1
The scientific method

Motivation

Most of you, the readers, have probably tuned in to the news at one point or another to hear the anchor person begin a story with the phrase "Today scientists reported," probably followed by a description of something completely outside of your everyday experience: colliding microscopic particles, genetic engineering, or statements about the behavior of giant galaxies or the Big Bang. How do scientists arrive at their conclusions on such topics? And perhaps more importantly, why should *you* believe anything they say?

The answers are important, for the topics covered in this book are indeed such things as the creation of the Universe, the violent death of massive stars, and so forth, which you will (probably) never personally experience. So an understanding of how science is done and why we trust the results is relevant to what will follow. Equally important is realizing that scientific results always represent a qualified "truth" (as opposed to an absolute truth). New discoveries that change how science views the universe are always just around the corner. The "scientific method" provides the recipe for generating a consistent set of unprejudiced ideas about how the Universe works; it is also the method by which the current ideas are replaced by new ones, based on observations of the real world.

In addition, the scientific method has a practical aspect that cannot, and should not, be neglected. Many aspects of our society have become increasingly intertwined with a manifold of technological improvements that lengthen and enrich life,[1] and we have come to expect (and demand) that every new such improvement will perform safely and reliably. To insure the safety and efficacy of these products most countries have instituted mechanisms to *test* whether these expectations are fulfilled and to insure that failing products are not distributed among the public. We expect these tests to be unprejudiced, repeatable and consistent, and so the standard approach in devising them must follow the scientific method.

Despite these protective mechanisms the market pullulates with wonderful life-enriching, disease-curing, products whose powers are said to derive from

some ancient magical or mystic principle; or else whose healing powers are supported by a befuddling mishmash of scientific "facts" judiciously taken out of context from appropriately chosen publications. Having had no unprejudiced, repeatable and independent verification of their properties, all such products (and their advocates) should be considered suspect. In the best cases these items are harmless, but sometimes they are dangerous,[2] and often they are fraudulent.

The intellectual and practical applications of the scientific method make it important for the public to have a basic understanding of the manner in which scientific theories are obtained, tested and accepted. With this type of knowledge the public can protect itself from non-scientific quackery, and impress upon the government the need to create and *enforce* policy against it. In many cases the lives of people depend on the reliability and thoroughness of this scrutiny.

It is not my intention to imply that all human intellectual pursuits should be centered on a scientific approach. But when gathering quantitative information about Nature, systematizing the data, and deriving from it a deeper understanding of the world around us, the scientific approach described below is the best method available.

Brief history

In his work "The History of Animals" the Greek philosopher Aristotle claimed that men and women had a different number of teeth, and that their internal organs were also different. These assertions were supported by long arguments. No verification was sought and, indeed, Aristotle believed that none was required, since the arguments were based on the author's philosophy, which was consistent and perfectly satisfactory. The possibility that these statements might actually be *wrong* never occurred to Aristotle, nor did the thought that careful observation and experimentation might be needed to decide the issue.

Early Greek philosophers (e.g. Thales of Miletus) noted that, in many instances, Nature exhibits a regular behavior. Most noticeable among these are, perhaps, the change of day into night and the progression of the seasons. This eventually gave rise to the idea of natural phenomena being determined by a set of rules, or by logical conclusions derived from these rules. The realization that Nature is not driven by the whims of the gods marks the beginning of science.

Accepting the existence of rules that describe the world around us, it is natural to ask how can we best discover these rules. One possible method is to argue that since the world around us can be explained by drawing conclusions from a set of first principles, these first principles ought to be part of a coherent and logically constituted philosophy. So all we have to do is guess the right philosophy, from it the behavior of the whole of Nature can be deduced.

The notion that the properties of the world around us can be obtained from a series of ideas, a philosophical edifice, is called the *deductive* method of reasoning.

Deductive reasoning

Argument is based on a rule, law, principle, or generalization. In other words, "I'm right because I said so."

Aristotle conducted some of the earliest studies of the world around us, interpreting his observations using this deductive approach. He did this by choosing a set of first principles, which he considered eminently clear, obvious and natural. Unfortunately they were also inappropriate for the task, as illustrated by the above example. Still, such was Aristotle's preponderance, that his deductive method was blindly followed for more than 13 centuries after his death, even today it is used (perhaps unwittingly) by many people. Deductive reasoning might be a good rule to follow when dictating human activities (*a la* Sherlock Holmes), but it is a very unreliable way of obtaining information about Nature.

