

Evaluating Scientific Theories

Russell Berg has fifteen criteria for scientificness and he knows how to use them

The 'scientific method' is a group of methods and procedures. But since Thomas Kuhn argued in the 1960s that the concept of 'falsification' formulated by Karl Popper is insufficient on its own to determine the scientificness of an idea, there has been no method of distinguishing scientific theories from non-scientific ones. Kuhn himself muddied the waters by rejecting the established rules for determining scientific results, to broaden the conception of science to include economics and psychoanalysis. The problem with this, as Kuhn admitted, was that it makes it extremely difficult to distinguish between science and pseudo-science. Examples of the consequences are that in America creationists are arguing that Creation Science and Darwinian Evolution should be given equal time in school biology lessons. Alternatively, theoretical physicists have produced concepts such as string theory, justified purely by its mathematical elegance, without any experimental evidence. This is perhaps also pseudo-science.

As if this is not enough, scientific ideas such as Marshall's theory that stomach ulcers and stomach cancer are caused by a bacterium were shunned for many years due to the combined efforts of vested interests (ie pharmaceutical companies), plus senior doctors' and scientists' fixed beliefs about the possibility of microbes surviving in low pH, despite the evidence. Meanwhile, alternative medicine with little scientific merit – homeopathy, aroma therapy etc – is funded by the NHS. What have the philosophers of science been doing all this time?

From a utilitarian perspective a method for quantifying scientificness would be worthwhile if it leads to a clearer distinction between science and pseudo-science, rejection of ineffective and unscientific medicine and a better grasp of the scientific method amongst the general public. It would mean new theories being judged on their scientific merit rather than being hyped or hindered by vested interest and subjective prejudice. I see no theoretical reason why the quantification of scientificness should be less reliable than the quantification of risk which currently takes place in health and safety and food safety.

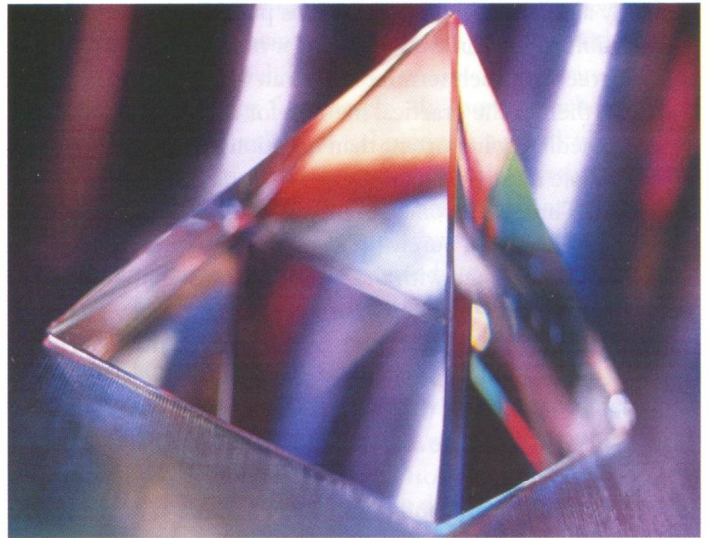
The next problem is what is the best method for quantifying the quality of being scientific. I've chosen a simple descriptive method so that as many people as possible may evaluate the evaluation. In a more academic exercise, I would have chosen a more enumerative approach which would provide significance levels when comparing theories for scientific quality, such as non-parametric enumerative statistics, discussing the merits of a Wilcoxon test against each criteria vs a Kruskal-Wallis one-way analysis of variance by ranks, or even the Friedman two-way analysis of variance by ranks. But that's for another day.

However to obtain a better tool for a job, we have to start with a basic tool. The wheel had to be invented before the pneumatic tyre. Therefore, the following fifteen criteria may be used to evaluate the scientificness of theories, and a theory can be scored against each criteria. When the aggregate score is known, the theory will have a 'Scientific Quotient' (SQ).

