2 Scientific method

> *It is true that much time and effort is devoted to training and equipping the scientist's mind, but little attention is paid to the techniques of making the best use of it.*
>
> Beveridge, 1957, p. iv

In order to attempt to resolve the obvious disagreement among scientists concerning the scientific method, some understanding of what is implied by the words is needed, and therefore, definitions of the words *science* and *method* are required. The word science is derived from the Latin words scientia, knowledge, and scientificus, making knowledge (Little *et al.*, 1964, p. 1806). Hence, science involves the business of discovery and the production of new knowledge. It is an activity that supposedly creates objective knowledge. That is, science is the attempt to learn the truth about those parts of nature that are explorable (Chargaff, 1978, p. 156).

The word method is defined in *The Oxford dictionary* (Little *et al.*, 1964, p. 1243) as 'a special form of procedure adopted in any branch of mental activity, whether for exposition or for investigation'. Hence, method is a way of doing anything according to a regular plan. So, if there is a method there is a procedure or a systematic way of pursuing a goal. Interestingly enough, the scientific method is not defined in this very comprehensive dictionary, but in *The American heritage dictionary* (Morris, 1981, p. 1163) it is defined as 'the totality of principles and processes regarded as characteristic of or necessary for scientific investigation, generally taken to include rules for concept formation, conduct of observations and experiments and validation of hypothesis by observation and experiments'. The three major components of this definition are, 1. concept formation, which involves generation of hypotheses, 2. observation and experiments, which involve the procedures of data collection, and 3. testing of hypotheses by observation and experimentation. This short statement adequately describes a scientific method. Data collection and testing are critical components of this definition, but because of the diverse nature of 'science', it is not possible to state a single method that applies to all sciences.

METHODS

Beveridge (1957) says that research is a complex and subtle task. It is difficult to teach and, therefore, one should just go do it. In effect, one should learn by one's mistakes in carrying out research, but nevertheless, he wrote his book *The art of scientific investigation* to assist the young scientist. The title of his book implies that there is a great deal of subjectivity in the activities of science, and indeed, if we take

the definition of method to mean a formal procedure or a specific way of proceeding then this subjectivity invalidates the definition of scientific method. Techniques and procedures are going to differ for the different sciences, and when one considers the great range of scientific endeavors it is no wonder that there is disagreement about whether or not there is a scientific method.

Nevertheless, we know that certain methods are not appropriate, such as the methods of Lysenko: 'distortion of facts, demagoguery, intimidation, dismissal, reliance on authorities, eyewash, misinformation, self-advertising, repression, obscurantism, slander, fabricated accusation, insulting name calling, and physical elimination of opponents . . . ' (Medvedev, 1969, pp. 191–2). In most scientific circles such methods would not be effective, but with Stalin's support they were effective in the USSR. Such techniques will not personally enhance a reputation, and they are not recommended for the pursuit of science.

Beveridge (1980, p. 55) does outline a scientific method in five steps, but he also includes some problems that are associated with each as follows:

1. recognition and formulation of the problem, but the statement of the problem may be incorrect,
2. collection of relevant data, but in many cases it is difficult to know what data are relevant,
3. induction of a hypothesis, but induction is unreliable,
4. deduction and testing of the hypothesis, but there can be error during testing,
5. revision of the hypothesis to give a more comprehensive explanation of the results, or starting again at step 3 by stating additional hypotheses, but the results may be misleading.

Earth scientists should disagree with Beveridge's outline, which involves development and testing of a single hypothesis, because the 'method of multiple working hypotheses', as stated by Chamberlin (1890, 1897) involves the formulation of as many hypotheses as possible, which are then selectively eliminated or combined to develop an explanation of the phenomenon under consideration. When only one hypothesis is generated and an attempt is made to demonstrate its correctness, it becomes a 'ruling hypothesis', which dominates the thinking of an investigator and may lead to serious error. The preferred method then is to develop as many explanations of a phenomenon as possible. Through the process of data collection these hypotheses are either modified or eliminated until a solution is developed, or perhaps until multiple explanations or hypotheses are combined to obtain a composite solution or theory.

Figure 2.1 is my attempt to organize a statement of the scientific method that generally follows Beveridge's approach (see Harvey, 1969, p. 34, Haines-Young and Petch, 1986, p. 25). This is a six-step procedure, but unlike Beveridge's the first step is preparation, which involves training and literature review. Without this background it is not possible to recognize a significant problem. In step 3 the problem statement may come directly from the literature, where it has been formulated by another, or it can come from observation of natural phenomena (2a) or by recognition of an anomaly or trend in data collected for another purpose (2b). For example, when a prepared earth scientist views the landscape on aerial photographs or from an aircraft, a great range of landform types, and vegetational and erosional patterns will be observed, and hypotheses can be developed to explain this variability. Also, a vast amount of data has been collected by the Water Resources Division of the Geological Survey primarily for water management purposes, but anyone examining these data will discover water and sediment discharge variations and anomalies that are of scientific and practical importance. Again, hypotheses can be formulated in an attempt to explain these deviations.

