The nature of scientific theory

Siddharth G. Chatterjee

This article explores Einstein's views on the nature of scientific theory, and is directed towards students and researchers in the physical sciences and engineering. A majority of scientific theories belong to the constructive type whose axiomatic foundation consists of empirically observed principles and imaginary or speculative hypotheses. On this base is erected a logical superstructure (e.g. calculus) whose object is to make concrete predictions which can be tested against experiential data. A favourable comparison will point to validity of the hypotheses while an unfavourable one will require their revision. A constructive theory holds temporarily and will eventually be replaced with a more comprehensive approach. In contrast, in a theory of principle, the axiomatic foundation consists solely of principles of nature that are derived from human experience. Unless these principles are found to be false in the future, the security of such a theory is guaranteed. An example of a constructive theory is the kinetic theory of gases while classical thermodynamics and the theory of relativity belong to the class of principle theories.

This essay, which lays out the nature of a scientific theory, is partially motivated by a recent article by Joyce Lucas-Clark entitled 'Framing the discussion: what to tell students about science'[1]. It is intended for students of science and researchers who will obtain an overall and coherent view of what a scientific theory is and discover that its characteristic features, discussed in this commentary, are common in the physical sciences and engineering.

Lucas-Clark writes about science as follows:

> 'Science refers to knowledge that can be demonstrated, objectively, in the concrete, factual realm of reality. Matters that are inherently outside of objective reality, those things that are subjective or internal and cannot be demonstrated objectively, are outside of the realm of natural science.'

Further on, she states:

> 'Scientific inquiry is characterized by objectivity. A scientist, ideally, does not become attached to a certain hypothesis and seek to prove it. Rather, he or she adopts multiple working hypotheses and seeks to eliminate them. Scientists also accept that there will always remain a degree of uncertainty in their conclusions. It is the hallmark of a good scientific principle or theory that one can state clearly how it could be disproved.'

In this paper, we will subject these thoughts to a critical reflection.

All scientific knowledge of objective reality is acquired through the senses, and its interpretation by the mind of a perceiving human subject. Without the human being, there would be no such thing called 'science', since science is a creation of humans. All data from the objective world are thus filtered, coloured and processed through the physiological and psychological (i.e. conceptual) lenses of the perceiver. Since the most important task of science is to impart order into the panoply of external observations and bring them into a coherent or logical scheme so as to make them understandable to the human mind, let us examine the nature of a scientific theory. In his essay 'What is the theory of relativity?', Einstein[2] distinguished between two types of scientific theories. He wrote:

> 'We can distinguish various kinds of theories in physics. Most of them are constructive. They attempt to build up a picture of the more complex phenomena out of the materials of a relatively simple formal scheme from which they start out. Thus the kinetic theory of gases seeks to reduce mechanical, thermal, and diffusional processes to movements of molecules – i.e., to build them up out of the hypothesis of molecular motion. When we say that we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question.'
>
> 'Along with this most important class of theories there exists a second, which I will call "principle-theories". These employ the analytic, not the synthetic, method. The elements which form their basis and starting-point are not hypothetically constructed but empirically discovered ones, general characteristics of natural processes, principles that give rise to mathematically formulated criteria which the separate processes or the theoretical representations of them have to satisfy. Thus the science of thermodynamics seeks by analytical means to deduce necessary conditions, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible.'
>
> 'The advantages of the constructive theory are completeness, adaptability, and clearness, those of the principle theory are logical perfection and security of the foundations. The theory of relativity belongs to the latter class.'

According to Einstein, at the base of most scientific theories (i.e. of the constructive type), there are certain hypotheses whose nature is arbitrary, i.e. they are speculative constructions or free inventions of the human mind[3]. The validity of such hypotheses cannot be proved directly, but is indirectly confirmed by agreement between predictions of the theory and external facts. In a letter written to his friend Maurice Solovine on 7 May 1952, Einstein[4] offered a sketch of the nature of a scientific theory. Figure 1, taken from the work of Mittal[5] on the kinetics of hemicellulose extraction in the autohydrolysis of sugar maple wood, shows a practical adaptation of Einstein's scheme.

