On the Problem of Truth and Understanding in Science

David Bohm

If one were to ask for a statement of the essence of Professor Popper's contribution to the theory of scientific research, it could probably best be expressed in the sentence: "Science is a part of the search for truth." To one who has not gone very deeply into the question this may seem self evident. However, on thinking more carefully one finds that the problem of truth is a slippery one, full of pitfalls. It is therefore perhaps not so surprising that scientists and philosophers of science have been attracted to a wide range of suggestions with the aim of avoiding this problem. To achieve this, they have tried to refer the essential content of science to something less vague and difficult to grasp, less "metaphysical," more solid, definable, and "positive" than the elusive notion of truth.

As a typical example of such efforts one can take the rather common idea that scientific theories are in essence just convenient and useful ways of summarizing, ordering, and organizing what we know of nature through past experiences. This point of view, in which the notion of truth is in effect replaced by that of utility and convenience, has most systematically been embodied in the positivist philosophies that were current during the nineteenth century (notably those of Mach and Comte). Later exponents of the positivist philosophy were, however, no longer satisfied with these ideas, but still felt that the notion of truth had to be replaced by something else that could be stated definitely and explicitly. Thus, the Vienna School proposed that the essential content of scientific theories was in their verifiability, i.e., in the procedures by which it could be seen that they are true. A closely related point of view had meanwhile arisen in America, pragmatism, in which the notion of truth was replaced by that of "what works." Scientific theory was thus conceived as nothing more than an instrument in man's efforts to control nature. Later there arose the allied view of operationalism, which identifies the content of a theory with statements about operations that a scientist can carry out with the aid of suitable equipment.

The tendency to try to replace the notion of truth by something else that can be given an explicit definition has continued into many more modern forms, which, however, need not concern us here. This tendency has penetrated very deeply into all thinking about science, both by the layman and by the scientist himself. For example, even the very form of theories in physics frequently reflects the commonly accepted presupposition that the essential purpose of physical laws is to enable us to predict the results of experiments. Once again, the question of truth is avoided, and it is replaced by something else on which one can put one's finger more easily, i.e., prediction.

Professor Popper has consistently criticized the point of view described above. Indeed, it was just in his efforts to show the inadequacy of the suggestions of

the Vienna School to substitute verifiability for truth that he was led to formulate his by now very well known thesis that no theory can ever be verified conclusively by any number of observations, however large, and that in fact, the question of falsifiability is in many ways more relevant than that of verifiability. Professor Popper also pointed out, of course, that it is just as impossible to falsify a theory conclusively as it is to verify it. But his statements concerning the impossibility of conclusively falsifying a theory are frequently overlooked, with the result that his views are distorted into what is, in effect, a continuation of the Vienna School in a modified form. Thus, it is now often asserted of his position that he maintains that the essential content of a scientific theory is not in its verifiability but, rather, in its falsifiability.

In fact, however, Professor Popper consistently stresses that the essential question to be studied is that of truth. While admitting that this is a very difficult question, he has shown throughout his whole life's work that if we do not tackle it, we will not even be able to make a start on the question of what is the essential point of scientific research.

It is with the above general situation in mind that I should like to go a bit more deeply into the problem of truth and its place in science. This problem is so closely allied with that of understanding, however, that it seems to be necessary to consider them as inseparable. Since there has already been a great deal of discussion of truth and comparatively little concerning understanding, it will be helpful to begin with understanding and then to return to the problem of truth later. Let us start by asking ourselves a simple question: What is understanding? In response to this question, there immediately arises another question: Can understanding ever really be defined? For example, suppose that one were to give an exhaustive definition of understanding. At the end, one could ask: "Have you understood?" In this way one shows that no definition of understanding is actually possible, which does not presuppose that the hearer already understands what is meant by understanding (even though he cannot put it into words).