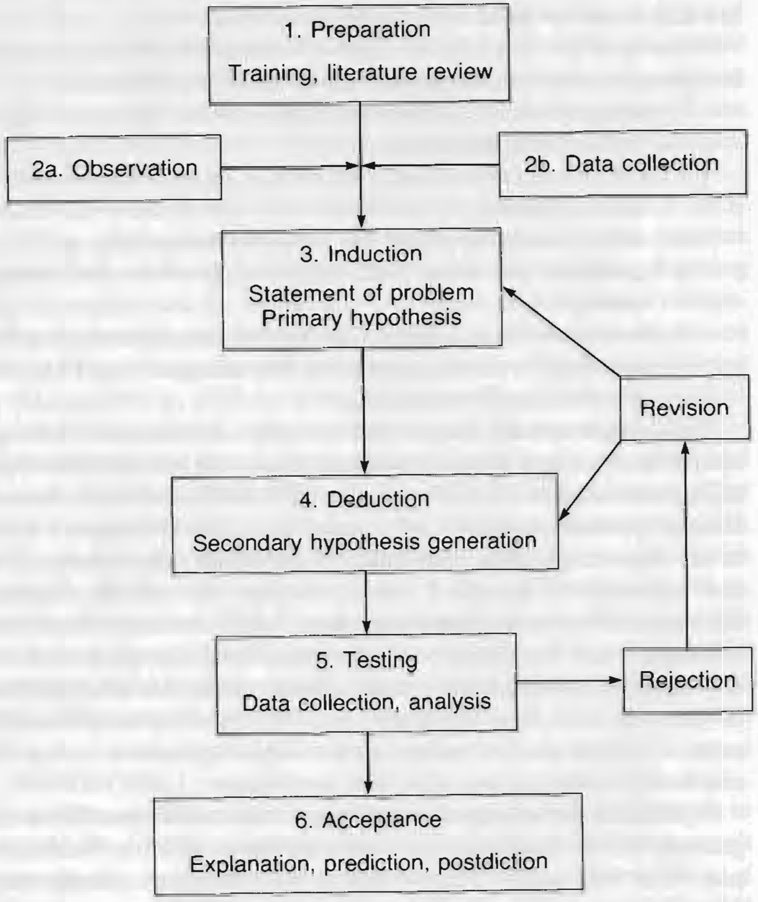

Figure 2.1. An outline of a scientific method.

From training (1) and experience (2) a problem is recognized by induction (Fig. 2.1), which can be considered to be a primary hypothesis (3). With the statement of the primary hypothesis deductions are made (4), which can be referred to as secondary hypotheses. These can be tested, and this testing will lead to either acceptance or falsification of the primary hypothesis.

Step 5 is the real work of testing hypotheses by field work and data collection, experimentation, etc. Falsification of a hypothesis may lead to its total rejection and failure or, more likely, revision of the hypothesis and renewed testing. Acceptance of the hypothesis will lead to explanation of the problem stated in step 3, and to acceptance because there is a high probability that the hypothesis is correct (6).

Figure 2.2 is an attempt to illustrate steps 4, 5 and 6 of Fig. 2.1. Given a problem (P), hypotheses (H) are generated in order to find an explanation or solution (ES) that is based upon a most probable hypothesis. When only a single ruling hypothesis is developed (*a*), the probability of being wrong is increased, and, of course, the researcher's reputation is at risk if it proves to be wrong. The creation of multiple hypotheses avoids this probability, either in a multiple sequential mode (*b*), with one hypothesis following another as weaknesses are found in each, or in a multiple parallel mode (*c*) with a number of hypotheses being developed and tested simultaneously. Finally, there is the combining of multiple explanations or