We see that the foundation of the theoretical edifice consists of two types of

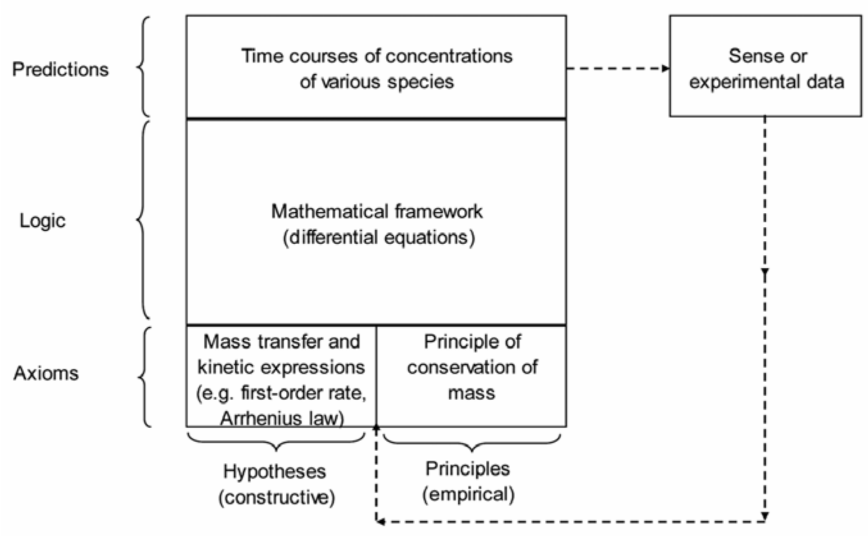

Figure 1. Structure of a scientific theory of the constructive type[5].

'blocks' or axioms. In the first type are empirically observed principles like the principles of conservation of mass, momentum and energy, which are generalizations of centuries of human experience. The second type of 'block' consists of hypotheses that are a priori, imaginary constructions of the human mind, whose validity as mentioned before, cannot be confirmed directly. What types of hypotheses are advanced, and there is freedom in doing so, depend purely on human intuition, and rest on a sympathetic understanding of the experiential data. There is no logical way of formulating or deriving the hypotheses, and it is here that a subjective factor enters into a scientific theory. We will say more on this aspect later. On the axiomatic foundation, which consists of principles and hypotheses, is erected a logical procedure or superstructure (e.g. mathematical reasoning) whose aim is to arrive at exact theoretical statements. Note that once a type of logical system has been adopted (e.g. calculus), the rules of that particular system have to be followed strictly, very much like the rules of grammar in language, i.e. there is no freedom of choice in this realm. The logical procedure finally culminates in predictions of the theory, whose test is a comparison between such predictions and sense or experiential data. If the comparison is favourable (and there can be some subjectivity in judging this), the theory is considered to be valid; otherwise the hypotheses have to be revised and the whole procedure repeated. Thus, in Figure 1, the motion is clockwise, viz. there is no logical method capable of revealing or deducing the axioms of a theory starting from empirical facts, i.e. of going in the counterclockwise direction. As Einstein[4] comments:

> 'Psychologically the A's (axioms – author) depend on the E's (experiences – author). But there is no logical route leading from the E's to the A's, but only an intuitive connection (psychological), which is always "re-turning".'

No theory that has constructive elements in its base can be final, for there is always the chance that future measurements will disprove its predictions. This fact attaches a pathos to a scientific theory since it is tenuous and holds the attention of the world only briefly, and which will eventually be superseded by a more comprehensive approach. This is in contrast to a mathematical theorem like that of Pythagoras, which is eternal since it does not depend upon external experience for its truth, and can be proved from logical considerations alone. Another feature of a constructive theory is that there can be more than one such theory based on different speculative constructions or premises that agree with experimental data taken at a particular level to a prescribed degree of tolerance. Since direct verification of the various hypotheses of the theory or measurement at a finer level is often not possible due to limitations in technology, there is no way of privileging one theory over the other, which are thus essentially of a fictitious character[3]. Frequently, there are assumptions made in a theory that are mutually contradictory, e.g. the theory may contain both true and false propositions. One example is the assumption of interfacial equilibrium in the dissolution of a gas into a liquid, which actually is a rate (i.e. non-equilibrium) process. Another example is the concept of an ideal gas, which contains the true proposition that the gas is made up of molecules, and the false proposition that the molecules are perfectly elastic spheres[6]. Such assumptions are invoked in order to simplify the theoretical treatment, which would otherwise become complicated. We thus arrive at the conclusion that a constructive theory has both subjective and objective elements, and is a creation of both the intuitive and logical faculties of the human mind.