Indeed, a person who did not understand what is meant by understanding could not enter into a conversation, and he would hardly be called human in any respect whatsoever. We see then that understanding is implicit, in the sense that it is logically prior to all words and thoughts. The recognition that the most fundamental things are implicit and incapable of being directly and positively asserted is, however, not entirely new. Thus, in geometry, it is well known that an explicit definition of points and lines and other such basic concepts is impossible. Rather, one begins by simply stating their existence, while their properties are only implicit in the axioms that are assumed to relate them. The choice of axioms is guided to some extent by vaguely and intuitively grasped common experiences with space, but finally, the properties of the points, lines, and other entities must be explicated (or made explicit) by a working out of the consequences of the axioms. And when this is done, the results are in general not a mere recovery in a more precise form of what was

already known in common experience, but rather, they are often in many ways new and even surprising.

We may say then that in fundamental questions, it is always necessary to approach the object of our studies obliquely, by implication rather than by positive assertions and definite conclusions, so that the positivist goal of trying to state precisely what is happening is never appropriate in a really deep problem. And no problem can be deeper than that of what is meant by understanding. For every form of human activity, including science, mathematics, and philosophy, presupposes understanding. We are therefore led to study this problem in an oblique and indirect fashion.

Guided by our over-all experience, let us try to find some general characteristics which throw light on a few aspects of understanding, at least by implication.

It is helpful to approach the problem by considering the example of the circle. Thus, at first people studied circles empirically, amassing many items of information about them (e.g., that the circumference is approximately 22/7 times the diameter, that certain chords are related to the diameter in certain ways, etc.). This may be called the stage of associative thinking; for the circle is treated simply by associating together a large number of remembered properties that were discovered empirically. Of course, such a mode of thinking is good enough for a wide range of practical purposes. Indeed, it is by a similar type of thinking that we get through a great deal of everyday life. At a certain stage, however, a new geometrical theory was developed, and a circle was regarded as a curve traced by a point moving equidistant from a fixed point, while a straight line was regarded as the result of moving a small straight segment in its own direction, etc.

Later, when this geometrical theory is explained to a particular individual, at first he tries to bring together the various parts of the argument. Then suddenly he says "I see," meaning by this that he understands. But what is it that is seen? In what is generally described as a very sudden process, a "click" or a "flash," one grasps the basic principle of the circle, which is to say, one sees it as a totality. Of course, this does not mean that one immediately knows the whole set of properties of a circle exhaustively and in full detail. Rather, it means that one sees the essential process by which the circle is generated, along with its various parts and aspects, which are now treated as sides of this totality, so formed that they are automatically in their proper relationships. In this way, one not only grasps the essential character of each circle, but also of all circles.

With the coming of understanding, then, there is (at least in the field under discussion) what amounts to a revolutionary change in the mode of thinking, in which the earlier associative thought about this field is set aside, and in which the field is comprehended as a totality. However, the fact that understanding involves such a fundamental change in thinking is perhaps generally not fully

realized consciously, or at least, its importance is frequently underestimated. We can bring out this importance a bit more sharply by considering a hypothetical device, which we shall call an "understanding machine."

Of course, it is common knowledge that calculating machines can now be made which can be programmed to do many tasks previously requiring human thought and attention. If one considers the possibility of an indefinite increase in the number of elements in such a machine (along with new and improved principles of construction and operation), one is quite naturally led to ask whether it would not be possible eventually to reproduce all of the properties of human thinking. It seems quite plausible that machines that could in effect "learn" by association could eventually be made to reproduce or perhaps even to excel human abilities in this field. But could a machine also come to "understand" in the sense that it could obtain a new "vision of totality"? For example, could it "learn" the principles of a science such as physics along with the basic experimental facts that are known at a given time, then "criticize" these principles, discover their weak points, and in a sudden "click" or "flash" (or in a series of them) "see" how newer and better theories could be developed? If it could do this, then it could "grasp" or "comprehend" its own system of programming and "suggest" new systems and perhaps even new principles for its own construction and operation.