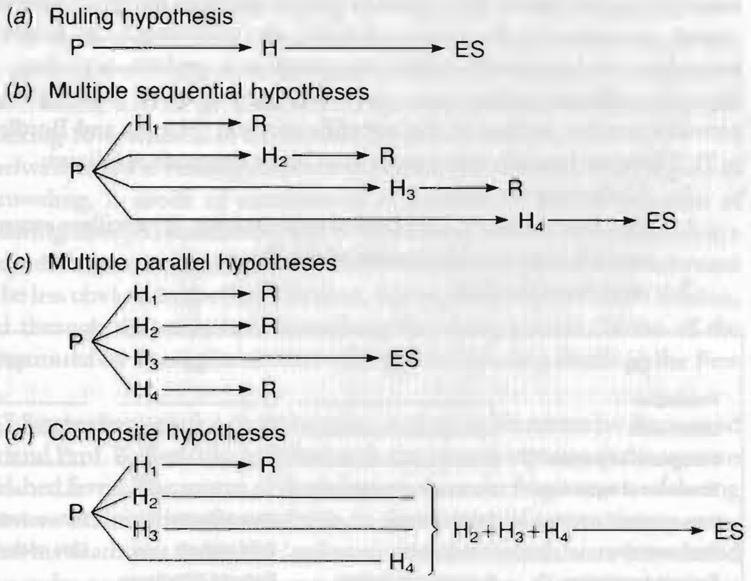

Figure 2.2. The paths to explanation (P = problem; H_1–H_x = hypotheses; H = ruling hypothesis; ES = explanation and solution; R = rejection).

hypotheses to develop a composite hypothesis (*d*). In fact, most explanations of complex phenomena will involve a combination of hypotheses, and many great controversies in science end in a compromise.

Part of the difficulty in outlining a scientific method is that there are at least four different levels of scientific activity (Fig. 2.3). Many investigators will enter Fig. 2.1 at step 3 with the problem stated by a supervisor (Fig. 2.3(*b*)). This is a lower level of scientific effort than the recognition of the problem as a result of experience and observation (Fig. 2.3(*a*)). An even lower level of research occurs when the investigator is presented with the hypothesis to be tested (FIg. 2.3(*c*)). The lowest level of effort is when the investigator is instructed to collect and analyze certain types of data. The results are then given to the principal investigator for evaluation. Depending upon what stages of the method one experiences (Fig. 2.1), one's perspective of the method will be affected. That is, someone entering at lower levels may never realize the importance and difficulty of problem recognition and hypothesis generation.

Differential diagnosis

Closely related to the method of multiple working hypotheses (Fig. 2.2) is the process of differential diagnosis in medicine. Harvey and Bordley (1970) deal with this topic, in a book that consists of 1238 pages of information on the diagnosis of various diseases. These authors define differential diagnosis as the *art* of distinguishing one disease from another and of selecting the disease which comes closest to explaining the clinical and laboratory findings. The physician must formulate his hypotheses, when he examines a patient, and these working hypotheses direct the diagnostic process. An outline of the steps in this diagnosis provides another outline of the scientific method (Harvey and Bordley, 1970, p. 7). There are basically three main steps in the diagnosis as follows:

1. Collecting the facts:
   (a) clinical history, (b) physical examination, (c) ancillary examinations, and (d) history of the course of the illness.
2. analysis of the facts:
   (a) critically evaluate the collected data, (b) list reliable findings in order of apparent importance, (c) select one or preferably two or three central features of the illness (d) list diseases in which the central features are encountered.
3. Reach a final diagnosis by selecting from the list of diseases either:
   (a) disease which best explains all the facts or (b) several diseases, each of which best explains the facts.
4. Review all the evidence both seemingly positive and seemingly negative with the final diagnosis in mind.

(*a*)	(*b*)	(*c*)	(*d*)
Preparation			
Observation			
Recognition of problem			
Statement of hypotheses	Statement of hypotheses		
Data collection	Data collection	Data collection	Data collection
Data analysis	Data analysis	Data analysis	Data analysis
Evaluate hypotheses	Evaluate hypotheses	Evaluate hypotheses	

Figure 2.3. Outline of four different levels of research.

The physician must keep an open mind and be aware of the many diseases that may have symptoms similar to those of the patient. To give some idea of the problems with and complexities of this procedure, Table 2.1 is a list of the causes of diarrhea, something that all readers should be familiar with. We see that the physician can have an abundance of hypotheses for the cause of diarrhea in a particular instance, and, of course, the causes can be multiple (Fig. 2.2). Although there are seven classes of acute diarrhea and six classes of chronic diarrhea with a total of 86 specific cases, the physician will probably give his patient one reason and prescribe for the problem.

A fine example of the use of multiple hypotheses can be obtained from the life work of Ignaz Semmelweis (1861) who used this method several decades before it was described by Gilbert and Chamberlin. Semmelweis was for a period in charge of obstetrics at the Vienna Lying-in Hospital, which was divided into two divisions. All medical students worked with the patients in the First Division and student midwives were assigned to the Second Division. In general treatment of patients was similar in both divisions, but in the First Division the mortality rate was a shocking 10%, whereas in the Second Division it was 3%.