Fifteen Criteria For Scientificness

1) Does the theory use natural explanations?

Thales of Miletus, the first recorded natural philosopher, believed that natural events have natural explanations, not divine. This rejection of explanations invoking gods or spirits led to the need for natural explanations and the development of the scientific method. Untestable supernatural explanations act as stoppers which prevent or retard further enquiry or research.



2) Does the theory use rational, inductive argument?

Rational *deductive* arguments are based on logical inference rather than appeal to authority. Rational *inductive* arguments are uncertain but plausible explanations based on evidence concerning cause and effect claims. A theory must use inductive argument to be scientific (*cf* 9). An early example is Anaximander's claim that man must have been born from animals of another kind, as humans alone require a long period of nursing.

3) Is the theory based on an analytical reductionist approach rather than a synthetic approach?

Reductionism is the attempt to understand complex things by analysing them in terms of their parts or simplest aspects. Reductionism was first used by Thales, when he claimed that all is water. A synthetic approach is the opposite of reductionism, in that it attempts to build a system of explanation from theory and usually results in added layers of complexity normally based on argument alone rather than substantial evidence. Examples are Plato's forms, Freudian psychoanalysis, Marxist historicism and string theory evoking extra dimensions.

4) Is the theory self-consistent?

According to Aristotle, the Principle of Non-Contradiction is the most fundamental principle of logic and thus of thought. The need for consistency is a manifestation of this principle.

Most theories are self-consistent, but occasionally a theory can be internally inconsistent. Such theories are however

sometimes useful as transitional ideas. Take Rutherford's solar system model of the atom, in which electrons are imagined to orbit the nucleus of the atom in a similar manner to planets orbiting the sun. This model is inconsistent because electrons orbiting the nucleus would emit electromagnetic radiation, which would result in loss of kinetic energy, causing the electrons to slow down and fall towards the nucleus, quickly colliding with it. But the solar system model was a useful stimulus for further thought about the structure of the atom.

5) Does the theory involve a mechanistic approach?

A mechanistic approach explains how a proposed idea works. This is in contrast to an approach which simply states that a situation *is so* (or less dogmatically, may be so). A good example of a mechanistic approach is the kinetic theory of gases. This states that as the temperature of a gas rises the molecules move faster so that they are more likely to collide; hence they become more reactive. This also explains why the pressure increases with temperature if the volume of a gas remains constant, as the molecules collide more frequently with walls of the container as the temperature rises.

By contrast, a non-mechanistic approach is often taken by extreme reductionism, such as Thales's claim that all is water. Sometimes a theory is formulated without an explanation of how it works, such as Newton's law of gravity and Darwin's theory of evolution; but good scientific theories will become mechanistic as new observations are obtained or ideas are proffered.

6) Are qualities given quantities?

Pythagoras first successfully assigned quantity to quality when he discovered that the pitch of a note depends on the length of the string which produces it: hence concordant intervals in musical scales are produced by simple numerical ratios. According to Arthur Koestler, this first successful reduction of quality to quantity was the first step towards the mathematization of human experience, and therefore was the beginning of science.

7) Is the theory the simplest way to explain the data?

The first person to formulate this principle was William of Ockham, hence it's referred to as *Ockham's Razor*. (Ockham's formulation was '*entia non sunt multiplicanda praeter necessitatem*': 'entities should not be multiplied beyond necessity'.) It has been extended to the idea that the best interpretation of a phenomenon should make as few assumptions as possible. This principle is also referred to as the Law of Parsimony or the Law of Succinctness. Ockham used it to argue that ideal forms in the mind of God were unnecessary for entities in this world to exist.

In *Just Six Numbers*, Martin Rees, Britain's Astronomer Royal, discusses six physical constants fundamental to the structure of the universe, such as the speed of light. If any of these values were slightly different the universe would not be capable of supporting life. However, the probability of all six constants randomly having a value that would together give rise to a life-supporting-universe is very low, so how did it happen?