Let us now suppose that we wish to develop such an "understanding machine." Our first task would be to define what is "understanding"; for if we do not know what it is that we are trying to reproduce, we can hardly design a machine that will function as planned. But as we have seen, every definition of understanding is necessarily only partial, and at best implicit. For it leaves out of the account the very action of understanding that is needed on the part of the proposer and the hearer before this definition has any meaning. Moreover, it also fails to include a potential infinity of new ways of understanding that can develop in the future. So, at least in our present state of understanding, it would seem to be impossible to design an understanding machine, whereas we can in all probability eventually design a machine that reproduces the essential characteristics of associative thinking quite well.

It seems clear, then, that understanding is something radically and fundamentally different from associative thinking. However, it is only in the context of the fact that understanding actually takes place that the question of truth can arise. For if there is no understanding of what is being thought and said, what meaning can there be to asking whether it is true or false? But when we understand deeply enough, then, as we can see by actually experiencing it as a fact, there is the possibility of perceiving the truth or falsity of what is being thought and said. Truth, like understanding, is basically implicit. If one were to give a definition of truth, at the end one could say "And this is the true definition of truth," thus showing that the hearer had to understand the meaning of truth before communication was even conceivable. This means that all efforts to reduce science to some definite, explicit, "positive" characteristics are bound to fail, because none of these makes sense

except in a context in which the undefinables, “truth” and “understanding,” are present from the very beginning.

The need for the falsifiability of theories, as emphasized by Professor Popper, already shows that the question of trying to determine the meaning of truth cannot properly be treated by the direct approach. Thus, if every really acceptable theory must be falsifiable, then it seems almost certain that, in time, such a theory will actually be falsified as more accurate experiments are done in broader domains and in new contexts. (As has, in fact, already happened to an extremely large number of theories in every branch of science.) In other words, it is hardly to be expected that any theory with a real predictive content will not in some way turn out to be false. But if all our theories are thus very likely to be false, then where is truth, and what is it that we are really searching for?

Here, once again, we must recall that there can be no explicit definition of truth, and that light on questions of this kind can best be obtained by proceeding in an oblique and indirect fashion. The first thing to do is to recognize that while almost anything that we say in our theories may eventually be shown to be in some way false, these theories will nevertheless continue to express a great many true relationships. For example, in the early phases of the development of physics, gases were often treated as perfectly continuous distributions of matter. Yet, even though the falsity of that conception is implied by the atomic theory developed since, the relationships of temperature and pressure obtained in terms of the former treatment are true (at least within a certain degree of approximation). True relationships can never be obtained from self-contradictory theories. For example, though we write the self-contradictory statements, $x=2x$, $y=2y$, this contradiction will not matter in the quantities which depend only on the ratio x/y.

It is clear that the appearance of some experiments confirming a theory still leaves plenty of room for false implications in the theory in question, as well as for self-contradictions which are subtle enough to have escaped notice thus far. The falsity of such implications and the existence of such contradictions may be shown up by future experiments. As a rule, however, experiments are by themselves sufficient for doing this only in very specific and narrowly delimited problems. Thus, it is well known that, with regard to theories having a broad domain of applicability, it is almost impossible to find a set of experiments that clearly and unambiguously show the theory to be false, because it is always possible that a modification of some inessential feature could bring the theory back into agreement with experiment. For this reason it frequently happens that a set of experiments which can be used for falsifying the theory may be available for a long time, while their true significance is not realized. In such cases, the crucial step in falsifying the older theory comes with an understanding of the implications of these experiments.

The development of the theory of relativity furnishes a very good example of the above point. Thus, the Michelson–Morley experiment and many others had

been available for some time, but there had been a widespread and not unnatural feeling that some change in the theory of the ether or some modification in our ideas concerning the mode of propagation of light could eventually account for these results. Einstein's basic contribution was his new way of understanding the problem. For he saw that the space-time coordinates are not something self-existent and wholly independent of the laws of physics, but rather that these coordinates must express relationships between physical entities (e.g., as manifested in the readings of clocks and measurements made with rulers). As a result, he was able to see that the essential point at issue in all these experiments was that the space-time frame of reference and the laws of physics must be interconnected very fundamentally. Therefore, it was through understanding the whole situation in a new way that it was possible for Einstein to perceive the falsity, not of some specific and detailed aspect of the old theory, but rather of the Newtonian conception of absolute space and time in its totality.