Semmelweis tested several hypotheses to explain this disparity with regard to 1. overcrowding, 2. mode of examination of patients, 3. diet, 4. position of women during labor, 5. ventilation, and 6. cleanliness, and he eliminated all as a cause of the difference in the death rates. He even made the priest, who delivered last rites, be less obvious in the First Division, where, unlike the Second Division, he passed through the ward before reaching the dying patient. None of the changes instituted by Semmelweis were effective in reducing deaths in the First Division.

In 1847 Semmelweis took a short vacation, and upon his return he discovered that his friend Prof. Kolletschka had died with the same symptoms as the women with child-bed fever. The source of the infection was a cut finger received during a post-mortem examination. Semmelweis, in the absence of a germ theory, concluded that his death was caused by 'cadaveric particles, which were introduced into his vascular system'. The next step in his analysis was the assumption that medical students who went directly from autopsies to the hospital introduced the cadaveric particles to the pregnant women. He examined hospital records and

Table 2.1. *Causes of diarrhea (from Harvey and Bordley, 1970, p. 445)*

- I. Acute
  - A. Intrinsic disease of the gastrointestinal tract
    - 1. Appendicitis
    - 2. Diverticulitis
    - 3. Ischemia of the bowel
    - 4. Radiation enteritis
    - 5. Acute ulcerative colitis
    - 6. Pseudomembranous enterocolitis
    - 7. Partial intestinal obstruction (e.g. intussusception)
    - 8. Acute exacerbation of chronic diarrhea
  - B. Infections of the intestinal tract
    - 1. Viral enteritis
    - 2. Salmonella enteritis
    - 3. Shigella enteritis
    - 4. Amebic colitis
    - 5. Cholera
    - 6. Staphylococcal enterocolitis
    - 7. Giardiasis
    - 8. Clostridium welchii
  - C. Toxins and poisons
    - 1. Staphylococcus
    - 2. Clostridium botulinum
    - 3. Clostridium welchii
    - 4. Heavy metals
    - 5. Mushrooms
    - 6. Carbon tetrachloride
  - D. Drugs
    - 1. Cholinergic drugs
    - 2. Ganglionic blocking agents
    - 3. Antibiotics
    - 4. Colchicine
    - 5. Digitalis
    - 6. Iron
    - 7. Para-amino salicylic acid
    - 8. Antimalarials
  - E. Generalized disorders or diseases affecting the intestine
    - 1. Trichinosis
    - 2. Malaria
    - 3. Uremia
    - 4. Cholemia
    - 5. Pernicious anemia
    - 6. Adrenal insufficiency
    - 7. Hyperthyroidism
    - 8. Diabetes mellitus
    - 9. Psittacosis
    - 10. Weil's disease
    - 11. Carcinoid syndrome
  - F. Acute exacerbation of chronic diarrhea
  - G. Psychogenic causes
- II. Chronic
  - A. Intrinsic disease of the gastrointestinal tract
    - 1. Ulcerative colitis
    - 2. Diverticulitis
    - 3. Irritable colon
    - 4. Colonic neoplasms
    - 5. Partial intestinal obstruction
    - 6. Inadvertent gastro-ileostomy
    - 7. Regional enteritis
    - 8. Carcinoma of the stomach
    - 9. Familial polyposis of the colon
  - B. Infection of the gastrointestinal tract
    - 1. Amebic colitis
    - 2. Actinomycosis
    - 3. Tuberculosis

Table 2.1 (*cont.*)

-
  -
    - 4. Giardiasis
    - 5. Strongyloidiasis
  - C. Drugs and poisons
    - 1. Cathartics
    - 2. Thyroid
    - 3. Digitalis
    - 4. Iron
    - 5. Mercury
  - D. Generalized disorders or diseases affecting the intestine
    - 1. Endocrine and metabolic
      - a. Uremic colitis
      - b. Hyperthyroidism
      - c. Addison's disease
      - d. Hypervitaminosis D
      - e. Diabetes mellitus
      - f. Cushing's syndrome
      - g. Amyloidosis
      - h. Hypoparthyroidism
    - 2. Connnective tissue diseases
      - a. Systemis Iupus erythematosus
      - b. Progressive systemic sclerosis
      - c. Polyarteritis'
    - 3. Neoplastic disease
      - a. Carcinoid syndrome
      - b. Sympathomimetic amine-secreting tumors
      - c. Zollinger-Ellison syndrome
      - d. Carcinoma of the pancreas
      - e. Carcinoma of the liver
      - f. Chronic myeloid leukemia
      - g. Pancreatic adenoma without gastric hypersecretion
      - h. Nodular carcinoma of thyroid with metastases
    - 4. Deficiency diseases
      - a. Pellagra
      - b. Pernicious anemia
    - 5. Other
      - a. Cirrhosis of the liver
      - b. Chronic cholecystitis
      - c. Chronic pancreatitis
      - d. Lymphogranuloma venereum
      - e. Protein-losing enteropathy
  - E. Malabsorption symdrome due to various causes
  - F. Psychogenic

determined that the hospital opened in 1784 and from that time until 1841 the mortality rate was about 1 per cent, but after 1841, when autopsies were begun, the mortality rate increased to 8 per cent in 1841, 15.8 per cent in 1842, and 9 per cent in 1843.

