



*Theory of Knowledge*  
(for *The Encyclopaedia Britannica*)

**Bertrand Russell**

THEORY OF KNOWLEDGE is a product of doubt. When we have asked ourselves seriously whether we really know anything at all, we are naturally led into an examination of knowing, in the hope of being able to distinguish trustworthy beliefs from such as are untrustworthy. Thus Kant, the founder of modern theory of knowledge, represents a natural reaction against Hume's scepticism. Few philosophers nowadays would assign to this subject quite such a fundamental importance as it had in Kant's "critical" system; nevertheless it remains an essential part of philosophy. It is perhaps unwise to begin with a definition of the subject, since, as elsewhere in philosophical discussions, definitions are controversial, and will necessarily differ for different schools; but we may at least say that the subject is concerned with the general conditions of knowledge, in so far as they throw light upon truth and falsehood.

It will be convenient to divide our discussion into three stages, concerning respectively (1) the definition of knowledge, (2) data, (3) methods of inference. It should be said, however, that in distinguishing between data and inferences we are already taking sides on a debatable question, since some philosophers hold that this distinction is illusory, all knowledge being (according to them) partly immediate and partly derivative.

## **I. THE DEFINITION OF KNOWLEDGE**

The question how knowledge should be defined is perhaps the most important and difficult of the three with which we shall deal. This may seem surprising: at first sight it might be thought that knowledge might be defined as belief which is in agreement with the facts. The trouble is that no one knows what a belief is, no one knows what a fact is, and no one knows what sort of agreement between them would make a belief true. Let us begin with belief.

# Belief.

Traditionally, a "belief" is a state of mind of a certain sort. But the behaviourists deny that there are states of mind, or at least that they can be known; they therefore avoid the word "belief", and, if they used it, would mean by it a characteristic of bodily behaviour. There are cases in which this usage would be quite in accordance with common sense. Suppose you set out to visit a friend whom you have often visited before, but on arriving at your destination you find that he has moved, you would say "I thought he was still living at his old house." Yet it is highly probable that you did not think about it at all, but merely pursued the usual route from habit. A "thought" or "belief" may, therefore, in the view of common sense, be shown by behaviour, without any corresponding "mental" occurrence. And even if you use a form of words such as is supposed to express belief, you are still engaged in bodily behaviour, provided you pronounce the words out loud or to yourself. Shall we say, in such cases, that you have a belief? Or is something further required?

It must be admitted that behaviour is practically the same whether you have an explicit belief or not. People who are out of doors when a shower of rain comes on put up their umbrellas, if they have them; some say to themselves "it has begun to rain", others act without explicit thought, but the result is exactly the same in both cases. In very hot weather, both human beings and animals go out of the sun into the shade, if they can; human beings may have an explicit "belief " that the shade is pleasanter, but animals equally seek the shade. It would seem, therefore, that belief, if it is not a mere characteristic of behaviour, is causally unimportant. And the distinction of truth and error exists where there is behaviour without explicit belief, just as much as where explicit belief is present; this is shown by the illustration of going to where your friend used to live. Therefore, if theory of knowledge is to be concerned with distinguishing

truth from error, we shall have to include the cases in which there is no explicit belief, and say that a belief may be merely implicit in behaviour. When old Mother Hubbard went to the cupboard, she "believed" that there was a bone there, even if she had no state of mind which could be called cognitive in the sense of introspective psychology.

## **Words.**

In order to bring this view into harmony with the facts of human behaviour, it is of course necessary to take account of the influence of words. The beast that desires shade on a hot day is attracted by the sight of darkness; the man can pronounce the word "shade", and ask where it is to be found. According to the behaviourists, it is the use of words and their efficacy in producing conditional responses that constitutes "thinking". It is unnecessary for our purposes to inquire whether this view gives the whole truth about the matter. What it is important to realise is that verbal behaviour has the characteristics which lead us to regard it as pre-eminently a mark of "belief", even when the words are repeated as a mere bodily habit. Just as the habit of going to a certain house when you wish to see your friend may be said to show that you "believe" he lives in that house, so the habit of saying "two and two are four", even when merely verbal, must be held to constitute "belief" in this arithmetical proposition. Verbal habits are, of course, not infallible evidences of belief. We may say every Sunday that we are miserable sinners, while really thinking very well of ourselves. Nevertheless, speaking broadly, verbal habits crystallise our beliefs, and afford the most convenient way of making them explicit. To say more for words is to fall into that superstitious reverence for them which has been the bane of philosophy throughout its history.