A good example of the sort of logical traps hiding in deductive reasoning can be seen in what is known as a *syllogism*. A syllogism is a logically incorrect generalization. For example, one might state "All cats have fur," and then state "A dog has fur; therefore a dog is a cat." Though it may seem silly to us now, this type of argument resulting from the deductive approach formed the basis of science for centuries.

By the end of the Middle Ages the deficiencies inherent to the deductive method became increasingly apparent. It was during this period that the foundations for the modern approach to science were laid. While writing a series of treatises on papal power and civil sovereignty, William of Ockham published his "Razor": a prescription for cutting away any unnecessary ideas present in a description of Nature.[3] Later, the Renaissance witnessed the final demise of the deductive approach in scientific inquiries; in its stead an *inductive* approach was adopted, based on experimentation and careful analysis of the data.

Inductive reasoning

Arguments are based on experience or observation. In other words, "I'm right and I can do an experiment to prove it."

The shift in science from deductive to inductive reasoning was prompted by the various writings of Francis Bacon, and perhaps more forcefully by the *results* obtained by Galileo and Newton. The same basic approach used then (with minor alterations) is still followed today in most research. The reliability of scientific results we have come to expect is due to the inductive approach. Through it, the US Food and Drug Administration can determine whether a given medicine is safe, and under which conditions it is dangerous. By following the scientific method, we obtained knowledge about the resistance of materials which is used by architects and civil engineers when declaring a building safe or unsafe. Rules obtained from the method ensured that the Galileo spacecraft,

setting out from Earth at the appropriate speed and direction, would rendezvous with the planet Jupiter on December 7, 1995, after a journey of over six years and 3 billion kilometers.

The basic assumptions behind the modern scientific method are the existence of a reality outside our own bodies, and that said reality can be understood by the human mind. It is, of course, conceivable that we are all characters in somebody's dream, but this is something we cannot prove or disprove (were it true one cannot but commend the dreamer on the richness and vividness of his/her imagination ... and hope the alarm-clock does not go off). The reality of the universe around us is something that ultimately must be taken on faith. Still one can provide plausibility arguments such as the following one by W. Churchill[4] that considers the possibility that our Sun is the result of our fertile imaginations:

> ... happily there is a method, apart altogether from our physical senses, of testing the reality of the Sun. It is by mathematics. By means of prolonged processes of mathematics entirely separate from their senses, astronomers are able to calculate when an eclipse will occur. They predict by pure reason that a black spot will pass across the sun on a certain day. You go and look, and your sense of sight immediately tells you that their calculations are vindicated ... We have got independent testimony to the reality of the Sun. When my metaphysical friends tell me that the data on which the astronomers made their calculations were necessarily obtained originally through the evidence of their senses, I say 'No'. They might, in theory at any rate, be obtained by automatic calculating machines set in motion by the light falling upon them without admixture of the human senses at any stage ... I am also at this point accustomed to reaffirm with emphasis my conviction that the Sun is real, and also that it is hot – in fact as hot as Hell, and that if the metaphysicians doubt it they should go there and see.

In this book I will follow the development of modern cosmology starting with ancient creation myths, and ending with the currently accepted ideas based on Einstein's theory of relativity and the currently accepted theories of elementary particle physics. Throughout the narrative I will point out the gradual shift from the deductive to the inductive approach, discussing the reasons behind this shift and its consequences.

Science's approach to knowledge

The scientific method can be described as a recipe for providing a level of understanding of a part of Nature. The recipe itself is simple, but, as is often the case, the devil is in the details: the way in which the recipe is applied and the context within which this application occurs. I will start by listing my version of the steps, and then I will discuss each separately.

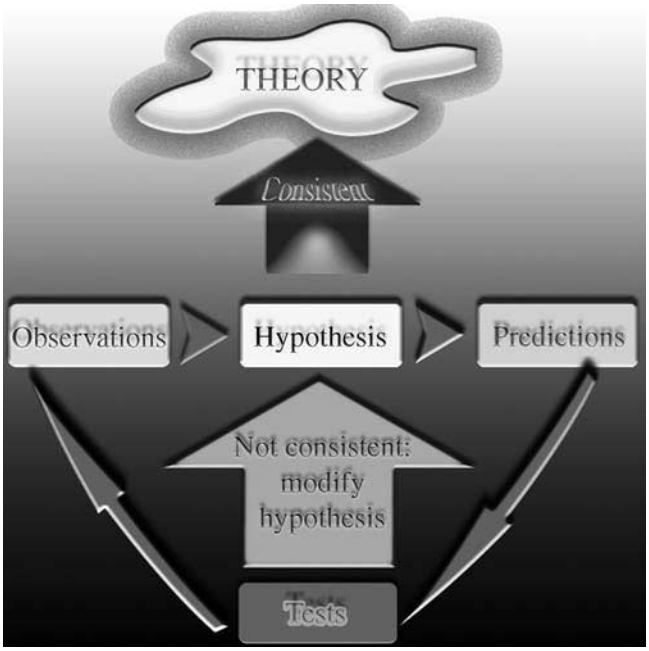

Figure 1.1 Diagram illustrating the scientific method.