The explanation for the disparity in death rates in the two divisions of the hospital was because medical students performed autopsies and went directly from the autopsy to the patients in the First Division. They did not examine patients in the Second Division. The student midwives did not attend autopsies, and therefore, they did not infect their patients in the Second Division.

Semmelweis forced the physicians and students to wash their hands in a solution of chlorine, and the mortality rates became comparable in the two divisions. Semmelweis' reasoning was correct, but unfortunately he could not convince the medical establishment, and his method was not generally followed, which was a great tragedy. Nevertheless, Semmelweis provides us with an excellent example of multiple working hypotheses, and the generation of a new hypothesis as new information becomes available, which appears to be a combination of methods (*b*) and (*c*) of Fig. 2.2.

Method in earth sciences

In order to function under the conditions imposed by their science, earth scientists have used the 'principle of uniformitarianism' (uniformity), and reasoning by analogy. Geikie (1905, p. 299) described uniformity as 'the present is the key to the past'. Europeans prefer the term actualism (actualisme, Aktulismus), which refers to explanation, postdiction and prediction based on the understanding of present processes (Jong, 1966; Rutten, 1971). There is currently considerable controversy concerning uniformitarianism (Albritton 1963, 1967; Gould, 1965, 1984), which has either been lauded as a basic principle of geology or as a worthless concept. For example, Challinor (1968) is of the opinion that the present tells us very little about the past and that, in fact, the past tells us about the present (Kitts, 1966, p. 143). The concept of uniformitarianism has also been greatly misinterpreted, and Shea (1982) concludes that 'most geologists do not understand the nature and correct meaning of what is said to be the basic principle of their science'.

The controversy surrounding the definition of uniformity is somewhat analogous to the reinterpretation of the US Constitution in the light of modern conditions. The fact that neither is defined rigorously leads to conflicting opinions. In order to remove this difficulty the restricted definition of uniformity as used by physicists, chemists and philosophers, can be employed. It is simply the assumption that natural laws are permanent; that is, under the same conditions a given cause will always produce the same results. Therefore, 'Regularities in the sequence of type of event found in any restricted region of space and time will continue to hold good in all regions of space and time.' (Harré, 1960, p. 117). This concept is necessary for extrapolation, from the present to the past and to the future.

Nothing stated above excludes the rare event such as meteor impact (Kyte *et al.*, 1988) and massive flooding as in the Channeled Scablands (Baker, 1973). They are improbable events that can be explained rationally. Such events probably can be safely eliminated from short-term predictions, but not for geologic time (Gretener, 1967, 1984), and they do not violate uniformity.

Analogy

Explanation of past events and prediction of future events (application of uniformity to the past and future) frequently requires reasoning by analogy (Gilbert, 1896; Leopold and Langbein, 1962), which is the recognition of similarity among different things. Analogy is indeed the basis of explanation and extrapolation for geologists, which is why we need experience and need to see as much as we can of different existing (actual) conditions so that comparisons can be made among a variety of geologic phenomena (Jong, 1966, p. 238). For example, planetary geologists use analogic arguments to explain features that cannot be inspected 'in the field' (e.g. Schumm, 1970; Baker and Milton, 1974).

Analogy was recognized as a significant component of scientific discovery by Francis Bacon, Booke, Kepler, Mach, Maxwell, Poincaré, Leatherdale (1974, pp. 4 and 13), and Medawar (1984, p. 109). Analogy in science is based on the assumption that if two things are alike in some respects then they must be alike in others. Frequently the similarity among things leads to a flash of insight and an inductive leap that solves a problem. Needless to say, many similarities may be misleading, and the recognition of analogous conditions is not sufficient to establish more than a working hypothesis. Indeed, Von Bertalanffy (1952, p. 200) insists that analogies are worthless but that homologies are useful because they involve more than superficial similarities, and they involve similarities in both structure and function. Extrapolation will be strengthened by use of homology because analogy is like a metaphor, whereas homology is like a synonym. Homology is a higher level of comparison than analogy. The discussion of the use of analogy and metaphor in science is extensive (e.g. Livingstone and Harrison, 1981).