## **Belief and Behaviour**

We are thus driven to the view that, if a belief is to be something causally important, it must be defined as a characteristic of behaviour. This view is also forced upon us by the consideration of truth and falsehood, for behaviour may be mistaken in just the way attributable to a false belief, even when no explicit belief is present—for example, when a man continues to hold up his umbrella after the rain has stopped without definitely entertaining the opinion that it is still raining. Belief in this wider sense may be attributed to animals—for example, to a dog who runs to the dining-room when he hears the gong. And when an animal behaves to a reflection in a looking-glass as if it were "real", we should naturally say that he "believes" there is another animal there; this form of words is permitted by our definition.

It remains, however, to say what characteristics of behaviour can be described as beliefs. Both human beings and animals act so as to achieve certain results, e.g. getting food. Sometimes they succeed, sometimes they fail—, when they succeed, their relevant beliefs are "true", but when they fail, at least one is false. There will usually be several beliefs involved in a given piece of behaviour, and variations of environment will be necessary to disentangle the causal characteristics which constitute the various beliefs. This analysis is effected by language, but would be very difficult if applied to dumb animals. A sentence may be taken as a law of behaviour in any environment containing certain characteristics; it will be "true" if the behaviour leads to results satisfactory to the person concerned, and otherwise it will be "false". Such, at least, is the pragmatist definition of truth and falsehood.

## **Truth in Logic.**

There is also, however, a more logical method of discussing this question. In logic, we take for granted that a word has a "meaning"; what we signify by this can, I think, only be explained in behaviouristic terms, but when

once we have acquired a vocabulary of words which have "meaning", we can proceed in a formal manner without needing to remember what "meaning" is. Given the laws of syntax in the language we are using, we can construct propositions by putting together the words of the language, and these propositions have meanings which result from those of the separate words and are no longer arbitrary. If we know that certain of these propositions are true, we can infer that certain others are true, and that yet others are false; sometimes this can be inferred with certainty, sometimes with greater or less probability. In all this logical manipulation, it is unnecessary to remember what constitutes meaning and what constitutes truth or falsehood. It is in this formal region that most philosophy has lived- and within this region a great deal can be said that is both true and important, without the need of any fundamental doctrine about meaning. It even seems possible to define "truth" in terms of "meaning" and "fact", as opposed to the pragmatic definition which we gave a moment ago. If so, there will be two valid definitions of "truth", though of course both will apply to the same propositions.

The purely formal definition of "truth" may be illustrated by a simple case. The word "Plato" means a certain man; the word "Socrates" means a certain other man; the word "love" means a certain relation. This being given, the meaning of the complex symbol "Plato loves Socrates" is fixed; we say that this complex symbol is "true" if there is a certain fact in the world, namely the fact that Plato loves Socrates, and in the contrary case the complex symbol is false. I do not think this account is false, but, like everything purely formal, it does not probe very deep.

## **Uncertainty and Vagueness.**

In defining knowledge, there are two further matters to be taken into consideration, namely the degree of certainty and the degree of precision. All knowledge is more or less uncertain and more or less vague. These

are, in a sense, opposing characters: vague knowledge has more likelihood of truth than precise knowledge, but is less useful. One of the aims of science is to increase precision without diminishing certainty. But we cannot confine the word "knowledge" to what has the highest degree of both these qualities; we must include some propositions that are rather vague and some that are only rather probable. It is important, however, to indicate vagueness and uncertainty where they are present, and, if possible, to estimate their degree. Where this can be done precisely, it becomes "probable error" and "probability". But in most cases precision in this respect is impossible.

## **II. THE DATA**

In advanced scientific knowledge, the distinction between what is a datum and what is inferred is clear in fact, though sometimes difficult in theory. In astronomy, for instance, the data are mainly certain black and white patterns on photographic plates. These are called photographs of this or that part of the heavens, but of course much inference is involved in using them to give knowledge about stars or planets. Broadly speaking, quite different methods and a quite different type of skill are required for the observations which provide the data in a quantitative science, and for the deductions by which the data are shown to support this or that theory. There would be no reason to expect Einstein to be particularly good at photographing the stars near the sun during an eclipse. But although the distinction is practically obvious in such cases, it is far less so when we come to less exact knowledge. It may be said that the separation into data and inferences belongs to a well-developed stage of knowledge, and is absent in its beginnings.

## **Animal Inference.**

But just as we found it necessary to admit that knowledge may be only a characteristic of behaviour, so we shall have to say about inference. What a logician recognises as inference is a refined operation, belonging to a high degree of intellectual development; but there is another kind of inference which is practised even by animals. We must consider this primitive form of inference before we can become clear as to what we mean by "data".