In a similar way, most of the experimental facts underlying the quantum theory had been available for a number of years when Niels Bohr saw a new way of understanding these facts on the basis of the notion of discrete energy levels in atoms and discontinuous processes of transition between these levels. Once again, it is clear that Bohr's new vision of the laws of physics falsified not a particular aspect of classical theory but the conception of continuous motion as a whole. Other examples of less comprehensive developments in our theories will occur to the reader. However, he will readily see that in most cases it is not the experiment itself that falsifies earlier theories and conceptions; rather, it is some new understanding which arises in response to reflection on the total situation, experimental and theoretical.

The question of seeing just what is false is particularly important to clarify. For when one realizes that the older point of view is false, one can also see in what way it is false, and in what way it continues to have truth in it. Thus, in the case of the theory of relativity, it was evident that while the Newtonian conception of absolute space and time is false in a basic sense, nevertheless at velocities small compared with that of light, it leads, at least within a certain level of approximation, to true relationships in a great many physical processes. (Similarly with regard to the connection between classical and quantum mechanics.) It may therefore be said that in addition to perceiving what is true and what is false, we can see the truth in the false (true relationships) and the truth about the false (the reasons why it is not true). When we go into the question of what is the real significance of falsification, it then becomes clear that in a fundamental problem much more is involved than merely to show that a given theory does not agree with experiment. For what actually happens is that the perception of the truth in the new conceptions and of the falsity of the old conceptions takes place in the same step, at the very moment of understanding. Thus, truth and understanding are really inseparable. Perhaps it could best be put by saying that they are themselves two sides of a totality.

In view of this essential connection between truth and understanding, what can we say about the question of an absolute truth? Is there an absolute truth which is, as it were, something accomplished, definite, and finished? If there is, then it would not be unreasonable to suppose that our theories can approach this truth. Such a light might even be infinite, so that man could never actually reach it or even get nearer to it, in any absolute sense. Nevertheless, at least he could still always be moving in its direction, or constantly accumulating more and more of it. If there is not such an absolute truth, the idea arises quite naturally that perhaps truth is only relative—for example, to the state of mind of a particular person, to the tastes, habits of thought, and general culture of a particular group during a particular epoch, or to the particular procedures that the main body of scientists may have chosen to adopt in their work in order to get results of a kind that they desire. It seems clear that all the various philosophical points of view such as positivism, pragmatism, operationalism, etc., do in fact try in this way to refer truth to something else and thus end up in effect by assuming that it is only relative. But a purely relative truth is in reality no truth at all, since it finally depends neither on itself alone nor on some more fundamental truth, but rather on something completely extraneous to the question of truth that happens to have been regarded as important by a particular person or by a particular group of people.

It seems that we are faced with a very difficult problem. The notion that truth is only relative simply makes no sense, for then there would be no reason to choose one criterion of what is important over another. Indeed, even the most thorough-going positivists, pragmatists, and others of similar schools are trying to maintain that their notion of the relativity of truth is the true one, and in this way they bring in absolute truth by the back door. (If they did not wish to suggest that at least their ideas on the subject were really true, there would evidently be no point in their saying or doing anything whatsoever concerning this problem.) On the otherhand, any attempt to avoid the arbitrariness of the notion that truth is only relative by asserting that there is an absolute truth is at bottom self- contradictory. For since there is no way to prove that this assertion is itself really true, it finally reduces to a belief or a hope, the foundation of which is only in the mental or emotional state of a particular individual or group of individuals. Therefore, the person who criticizes those who hold to the purely relative character of truth for basing everything on subjective ideas, preferences, and beliefs is doing just the same thing if he asserts the absolute character of truth.