A problem with using analogy in geomorphology is that, where erosion is moderate or slow, the modern landscape may be dominated by past events which greatly complicate the analogy (Douglas, 1980). Areas subjected to continental glaciation retain that influence, and on a different scale the patterns of some modern rivers reflect the influences of active tectonics on valley-floor slope (Adams, 1980; Burnett and Schumm, 1983). The question, then, becomes one of the validity of using modern analogy as a basis for extrapolation. There really is no other option, and the actualistic approach is mandatory. In fact, the concept of actualism can be modified and expanded to include the future as follows: the present provides insights into the past, and it influences future geologic processes and events, and the past influences both present and future geologic processes and events (Fig. 1.2).

Multiple hypotheses

Although, unlike other sciences, there are no books describing the scientific method as applied to geologic problems, there are several papers that stress the multiple hypothesis approach (Fig. 2.2), and in each case the procedure is illustrated by geomorphic examples (Gilbert, 1886, 1896; Chamberlin, 1890, 1987). Because these papers are of such importance they will be reviewed here.

Gilbert:

Gilbert (1886) characterizes his method of research as follows:

1. selective and concentrated observation;
2. empirical classification and grouping of facts;
3. development of a hypothesis by induction to explain observations;
4. testing the validity of hypothesis and revising until explanation is satisfactory.

An important contribution of Gilbert, which anticipates Chamberlin's 1890 paper on multiple working hypotheses, is his statement that the investigation 'is not restricted to the employment of one hypothesis at a time. There is indeed an advantage to entertaining several at once . . . ' (Gilbert, 1886, p. 286). He warns of the dangers of a single ruling hypothesis and stresses that the development of hypotheses depends on analogy.

The example that Gilbert uses to illustrate his method is the variable height of a Pleistocene shoreline of Lake Bonneville, Utah. A previous worker had determined the elevation of the shoreline at one location and assumed that it was horizontal. Gilbert measured the elevation at two locations and found that the elevations were not the same, and, therefore, the shoreline was not horizontal. He first assumed that 'a gentle undulatory movement of the crust' or folding was the cause, but he then observed a fault between the location of the measurements and this additional hypothesis, faulting, was added as a probable explanation. Gilbert carried out additional surveys, as part of his general study of the geology of Lake Bonneville, and he found that the displacement along the fault was not adequate to explain the difference in elevation of the shoreline, therefore, the faulting hypothesis was rejected.

As more information became available, he discovered that the maximum displacement of the shoreline occurred over the center of Lake Bonneville, and he proposed the deformation was somehow related to the drying up of Lake Bonneville to form Great Salt Lake. He eventually suggested that the warping was the result of what we now call isostatic adjustment of the crust to the removal of the load produced by evaporation of the water, but in order to prove this hypothesis additional surveying of the complete shoreline and surveys of other Pleistocene lakes were required. If all showed similar shoreline deformation the isostatic hypothesis would be strongly supported, and of course, it was (Crittenden, 1963), and the folding and faulting hypothesis was abandoned. Gilbert apparently used the method of multiple sequential hypotheses (Fig. 2.2(*b*)).

Chamberlin:

Shortly after Gilbert's paper appeared the famous paper of Chamberlin (1890) was published. This paper has received much attention during the last 100 years (Laudan, 1980), and Platt (1964, p. 350) describes the method of multiple working hypothesis as a 'great intellectual invention'. In any case, the paper was published twice by Chamberlin and reprinted at least three times.

Chamberlin's objective was to stress the need for multiple hypothesis generation, a procedure used by Gilbert (1886) and Semmelweis (1861) earlier. Chamberlin stressed the danger of a ruling hypothesis (Fig. 2.2(*a*)) as follows:

> The moment one has offered an original explanation for a phenomenon which seems satisfactory, that moment affection for his intellectual child springs into existence; and as the explanation grows into a definite theory, his parental affections cluster about his intellectual offspring, and it grows more and more dear to him, so that, while he holds it seemingly tentative, it is still lovingly tentative, and not impartially tentative. So soon as this parental affection takes possession of the mind, there is a rapid passage to the adoption of the theory. There is an unconscious selection and magnifying of the phenomena that fall into harmony with the theory and support it, and an unconscious neglect of those that fail of coincidence. The mind lingers with pleasure upon the facts that fall happily into the embrace of the theory, and feels a natural coldness toward those that seem refractory. Instinctively there is a special searching-out of phenomena that support it, for the mind is led by its desires. There springs up, also, an unconscious pressing of the theory to make it fit the facts, and a pressing of facts to make them fit the theory. When these biasing tendencies set in, the mind rapidly degenerates into the partiality of paternalism. The search for facts, the observation of phenomena and their interpretation, are all dominated by affection for the favored theory until it appears to its author or its advocate to have been overwhelmingly established. The theory then rapidly rises to the ruling position, and investigation, observation, and interpretation are controlled and directed by it. From an unduly favored child, it readily becomes master, and leads its author whithersoever it will. The subsequent history of that mind in respect of that theme is but the progressive dominance of a ruling idea.

