? Like the issue of competing standards, that question of values can be answered only in terms of criteria that lie outside of normal science altogether, and it is that recourse to external criteria that most obviously makes paradigm debates revolutionary. Something even more fundamental than standards and values is, however, also at stake. I have so far argued only that paradigms are constitutive of science. Now I wish to display a sense in which they are constitutive of nature as well.

Postscript: Revolutions and Relativism

One consequence of the position just outlined has particularly bothered a number of my critics. They find my viewpoint relativistic, particularly as it is developed in the last section of this book. My remarks about translation highlight the reasons for the charge. The proponents of different theories are like the members of different language-culture communities. Recognising the parallelism suggests that in some sense both groups may be right. Applied to culture and its development that position is relativistic.

But applied to science it may not be, and it is in any case far from mere relativism in a respect that its critics have failed to see. Taken as a group or in groups, practitioners of the developed sciences are, I have argued, fundamentally puzzle-solvers. Though the values that they deploy at times of theory-choice derive from other aspects of their work as well, the demonstrated ability to set up and to solve puzzles presented by nature is, in case of value conflict, the dominant criterion for most members of a scientific group. Like any other value, puzzle-solving ability proves equivocal in application. Two men who share it may nevertheless differ in the judgments they draw from its use. But the behaviour of a community which makes it pre-eminent will be very different from that of one which does not. In the sciences, I believe, the high value accorded to puzzle-solving ability has the following consequences.

Imagine an evolutionary tree representing the development of the modern scientific specialties from their common origins in, say, primitive natural philosophy and the crafts. A line drawn up that tree, never doubling back, from the trunk to the tip of some branch would trace a succession of theories related by descent. Considering any two such theories, chosen from points not too near their origin, it should be easy to design a list of criteria that would enable an uncommitted observer to distinguish the earlier from the more recent theory time after time. Among the most useful would be: accuracy of prediction, particularly of quantitative prediction; the balance between esoteric and everyday subject matter; and the number of different problems solved. Less useful for this purpose, though also important determinants of scientific life, would be such values as simplicity, scope, and compatibility with other specialties. Those lists are not yet the ones required, but I have no doubt that they can be completed. If they can, then scientific development is, like biological, a unidirectional and irreversible process. Later scientific theories are better than earlier ones for solving puzzles in the often quite different environments to which they are applied. That is not a relativist's position, and it displays the sense in which I am a convinced believer in scientific progress.

Compared with the notion of progress most prevalent among both philosophers of science and laymen, however, this position lacks an essential element. A scientific theory is usually felt to be better than its predecessors not only in the sense that it is a better instrument for discovering and solving puzzles but also because it is somehow a better representation of what nature is really like. One often hears that successive theories grow ever closer to, or approximate more and more closely to, the truth. Apparently generalisations like that refer not to the puzzle-solutions and the concrete predictions derived from a theory but rather to its ontology, to the match, that is, between the entities with which the theory populates nature and what is "really there."

Perhaps there is some other way of salvaging the notion of 'truth' for application to whole theories, but this one will not do. There is, I think, no theory-independent way to reconstruct phrases like 'really there'; the notion of a match between the ontology of a theory and its "real" counterpart in nature now seems to me illusive in principle. Besides, as a historian, I am impressed with the implausibility of the view. I do not doubt, for example, that Newton's mechanics improves on Aristotle's and that Einstein's improves on Newton's as instruments for puzzle-solving. But I can see in their succession no coherent direction of ontological development. On the contrary, in some important respects, though by no means in all, Einstein's general theory of relativity is closer to Aristotle's than either of them is to Newton's. Though the temptation to describe that position as relativistic is understandable, the description seems to me wrong. Conversely, if the position be relativism, I cannot see that the relativist loses anything needed to account for the nature and development of the sciences. ...

Further Reading:
Biography | Hegel | Lektorsky | Noam Chomsky | Einstein | Newton | Talcott Parsons

Philosophy Archive @ marxists.org