The steps in the method (Figure 1.1) are the following:

1. Observe some aspect of the Universe.
2. Invent a tentative description, called a *hypothesis*, which is consistent with what you have observed. This might range from a fine tuning of existing ideas to a complete revamping of accepted knowledge (sometimes called a "paradigm shift").
3. Use the hypothesis to make predictions.[5]
4. Test those predictions by experiments or further observations, and modify the hypothesis in the light of these results.
5. Repeat steps 3 and 4 until there are no discrepancies between theory and experiment and/or observation. Once this type of consistency is achieved the hypothesis is *validated* and accepted as a new theory.

This method has several very attractive features:

- It is unprejudiced in the sense that the data from experiments and/or observations is the sole arbiter of whether a hypothesis or theory survives.
- It does not require superhuman powers. Any person with the patience and money to perform the experiments can test the reliability of a theory.

There is a common misconception that the scientific method is the "method of science," that is, the *way* science is done, as opposed to *why* we believe scientific results. Many of you may recall your elementary school science fair projects, where you were told to follow the "scientific method," defined as the process by which hypotheses are verified. Perhaps in the course of your project, you

discovered something unexpected and exciting, but were disappointed to find that your grade was lowered because you were unable to prove your initial hypothesis, and so failed to faithfully follow the "scientific method." Such incidents are all too common in science education, and are all the more unfortunate because, in truth, the actual practice of science is anything but methodical, being rather fairly chaotic. The "scientific method" is the recipe from which we can make sense of this chaos, choosing hypotheses that will be kept as "theories," and discarding those that are inconsistent with Nature.

This recipe, however, also raises many questions. For example, how are hypotheses concocted? How do we select among competing hypotheses? How do we obtain predictions from the hypothesis? What if a hypothesis explains the initial data but has no predictions? I will deal with these issues below.

Paradigms

When studying a part of Nature the observer does so with a certain set of loose preconceived ideas, a certain amount of background knowledge that he/she uses to make sense of the observations. This is often called a *paradigm*. This background knowledge can range from the very basic (e.g. far-away things look smaller) to the complex (e.g. the propagation of light in matter is determined by the interaction of light with subatomic particles). Using this background knowledge, the observer attempts an explanation of the phenomena and produces a hypothesis. It is important to note that this hypothesis, though created within the context of a certain body of knowledge, need not *agree* with that knowledge. Thus Einstein studied the properties of light using Newton's ideas of motion, but the hypotheses that flourished into the Special Theory of Relativity disagreed with Newton's basic assumptions about space and time, which were almost universally accepted as providing the basic description of the workings of Nature; Harvey obtained his revolutionary description of blood circulation only after careful anatomical observations motivated by Galen's ideas, etc.

Hypotheses and theories

As with all creative processes it is difficult, if not impossible, to explain or even describe how hypotheses are generated. Researchers have historically come up with their hypotheses in every conceivable way: inspiration, laboriously investigating data, even dreams.[6] The *origin* of the hypothesis, however, is immaterial (except perhaps as a field of study in psychology), what matters is whether the hypothesis describes the data, and can be used to make new accurate predictions that are subsequently verified. If these conditions are satisfied the hypothesis is accepted, otherwise it is discarded.

A little thought shows that no hypothesis can be proved *absolutely* true, for that would entail testing it in *all* possible ways under *all* possible circumstances.

In contrast, a hypothesis *can* be proved false for it makes predictions that may or may not be verified by experiment. To be scientifically useful, hypotheses must be *falsifiable*. Thus *all* scientific theories are constantly in peril of being proven wrong by new data or observations: experiments are the sword of Damocles for theories. This is a positive aspect of all scientific investigations for it provides a natural mechanism for testing and improving scientific knowledge.