Possible explanations are:

- God gave the constants their values.
- The constants were set by another intelligent designer.
- The universe is a computer simulation.
- This universe is one of many in a multiverse of universes,

each with different values for these six constants.

e) This is the only universe, and the constants have their value by pure chance.

f) This is the only universe, and the values of the six constants are not independent but fundamentally linked together in ways which we currently do not understand, due to theories of physics which have not yet been formulated.

The present question is, which of these six theories is the simplest, all other things being equal? They would not be equal if we started to pick up information from another universe, or there was strong evidence for a yet-unknown theory of physics that explains how these constants are linked.

Theories a) to d) all involve extra entities not required by theories e) and f). So the question now becomes, is e) or f) the simpler theory? I think that saying that the six constants are linked actually produces a simpler model of the universe, so according to this interpretation, theory f) should be the one investigated first.

8) Does the theory conform to existing scientific understanding?

Scientific theories do not stand alone, but relate to other scientific theories, hence it is not adequate for a scientific theory to be merely self-consistent: the theory should also be consistent with the existing body of scientific knowledge. However, sometimes the evidence for an incompatible new theory is so overwhelming that an existing theory has to be amended, revised, or even dropped, so the situation isn't simple.

When Alfred Wegener first proposed Continental Drift in 1912 to explain why the coast of Africa seems to fit into the coast of South America like a jigsaw piece, the majority of geologists did not accept that masses as large as continents



could move round the surface of the Earth. However, after the Second World War, evidence was discovered that supported plate tectonics. Paleomagnetic studies found a striped pattern of magnetic reversals in the Earth's crust, which showed that the crust was moving around. Also, most seismic activity was found to occur along the lines where the plates would be colliding. The anti-mobilists' understanding had to be revised in the face of the new evidence.

A general rule of thumb is that the greater and the more fundamental changes required to existing scientific thinking, the more conclusive the evidence must be for the challenger theory to obtain scientific orthodoxy, as this will only be possible after the more established theories have been reviewed. It is unlikely that existing theories will be reviewed if a new conflicting theory is proffered without any substantial evidence.

9) Is the theory based on observed data?

The gathering of data is the first stage of the *inductive process* developed by Francis Bacon and Thomas Hobbes. It became the basis of Newtonian science, and empiricism generally.

This is where science parts from philosophy. In philosophy, theories can be based purely on speculation without the burden of data-gathering. Plato's division between body and soul and his theory of forms were products of speculation rather than observation or gathered data, for example. However, science is concerned with what may be observed.

10) Has the theory been tested?

At the beginning of the eighteenth century Georg Stahl proposed the existence of 'phlogiston' to explain why some substances burned and others do not. According to this theory, substances which burnt contained phlogiston, which was released by the fire. The problems were that phlogiston had never been isolated.

The quantification of qualities (see 6) had then barely entered chemistry. But Lavoisier tested the theory of phlogiston by carefully making measurements, and he found it wanting. Lavoisier showed that when metal is burned it *increases* in weight, and the air in a closed container suffers a corresponding loss of weight. So the metal doesn't lose phlogiston by burning it; rather, it gains something else. After further experimentation, Lavoisier proved that only one fifth of the air could support combustion, and he concluded that it was this 'oxygen' which combined with the metal during burning. The theory of gases had come into being, and the theory of phlogiston was dead.

There was a similar occurrence in 1948, when Hoyle, Bondi and Gold proposed the Steady State Theory to explain the observation of galaxies moving away from each other. They claimed that the universe had always existed in the state it was now, and that matter formed from nothing in the spaces between the galaxies, which coalesced into stars and new galaxies, pushing the others away and making space for more matter to form. The problem was this theory hardly made any predictions which could be tested (see 14) – except for the creation of matter between galaxies, which had never been observed and would be very difficult to observe in any case.