When a dog hears the gong and immediately goes into the dining-room, he is obviously, in a sense, practising inference. That is to say, his response is appropriate, not to the noise of the gong in itself, but to that of which the noise is a sign: his reaction is essentially similar to our reactions to words. An animal has the characteristic that, when two stimuli have been experienced together, one tends to call out the response which only the other could formerly call out. If the stimuli (or one of them) are emotionally powerful, one joint experience may be enough-, if not, many joint experiences may be required. This characteristic is totally absent in machines. Suppose, for instance, that you went every day for a year to a certain automatic machine, and lit a match in front of it at the same moment at which you inserted a penny-, it would not, at the end, have any tendency to give up its chocolate on the mere sight of a burning match. That is to say, machines do not display inference even in the form in which it is a mere characteristic of behaviour. Explicit inference, such as human beings practise, is a rationalising of the behaviour which we share with the animals. Having experienced *A* and *B* together frequently, we now react to *A* as we originally reacted to *B*. To make this seem rational, we say that *A* is a "sign" of *B*, and that *B* must really be present though out of sight. This is the principle of induction, upon which almost all science is based. And a great deal of philosophy is an attempt to make the principle seem reasonable.



Whenever, owing to past experience, we react to *A* in the manner in which we originally reacted to *B*, we may say that *A* is a "datum" and *B* is "Inferred". In this sense, animals practise inference. It is clear, also, that much inference of this sort is fallacious: the conjunction of *A* and *B* in past experience may have been accidental. What is less clear is that there is any way of refining this type of inference which will make it valid. That, however, is a question which we shall consider later. What I want consider now is the nature of those elements in our experiences which, to a reflective analysis, appear as "data" in the above-defined sense.

## **Mental and Physical Data.**

Traditionally, there are two sorts of data, one physical, derived from the senses, the other mental, derived from introspection. It seems highly questionable whether this distinction can be validly made among data; it seems rather to belong to what is inferred from them. Suppose, for the sake of definiteness, that you are looking at a white triangle drawn on a black-board. You can make the two judgments: "There is a triangle there", and "I see a triangle." These are different propositions, but neither expresses a bare datum; the bare datum seems to be the same in both propositions. To illustrate the difference of the propositions: you might say "There is a triangle there", if you had seen it a moment ago but now had your eyes shut, and in this case you would not say "I see a triangle"; on the other hand, you might see a black dot which you knew to be due to indigestion or fatigue, and in this case you would not say "There is a black dot there." In the first of these cases, you have a clear case of inference, not of a datum.

In the second case, you refuse to infer a public object, open to the observation of others. This shows that "I see a triangle" comes nearer to being a datum than "There is a triangle there." But the words "I" and "see" both involve inferences, and cannot be included in any form of words

which aims at expressing a bare datum. The word "I" derives its meaning, partly, from memory and expectation, since I do not exist only at one moment. And the word "see" is a causal word, suggesting dependence upon the eyes; this involves experience, since a new-born baby does not know that what it sees depends upon its eyes. However, we can eliminate this dependence upon experience, since obviously all seen objects have a common quality, not belonging to auditory or tactual or any other objects. Let us call this quality that of being "visual". Then we can say: "There is a visual triangle." This is about as near as we can get in words to the datum for *both* propositions: "There is a triangle there", and "I see a triangle." The difference between the propositions results from different inferences: in the first, to the public world of physics, involving perceptions of others; in the second, to the whole of my experience, in which the visual triangle is an element. The difference between the physical and the mental, therefore, would seem to belong to inferences and constructions, not to data.

It would thus seem that data, in the sense in which we are using the word, consist of brief events, rousing in us various reactions, some of which may be called "inferences", or may at least be said to show the presence of inference. The two-fold organisation of these events, on the one hand as constituents of the public world of physics, on the other hand as parts of a personal experience, belongs to what is inferred, not to what is given. For theory of knowledge, the question of the validity of inference is vital. Unfortunately, nothing very satisfactory can be said about it, and the most careful discussions have been the most sceptical. However, let us examine the matter without prejudice.

### **III. METHODS OF INFERENCE**

It is customary to distinguish two kinds of inference, Deduction and Induction. Deduction is obviously of great *practical* importance, since it

embraces the whole of mathematics. But it may be questioned whether it is, in any strict sense, a form of inference at all. A pure deduction consists merely of saying the same thing in another way. Application to a particular case may have importance, because we bring in the experience that there is such a case—for example, when we infer that Socrates is mortal because all men are mortal. But in this case we have brought in a new piece of experience, not involved in the abstract deductive schema. In pure deduction, we deal with  $x$  and  $y$  not with empirically given objects such as Socrates and Plato. However this may be, pure deduction does not raise the problems which are of most importance for theory of knowledge, and we may therefore pass it by.