We see that there is something paradoxical in the effort to regard truth within the framework of the system of categories, absolute versus relative. Perhaps, then, the difficulty can be resolved only if we begin by seeing the falsity of this whole way of considering the problem. For as we saw earlier, truth (like understanding) does not seem to be approachable through positive assertions with regard to its character. Let us therefore try once again to approach it more indirectly. The problem described above seems to arise, at least in part, in the effort to refer truth to something fixed, definite, and final, either a

subjective criterion which we are to choose once and for all, to be applied in research, or else an objective truth that is supposed to exist somewhere "out there" in a finished form, and which we are supposed to approach step by step, or to accumulate bit by bit. But it may be that truth is none of these. Perhaps it is something that has no fixed and final forms or limits within it, so that it cannot be known in its totality nor approached nor accumulated nor even referred to some definable criterion by which it can be recognized. Instead, what may happen is that both truth itself and the methods and criteria for establishing it must be understood afresh from moment to moment, because everything is always changing, so that the problem is, in some respects, fundamentally new on each occasion on which the question of truth is to be considered.

Of course, a great many aspects of the total situation that confronts us from moment to moment (and in which we are) do actually repeat their essential characteristics, at least on a large number of occasions and for a wide variety of conditions and contexts. To the extent that this happens knowledge established on the basis of past understanding will continue to be valid. However, the repetition of past situations is never complete. Indeed, any particular characteristic, no matter how firmly and securely established it seems to be, can change fundamentally, often in a totally unexpected way. Such changes are in fact always occurring naturally, as things move and develop of their own accord, as well as because of human activities. In this regard, it must be remembered that in the particular activity of scientific research, the indefinite repetition of the same experiment under identical (or essentially identical) conditions is of very little interest, even in those cases in which such a repetition is actually possible. Indeed, since theories are expressed in the form of universal laws, scientific research is always directed towards testing these laws by the study of their operation under new conditions, in new contexts, and within new degrees of approximation.

Moreover, these studies are carried out with the aid of experiments involving new techniques, new kinds of apparatus, and new modes of investigation. As a result, not only because of the general nature of the world and of human activities in this world, but also because of the peculiar nature of scientific research, the conditions in the total situation confronting us from moment to moment are certain not to undergo indefinite and effectively identical repetition. It is basically because of the externally changing character of the world and of our experiences with it that the question of falsification of existing theories is so important. For it then follows that the validity of past knowledge can in general be only partial (in the sense that it contains some true relationships). When a theory is falsified, the fact that its validity is only partial is what is actually established. But now, we can see more thoroughly into the full implications of this feature of scientific research. For because nothing remains completely identical (even in its essential characteristics) with what it was previously, an understanding that is adequate to the situation confronting us at a given moment will, in certain ways, cease to be adequate for the next. Truth will therefore slip out of our grasp, unless we are continually alert and

attentive to the ever-changing total situation, being ready always to perceive the falsity of our older ideas in newer conditions and contexts and continually to develop new ideas that are appropriate to new situations. Thus, we are led to consider truth, not as fixed and finished, but, rather, as coming into being anew from moment to moment. It is clear that truth as conceived above is not subjective, in the sense that it would have meaning only in relation to the arbitrary ideas, preferences, goals, and feelings of various individuals or groups of individuals. Nevertheless, to the actual situation prevailing in a given moment there will be special standpoints and perspectives, as well as a general "world view," which are peculiarly appropriate for the understanding of what is in that moment. If one holds too strongly onto these, then one may misunderstand something that arises in the next moment, and thus lose part of the truth. In this regard, a new situation orproblem may even require basically new general methods of investigation and general criteria for acceptable theories. For example, consider what is usually called "the scientific method." Is it possible, once and for all, to define exhaustively what this is? Evidently not, because this method has itself evolved, and is still evolving, in response to our being confronted with ever new kinds of problems. The scientific method of today contains aspects that were not present several centuries ago, and it seems very likely that in a few more centuries it will be very different again in many respects from what it is now. But this means that it must be continually changing. Perhaps the change is imperceptible in the short run but, nevertheless, it is clearly a real one.