However the alternative Big Bang Theory made testable predictions, one of the most important being that there would be background radiation from the Big Bang. The background

radiation was discovered by Penzias and Wilson by accident in 1964, in the microwave range, at about 3.5° above absolute zero. Also in the early 1960s, radio astronomer Martin Ryle discovered that the further away (and so back in time) he looked, the greater the percentage of radio galaxies. This showed that the universe had changed with time. The Steady State theory suffered a similar fate to the phlogiston theory.

11) Do the results of the tests plausibly support the theory?

Homeopathy was invented at the beginning of the 19th Century by Samuel Hahnemann, who proposed that ill people could be treated by medicines that would be harmful to healthy people. Even more controversial was his belief that the more dilute the medicine the more potent the vanishing drug. In contemporary homeopathy the solution is diluted to half its strength thirty times, making it unlikely that there is even one molecule of the 'active' ingredient in the final medicine. Homeopaths get round the problem of the lack of medicine in the medicine by claiming that water has memory. This conflicts with existing scientific understanding (see 8), yet testing by the double blind method does show that homeopathy is of some benefit. However, this benefit is of equivalent power to the placebo effect. Hence there is not adequate evidence for the claim that water has memory. (When homeopathy started, conventional medicine was less scientific and included many untested treatments which often did more harm than good, so the more 'neutral' homeopathy rapidly gained popularity. However, conventional medicine has progressed scientifically but homeopathy has not, being trapped in a blind alley.)

12) Are the experiments repeatable by different experimenters?

In 1989 two scientists in America, Fleischmann and Pons, claimed they'd achieved nuclear fusion at relatively low temperature – in a standard laboratory, rather than at the exceedingly high temperatures which occur in a star or a particle accelerator. If cold fusion is possible, the world's energy supply would be virtually limitless. However despite numerous attempts by other scientists, none succeeded in repeating their 'results'.

13) Can the theory be falsified?

Experiments can be set up to disprove some theories, but others might not be potentially falsifiable. Theories that cannot be disproved by experiments fall into two categories: those intrinsically immune to experimentation, and those that cannot be disproved by experimentation due to lack of technology.

The concept of falsification was formulated by Karl Popper when investigating the differences between *dogmatic* and *critical* thinking. Dogmatic thinkers, including the followers of Marx and Freud, try to interpret all events in terms of their favoured theory or beliefs, whilst a critical thinker tries to find the *flaws* in theories – *especially* their favoured ones. Popper gives Einstein as an example of a critical thinker, when Einstein said "If the red-shift of spectral lines due to the gravitational potential should not exist, then the general theory of relativity will be untenable."

14) Does the theory have predictive elements?

Without a predictive element, science would be an esoteric or speculative subject, the output of which would only be higher-definition 'Just So Stories'. It's the predictive element which

gives science its practical value, allowing us to say how materials will behave or what various reactions will produce. This made possible the technology which changed the world during the industrial and information revolutions. Physics underpins the technology of locomotives and jets.

As medicine has become more scientific it has been more successful. Dr Alexander Fleming observed the mould *Penicillium* retard the growth of the bacterium *Staphylococcus*, and predicted that penicillin could be used to treat bacterial disease. Also, Marshall's theory that stomach ulcers are caused by bacteria and hence are treatable by antibiotics, has proved correct.

15) How accurate are the predictions based on the theory?

Scientific theories are not the only explanatory systems that produce predictions. Long before there was science there were oracles, the most famous being the Oracle of Apollo at Delphi. However, her prophecies were not subject to the statistical analysis used to test modern scientific predictions. Also, like the quatrains of Nostradamus, Oracular predictions were ambiguous and relied on equivocation. When King Croesus of Lydia asked the Oracle what would happen if he went to war against Persia, the Oracle prophesied that a great empire would fall. She just didn't say whose great empire.

The predictions based on the laws of motion of Newtonian physics, for instance, are very different. These laws were used to accurately predict when Halley's comet would next be visible.