## **Induction.**

The important forms of inference for theory of knowledge are those in which we infer the existence of something having certain characteristics from the existence of something having certain other characteristics. For example: you read in the newspaper that a certain eminent man is dead, and you infer that he is dead. Sometimes, of course, the inference is mistaken. I have read accounts of my own death in newspapers, but I abstained from inferring that I was a ghost. In general, however, such inferences are essential to the conduct of life. Imagine the life of a sceptic who doubted the accuracy of the telephone book, or, when he received a letter, considered seriously the possibility that the black marks might have been made accidentally by an inky fly crawling over the paper. We have to accept merely probable knowledge in daily life, and theory of knowledge must help us to decide when it really is probable, and not mere animal prejudice.

## **Probability.**

Far the most adequate discussion of the type of inference we are considering is obtained in J. M. Keynes's *Treatise on Probability* (1921). So superior is his work to that of his predecessors that it renders consideration of them unnecessary. Mr. Keynes considers induction and analogy together, and regards the latter as the basis of the former. The bare essence of an inference by analogy is as follows: We have found a number of instances in which two characteristics are combined, and no instances in which they are not combined; we find a new instance in which we know that one of the characteristics is present, but do not know whether the other is present or absent; we argue by analogy that *probably* the other characteristic is also present. The degree of probability which we infer will vary according to various circumstances. It is undeniable that we do make such inferences, and that neither science nor daily life would be possible without them. The question for the logician is as to their validity. Are they valid always, never or sometimes? And in the last case, can we decide when they are valid?

## **Limitation of Variety.**

Mr. Keynes considers that mere increase in the number of instances in which two qualities are found together does not do much to strengthen the probability of their being found together in other instances. The important point, according to him, is that in the known cases the instances should have as few other qualities in common as possible. But even then a further assumption is required, which is called the principle of limitation of variety. This assumption is stated as follows : "That the objects in the field, over which our generalisations extend, do not have an infinite number of independent qualities; that, in other words, their characteristics, however numerous, cohere together in groups of invariable connection, which are finite in number." It is not necessary to regard this assumption as *certain*; it is enough if there is some finite probability in its favour.

It is not easy to find any arguments for or against an *a priori* finite probability in favour of the limitation of variety. It should be observed, however, that a "finite" probability, in Mr. Keynes's terminology, means a probability greater than some numerically measurable probability, e.g. the probability of a penny coming "heads" a million times running. When this is realised, the assumption certainly seems plausible. The strongest argument on the side of scepticism is that both men and animals are constantly led to beliefs (in the behaviouristic sense), which are caused by what may be called invalid inductions; this happens whenever some accidental collocation has produced an association not in accordance with any objective law. Dr. Watson caused an infant to be terrified of white rats by beating a gong behind its head at the moment of showing it a white *rat* (*Behaviourism*). On the whole, however, accidental collocations will usually tend to be different for different people, and therefore the inductions in which men are agreed have a good chance of being valid. Scientific inductive or analogical inferences may, in the best cases, be assumed to have a high degree of probability, if the above principle of limitation of variety is true or finitely probable. This result is not so definite as we could wish, but it is at least preferable to Hume's complete scepticism. And it is not obtained, like Kant's answer to Hume, by a philosophy *ad hoc*; it proceeds on the ordinary lines of scientific method.

## **Grades of Certainty.**

Theory of knowledge, as we have seen, is a subject which is partly logical, partly psychological; the connection between these parts is not very close. The logical part may, perhaps, come to be mainly an organisation of what passes for knowledge according to differing grades of certainty: some portions of our beliefs involve more dubious assumptions than are involved in other parts. Logic and mathematics on the one hand, and the facts of perception on the other, have the highest grade of certainty; where

memory comes in, the certainty is lessened; where unobserved matter comes in, the certainty is further lessened; beyond all these stages comes what a cautious man of science would admit to be doubtful. The attempt to increase scientific certainty by means of some special philosophy seems hopeless, since, in view of the disagreement of philosophers, philosophical propositions must count as among the most doubtful of those to which serious students give an unqualified assent. For this reason, we have confined ourselves to discussions which do not assume any definite position on philosophical as opposed to scientific questions.