An example of a falsifiable hypothesis is Newton's universal law of gravitation that predicts the planetary orbits. These predictions are verified, within experimental accuracy, for all members of the Solar System *except* Mercury. The orbit of this planet exhibits a slight deviation from the Newtonian predictions: Newton's hypothesis has been falsified. The General Theory of Relativity also provides predictions for all the planetary orbits, and these agree with observations, *including* the case of Mercury. Because of these (and many other) observations the General Theory of Relativity is now accepted as the best description of gravitation. Yet this does not mean that future experiments need to agree with its predictions. Should a discrepancy be confirmed, this theory will have been falsified and the search for a more accurate description of gravitational phenomena will begin.

An example of a non-falsifiable hypothesis is the one that claims the Moon is densely populated with little green men. These beings are then assumed to be perfectly in tune with human intentions so that whenever we attempt a way of finding them they would know of this in advance and thwart us. Should anyone on Earth look at the Moon, the green men have foreknowledge of this and they *all* hide. When the Apollo spacecrafts landed on the Moon they all moved to the dark side (of the Moon) beforehand, and obliterated all traces of their existence. If we were to observe the whole surface of the Moon simultaneously then, by the time our equipment is set, they would have dug tunnels deep into the Moon (again perfectly covering their tracks) and would stay there until we stopped observing. If we were to obliterate the Moon they would again know of our intentions ahead of schedule and leave, timing their trip so that we could not see them, etc. It is clear that by construction the existence of these little green men cannot be disproved, their existence has to be taken on faith and might be a matter for mystical or philosophical musing, but it has no place in a scientific discussion.

The experimental verification of hypotheses is paramount, and it is in this realm that many controversies arise. It is possible for an experiment to provide an apparent confirmation of a prediction, but this verification is later found to be the result of poor experimental design, a misinterpretation of the data, the result of extraneous effects that were not properly taken into account, etc. In order to minimize these potential problems scientists test predictions using many different experimental designs. The confirmation of a prediction by a given experiment will usually set off a flurry of new experiments also aimed at verifying this prediction, but using many different approaches. In addition the details of every such experiment are published, and the experimenters open

themselves to public scrutiny, thus insuring that the experiments are reproducible and unprejudiced.

On some rare occasions an experiment will verify a prediction that is *not* verified by other contemporary experiments. Yet, in the long run, this single result turns out to be correct. This is the case where the experimental techniques of the first researcher are much superior to any contemporaries'. Even though the positive claims are verified by future careful and more sophisticated experiments it is undeniable that the original positive results are criticized and even ridiculed by contemporaries. The pain associated with the creation or confirmation of scientific theories is almost always relegated to footnotes, if mentioned at all.[7]

The predictions made by the theory should apply to all phenomena that the theory claims to describe. If the majority of such predictions, but not all, are confirmed, the hypothesis should be modified in order to properly describe the discrepancies, or to explain why they do not fall within its scope. Theories cannot pick and choose the phenomena they aim to describe by selecting successes and rejecting failures.

Box 1.1

Continental drift

The idea of continental drift, proposed by the German geophysicist Alfred Wegener (1880–1930), was originally motivated by the observation that fossils of identical plants and animals are found on opposite sides of the Atlantic. The standard explanation at that time (1911) postulated that land bridges, now sunken, had once connected far-flung continents, but this could not explain certain geological features such as the close fit between the coastlines of Africa and South America, the match between the Appalachian mountains of eastern North America with the Scottish Highlands, and the fact that the distinctive rock strata of the Karroo system of South Africa were identical to those of the Santa Catarina system in Brazil. Using these arguments and a variety of fossil records, Wegener boldly proposed that all these observations could be explained by assuming that about 300 million years ago all the continents themselves had once been massed together in a single super-continent he called *Pangea* (from the Greek for "all the Earth") which has since been fractured into the current geography. Wegener was not the first to suggest this, but he was the first to present extensive evidence for it.

Though some scientists supported this hypothesis, the reaction from the geophysical community was almost uniformly hostile, and often exceptionally harsh and scathing. Part of the problem was that Wegener provided no convincing mechanism for how the continents might move: he envisioned the continents plowing through the Earth's crust driven by tidal and centrifugal forces, but his opponents noted that these forces are too weak for the task, and that the shape of

Box 1.1 (cont.)