Unfortunately not all theories which claim to be scientific are as accurate in their predictions as Newton's. Marxist theory (which Marxists claim to be scientific) claims that it can predict future historical periods: in Marxist theory the feudal period is succeeded by the capitalist period, which is succeeded by the socialist period, which in turn is succeeded by the communist period. But according to Marxist theory the countries which would be the first to undergo socialist revolution would be the advanced capitalist ones, Britain, Germany or the United States, not the peasant-based economies of Russia or China. This prediction failed, even though it was a very broad theory.

Critical Qualifications Of The Criteria

Let us briefly compare some well-known theories by assigning scientific quotients according to each of these criteria:

	Evolution	Creationism	ID (Score out of 10.
1. Natural Explanation *	9	1	8 Stars indicate a
2. Rational Argument *	8	6	8 necessary criterion.)
3. Reductionist Approach	9	2	2
4. Self-Consistent *	10	10	10
5. Mechanistic Approach *	10	1	1
6. Qualities in Quantities	6	1	1
7. Simplicity	8	3	4
8. Conformity	9	2	4
9. Data Based *	9	2	3
10. Tested and Verified *	9	1	6
11. Supported by Test Results	6	1	4
12. Repeatability	1	1	1
13. Falsification	6	1	2
14. Predictive Elements	6	1	1
15. Accuracy of Predictions	4	1	1
TOTAL/150	110	34	56
SQ:	73	23	37
VERDICT: SCIENTIFIC?	YES	NO	NO

A disadvantage of this approach is the subjectivity in the weighting of the criteria and the scoring process. However this problem can be offset by choosing an expert panel to evaluate the theory against the criteria. (This is not meant to exclude an amateur from calculating a scientific quotient.)

There are other complications too. History shows us that whether or not a theory is scientific can change in the light of new evidence or new techniques. What is currently not testable can become testable, for example. The first six criteria given are intrinsic properties of theories, not alterable by new data or techniques. The criteria of *simplicity*, *conformity*, *falsification* and *predictive elements* are transitional, insofar as new data and techniques are highly unlikely to change this part of a theory's nature with time. The remaining five criteria are extrinsic properties that *are* likely to change as new data is gathered or new techniques become available.

The aspects of a theory's scientificness are not independent. For example, just because a theory is based upon observed and gathered data it does not necessarily mean that the theory is accurate or is the simplest (see 7). Moreover, the criteria are not of equal weight. Some of the criteria given above are necessary for a theory to be scientific, others more amorphaously influential. We can combine this scientific quotient scoring system with a star system in which all the *necessary* criteria for a theory being scientific are given a star (as shown), and so theories are unscientific if they do not pass all the starred criteria. These criteria include: Is the theory self-consistent? Is the theory based on data? Has the theory been tested? etc. However, a star system alone would not distinguish the degree of fulfilment of criteria between two competing theories, unlike the Scientific Quotient system. Before the background radiation from the Big Bang was discovered it was inconclusive which was the stronger theory. However, using the Scientific Quotient system, I think the Big Bang theory would still have had a higher score. It would have fared better on *simplicity*, a single creation then expansion being a simpler explanation than the continuous creation of matter. Also, at that time the Big Bang theory was more in tune with the rest of physics than matter being formed in interstellar space (violating the first law of thermodynamics), and so had a stronger fulfilment of criterion 8.

Furthermore, many theories at the boundaries of science would cease to be scientific having failed to obtain stars for 'Has the theory been tested?' Currently string theory and multiverse theory would fall into that category. And by the mechanistic criterion, Darwin's theory of evolution by natural selection could have been said to be unscientific until Watson and Crick discovered DNA. I would think it fairer to say these are untested or otherwise incomplete rather than claim that they are unscientific. If we acknowledge that some of the necessary criteria for being scientific are extrinsic (dependent on factors other than the theory itself), the claim that whether a theory is scientific or not could change with time. Or perhaps we can augment our vocabulary and say that there are *immature* scientific theories. As I say, this theory of evaluation is itself in its preliminary stages.

© RUSSELL BERG 2009

Russell Berg studied at the University of Leeds and is currently working as a food microbiologist.