The notion that we can completely separate the mode of understanding from the object of understanding is then evidently false. If our understanding is to follow its object and thus to remain true, it is necessary that even our methods of research, criteria for truth, and general perspectives and "world view" shall be free to change from one occasion to another. For it must be seen that in each concrete situation and problem, the proper (i.e., the true) mode of understanding is an essential and indispensable part of what is to be meant by truth in that situation and problem. Both the notion of a fixed and definite absolute truth and that of a fixed and definite method for establishing and understanding truth are therefore not valid. Rather, each truth must contain within it the true mode for its being established and understood, in such a way that, without this mode, that truth has no meaning. If truth has the character described here, it may be that our whole mode and level of understanding things in general, which has evolved in response to a certain range of specific kinds of problems, practical and theoretical, is not adequate to the problem of understanding truth itself; i.e., of grasping the basic principle of truth. To do this would require of us not the mere repetition of some set of words, but rather (as in the case of the circle, discussed earlier) a real act of comprehension, in which truth would be seen as a totality—coming into being, as it actually does, from moment to moment, but always with radical differences in its essential characteristics. It seems evident that to understand truth in this way would be an extraordinarily difficult task. But if this is the way that truth really is, then such an understanding may well be just what is needed for seeing its basic principle. Our customary excessively narrow and

limited approach to the problem may therefore be what is responsible for the confused and self-contradictory nature of most of our ideas on the subject.

We may illustrate what is meant here by a kind of analogy. Generally speaking, science develops by a series of small steps of understanding that may be compared to the lighting of candles, each of which illuminates some small domain in the surrounding darkness for a short period of time (until it burns out). Every now and then there is a flash of understanding, which, like lightning, illuminates a whole field of study, and so brings about a fundamental change in our basic conceptions in that field. It is with the memory of what was seen during these occasional flashes that we usually try to direct our steps beyond the regions rendered visible by our candles. Meanwhile, however, the situation may have in some ways changed fundamentally, so that this earlier vision is no longer completely appropriate. To deal with this problem, perhaps it will be necessary eventually to emerge into a kind of understanding analogous to the steady sunshine, in the light of which there can be a full and adequate response in each moment to the whole of the ever-changing situation because no feature is, as it were, allowed to fall into a shadow dark enough to obscure its significance.

Such an unrestricted mode of understanding would probably enable us at each moment to see truth as an integral whole. On the other hand, our present modes of understanding are not only too rigid and fixed to follow the actual situation confronting us in its rapid and often unexpected changes, but also, being generally fragmented, specialized, and otherwise only partially valid, they must introduce a further kind of basic falsity by splitting the whole truth into pieces. Naturally, these kinds of falsity are not important in a wide variety of problems of relatively limited scope, because (as we have seen quite generally in science, for example), conceptions which are demonstrated to be false under new conditions or in a broader domain can still express true relationships under an approximate repetition of older conditions, or when the domain under consideration is restricted to what it was formerly. Nevertheless, it does not seem reasonable to suppose that a mode of understanding which inherently and unavoidably breaks truth into pieces and fixes the form as that of the pieces is likely to be adequate for the task of grasping truth as a totality in the kind of essential movement and process by which it is always coming into being anew, without fixed limits or characteristics.

Of course, it must be recognized that there exists a widespread feeling that truth really does not have the ever-changing and elusive character that we have suggested for it here. Or if it is admitted that truth is in fact an elusive and difficult notion, it is often felt that as far as scientific research is concerned, we can avoid such questions in the manner discussed in the beginning of this article; i.e., by basing everything on something else that is easier to grasp in a definite way (such as verifiability, falsifiability, instrumentalism, etc.). To a certain extent, such feelings arise out of a very natural desire to put things on some solid foundation, which can be asserted positively once and for all, so that we can then get on with our other tasks,

knowing that at least we are secure in our foundations. But as we have seen, there is good reason to suppose that this cannot be done. If this is so, we may be confusing ourselves by the effort to do the impossible; and as a result, we may be making our actual position more insecure than it really needs to be. In any case, what task could be more important than the development of a mode of understanding that will be adequate to the problem of perceiving what is the meaning of truth and how all our activities must be related to it?

(From the Collection of Essays in honour of Karl Popper *The Critical Approach to Science and Philosophy,* Macmillan 1964 p212)