In a theory of principle on the other hand, the axiomatic foundation consists solely of general principles that are directly derived from human experience. Unless these principles are found to be false in the future, the security of such a theory is assured since it will be impossible to overthrow it. Einstein cited the examples of classical thermodynamics and relativity as theories belonging to this class. The former is based on the impossibility of a perpetual motion machine, while the latter is based on the principle that the laws of physics are the same for all observers – both of which are empirical observations. To this list of examples can be added Marx's profound critique of political economy[7]. The theory of capital, as developed by Marx, hinges on the single concept of value, which is the socially necessary labour time required to produce a commodity. This concept is deduced from the observation of the exchange of commodities between individuals in a society under the sway of the market, i.e. it is a 'real' abstraction.

An extensive discussion of the distinction between constructive and principle theories has recently been provided by Frisch[8], who has also proposed a further refinement of Einstein's views. According to theoretical physicist David Bohm[9], the word 'theory' is derived from the Greek 'theoria', which has the same root as 'theatre'. It means to view or to make a spectacle, and is a form of insight, i.e. a way of looking at the world, and not a form of knowledge of how the world is.

Theories (i.e. insights) are neither true nor false – they are clear in certain domains and unclear when extended beyond these domains.

To conclude, the remarks on the nature of scientific theory advanced in this paper, which also have a bearing on the philosophy of science, may not be applicable to sciences like biology and related disciplines. Over the past few decades, an extensive literature has grown pertaining to the structure of scientific theory in biology. There are two ways to present a philosophy of biology in the opinion of Mayr[10]. One can base it on the chief philosophical concepts of biology like population thinking, the dual causation of biological processes, teleology or its absence, reductionism, modes of selection, pluralism, prediction, emergence, etc. An alternative foundation would be that of major biological phenomena or processes like the nature of life, genes, phenotype development, gene regulation, speciation, adaptation, biodiversity, extinction, etc. Unlike physics or chemistry, laws are rare or nonexistent in biology since exceptions to such 'laws' are very common according to Mayr[10], and Craver and Darden[11]. We refer the reader, interested in such matters, to the works of these and other philosophers of biology like David Hull[12].

1. Lucas-Clark, J., *Thought & Action*, 2010, **26**, 123–125.
2. Einstein, A., In *Ideas and Opinions*, Crown Publishers, New York, 1954, pp. 227–232.
3. Einstein, A., In *Albert Einstein: Philosopher–Scientist* (ed. Schilpp, P. A.), The Library of Living Philosophers, Evanston, Illinois, 1949.
4. Einstein, A., *Letters to Solovine*, Citadel Press, New York, 1993, pp. 135–139.
5. Mittal, A., Ph D thesis, SUNY College of Environmental Science and Forestry, Syracuse, New York, 2006, p. 67.
6. Giere, R., *Stud. Hist. Philos. Mod. Phys.*, 2011, **42**, 211–212.
7. Marx, K., *Capital*, International Publishers, New York, 1967, vols 1–3.
8. Frisch, M., *Stud. Hist. Philos. Mod. Phys.*, 2011, **42**, 176–183.
9. Bohm, D., *Wholeness and the Implicate Order*, Routledge, London, 2002.
10. Mayr, E., *Science*, 1999, **285**, 1856–1857.
11. Craver, C. F. and Darden, L., *Stud. Hist. Philos. Biol. Biomed. Sci.*, 2005, **36**, 233–244.
12. Hull, D. L., *Science*, 1999, **284**, 1131–1133.

Siddharth G. Chatterjee is in the Department of Paper and Bioprocess Engineering, SUNY College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, New York 13210, USA.
e-mail: schatterjee@esf.edu