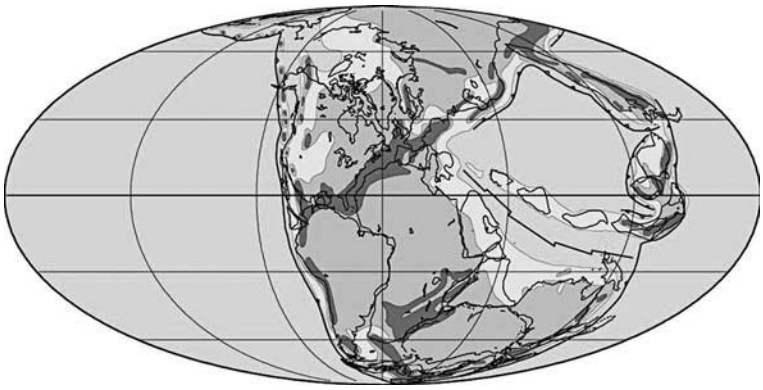

Figure 1.2 A modern reconstruction of the original super-continent Pangea. (courtesy of C. R. Scotese, PALEOMAP Project (www.scotese.com))

the continents would be distorted beyond recognition should they plow through the hard crust. In addition Wegener's original data contained errors that led him to make impossible claims (e.g. that North America and Europe were moving apart at over 250 cm per year – about 100 times faster than the actual number).

Despite this unpromising beginning, Wegener's basic idea has been fully vindicated by modern investigations. These experiments show that the Earth's crust is composed of a series of mobile plates, called *tectonic plates*, floating on a bed of molten rock that behaves as a viscous fluid (a result of the enormous pressures it is subjected to). The various continents lie atop the tectonic plates and so move about with them. When plates separate the underlying molten rock oozes out (as happens in the middle of the Atlantic Ocean); and when plates collide they generate regions of intense earthquake activity (as in the west coast of North America), and often give rise to stupendous mountain ridges, such as the Himalayas.

Trivial as the above may sound, such practices are sometimes adopted as policy in pseudo-scientific research. For example, the official policy of the *Journal of Parapsychology* is to reject publications containing negative results since these are considered as failures and hence of little use; the editors of this journal assume a priori the existence of parapsychological phenomena and deny they should be subject to testing.

Hypotheses and theories within the exact sciences (and in some cases within the social sciences) are often cast in mathematical language. The derivation of the predictions comes then from manipulation of the mathematical expressions within the theory as determined by logic. Such manipulations correctly describe Nature in tantalizing fashion (in E. Wigner's words, "the unreasonable effectiveness of mathematics") and there is no obvious reason why this should be so, but it is. One advantage is that the predictions and assumptions are very

precise; a disadvantage is that the mastering of the techniques requires time and effort.[8]

Ockham's Razor

A description of a phenomenon often requires a set of initial conditions in order to provide predictions. For example, in order to predict the motion of all bodies in our Solar System, we need to know the positions and velocities of all of them at one time (any one instant will do). With this information the future behavior of the planets can be predicted by solving Newton's (or Einstein's) equations. One might, of course, ask what circumstances dictated those initial positions and velocities, and even formulate a hypothesis for that. But one should not use the accuracy of Newton's law of gravitation to justify the second hypothesis. For example, suppose we claim that the planets have their present orbits because a race of advanced aliens put them at the appropriate places and gave them the required initial pushes. So far this is a valid hypothesis. But this does not warrant the claim that our alien hypothesis is *supported* by the great accuracy with which we can predict the planetary motion. We can predict planetary motion thanks to Newton's genius in devising his law of gravity, this has *nothing* to do with how the planets got where they are. The aliens (or lack thereof) are irrelevant in predicting the planets' motion.

Top-loading a theory with a new hypothesis and claiming the successes of the theory justify the hypothesis is common, but it is clearly wrong. This was repeatedly emphasized and justified by the medieval philosopher/theologian William of Ockham, who summarized this reasoning with the statement:

> *Pluralitas non est ponenda sine neccesitate*.
> (Entities should not be multiplied unnecessarily.)

The above is referred to as Ockham's "Razor," because it "cuts" between those components necessary for a theory, and those which do not add anything except complication. Ockham's Razor does *not* mean "always keep hypotheses simple," rather "eliminate the dead wood." Einstein's theory of gravitation is considerably more complicated than Newton's, yet it provides a much better description of Nature. Neither of them attests to the presence of aliens in our universe.

The validity of scientific theories

Scientific theories have, by their very nature, a limited applicability. As I mentioned above, even if we find a theory that is absolutely true in some metaphysical sense, we would be unable to demonstrate this admirable quality: we cannot possibly verify the infinity of the predictions such a theory would provide.

When investigating a new set of phenomena, researchers come armed with the available successful theories. Using these theories they try to describe the new