Even when the investigator tries to be unbiased and to test a single hypothesis objectively, it is easy for it to become a ruling hypothesis. The solution is clear; multiple hypotheses must be proposed (Fig. 2.2(*c*)) and

> The investigator thus becomes the parent of a family of hypotheses: and, by his parental relation to all, he is forbidden to fasten his affections unduly upon any one. Having thus neutralized the partialities of

> his emotional nature, he proceeds with a certain natural and enforced erectness of mental attitude to the investigation, knowing well that some of his intellectual children will die before maturity, yet feeling that several of them may survive the results of final investigation since it is often the outcome of inquiry that several causes are found to be involved instead of a single one. In following a single hypothesis, the mind is presumably led to a single explanatory conception. But an adequate explanation often involves the coordination of several agencies, which enter into the combined result in varying proportions. The true explanation is therefore necessarily complex. Such complex explanations of phenomena are specially encouraged by the method of multiple hypotheses, and constitute one of its chief merits. We are so prone to attribute a phenomenon to a single cause, that, when we find an agency present, we are liable to rest satisfied therewith, and fail to recognize that it is but one factor, and perchance a minor factor, in the accomplishment of the total result.

Not only is the method of multiple hypotheses a significant contribution, but the resulting statement by Chamberlin that complex or composite explanations of a feature are probable is equally significant (Fig. 2.2(*d*)), but this point has been ignored in the discussions of Chamberlin's contributions. Chamberlin uses as an illustration the deep basins of the Great Lakes, which had been previously explained as river valleys, glacial excavations, or as the result of downwarping. Chamberlin states that all of these explanations are partly correct.

> The problem therefore, is the determination not only of the participation, but of the measure and the extent, of each of these agencies in the production of the complex result. This is not likely to be accomplished by one whose working hypothesis is pre-glacial erosion, or glacial erosion, or crust deformation, but by one whose staff of working hypotheses embraces all of these and any other agency which can be rationally conceived to have taken part in the phenomena.

A final quotation from the 1897 (p. 165) paper provides an adequate summary of Chamberlin's views.

> The studies of the geologist are peculiarly complex. It is rare that his problem is a simple unitary phenomenon explicable by a single simple cause. Even when it happens to be so in a given instance, or at a given stage of work, the subject is quite sure, if pursued broadly, to grade into some complication or undergo some transition. He must therefore ever be on the alert for mutations and for the insidious entrance of new factors. If, therefore, there are any advantages in any field in being armed with a full panoply of working hypotheses and in habitually employing them, it is doubtless the field of the geologist.

Gilbert:

Gilbert's (1896) later paper is his attempt to illustrate in some detail the method of multiple hypotheses, which, ironically, in this case led to an incorrect conclusion. Gilbert again stresses the role of analogic reasoning in hypothesis generation.

The example he selects to illustrate the approach is his investigation of a large crater southeast of Flagstaff, Arizona called Coon Butte and now known as Meteor Crater. This large approximately circular crater is 'a few thousand feet broad and a few hundred feet deep' with a rim rising above the plain of Kaibab limestone. Although the San Francisco volcanic field is nearby, the crater is comprised of limestone and sandstone. However, meteor fragments are found nearby. Four hypotheses for the crater and iron fragments were proposed by Gilbert and others as follows:

1. it is a volcanic explosion crater from which the fragments were ejected;
2. it is a breached dome over a laccolithic intrusion at depth;
3. it is a crater formed by steam explosion;
4. it is a meteor crater.

A magnetic survey by Gilbert showed that a meteor was not buried in the crater, and a calculation of the volume of ejected material revealed that the ejected material would fill the crater leaving no space for a meteor (Fig. 2.4(*a*)). As there are no igneous rocks associated with the crater, all hypotheses are rejected except that of a phreatic explosion. Elsewhere in the world similar appearing craters (Maars) were formed by explosions, and a nearby volcanic field indicated that a source of geothermal energy existed (Fig. 2.4(*b*)); therefore, Gilbert accepted the volcanic origin of Meteor Crater. If he had known that the impact of the meteor at high velocity causes an explosion (Cooper, 1977) and that a relatively small meteor could have formed Meteor Crater, his conclusion undoubtedly would have been different.

The example of Gilbert's study of Meteor Crater shows that in spite of a determined effort to use multiple hypotheses erroneous conclusions can result. This is especially true when the records are fragmentary, as in most studies of structural geology, stratigraphy, sedimentation and historical geomorphology.

Douglas Johnson (1933, 1940) elaborated on Gilbert's and Chamberlin's suggestions and described a seven-step method of scientific analysis that differs little from others discussed above. More recently Platt (1964) urges the systematic application of multiple hypotheses with careful design of crucial experiments that will exclude some hypotheses. Mosley and Zimpfer (1976) stress, in addition, that composite explanations are often required when dealing with complex geomorphic systems. They use an analogy with analysis of variance and conclude that an explanation of a complex phenomenon, such as river meandering, requires partial answers from geology, geomorphology, hydraulics, geotechnical engineering, and hydrology (Fig. 2.2(*d*)).

If a ruling hypothesis proves correct then all credit to its parent. However, if it is in error the consequences to a career may be unfortunate and significant. The multiple hypothesis approach, although difficult, is an ideal that should be the goal of the earth scientist.

Analogy with scientific paper

It may be of some value to consider that a well written scientific paper is analogous to the scientific method. Even though the scientific paper does not describe the mistakes made and the false starts during the performance of research, it provides an idealized example of how science should be carried out (Table 2.2). For example, the introduction of the scientific paper includes a statement of the problem, a review of relevant literature, and a statement of the objectives of the specific study. In research this is the period of preparation, problem identification, and hypothesis generation (Fig. 2.1). The second part of a paper is a description of the procedure followed, which in research involves development of a work plan to be followed for data collection and analysis. The third part of the paper is the presentation of the results of data collection and analysis, which is obviously the next step in the research program. In the paper the final conclusions are presented, but in reality only if we are very lucky is the original hypothesis supported, and it may be necessary to modify the hypothesis and to generate other hypotheses which will then be tested.

Figure 2.4(*a*). G. K. Gilbert's experiments on crater formation by impact. Balls of clay were thrown against a slab of clay. Different size craters are due to different impact velocities. Note that the clay ball occupies a large part of the crater volume. (G. K. Gilbert photograph 842, circa 1891, US Geological Survey Photolibrary, Denver.)

Unfortunately, as noted above, the scientific literature gives little hint of the trials of scientific research, and the papers that are produced only describe successful methods, successful data analysis and conclusions which support the original or modified hypotheses. Of course, this is because of page limitations and the inability of editors to permit a discussion of the full amount of work done by a scientist, which would include the normal confusion and disappointments inherent in research. Therefore, the despair of success frequently experienced by a young researcher may seem to him unique, whereas it is a common attribute of research. The senior scientist in such cases must advise the young investigator to proceed with his work because the hypothesis is worthy of investigation and because the end results will be of value, even in the disappointing circumstance that they do not support the stated hypotheses. Unfortunately, the scientific literature gives a very misleading perspective of science. It isn't as easy as the papers suggest.

Fig. 2.4(*b*). G. K. Gilbert contemplating cinder cones near Flagstaff, Arizona. (G. K. Gilbert photograph 801, 1891, US Geological Survey Photolibrary, Denver.)

Table 2.2. *Comparison of the scientific method to the outline of a scientific paper*

Scientific paper	Scientific method
1 Introduction	Preparation
Statement of problem	Statement of problem
Literature review	Literature review
Statement of objectives	Statement of objectives
2 Procedure	Work plan
Date collection	Data collection
3 Results	Data analysis
Data analysis	Evaluation of results
Results of analysis	
4 Summary and conclusions	Accept, reject or modify hypothesis

CRITICISM OF THE SCIENTIFIC METHOD

Although there is agreement that a general scientific method or approach exists, there are critics who convincingly argue that the method is only as powerful as the objectivity of the individual using it. Although the multiple hypothesis approach is obviously the way to proceed, the difficulties of eliminating bias are nicely stated by N. L. Bowen (1948, p. 79) as the 'method of multiple prejudices'. This is in contrast to multiple working hypotheses where

> Chamberlin thus had in mind a happy situation where an individual investigator diligently sought all reasonable processes that might lead to an observed relation, carefully considered the full consequences of each process envisioned, impartially compared these deduced consequences with the re-examined facts and thus reached a conclusion as to the probable process or group of processes that were operative. 'Tis a consummation devoutly to be wished; yet one . . . cannot fail to wonder whether any individual is capable of such detachment. Each of us is, of course, quite sure that he himself has done just what Chamberlin recommends, but he is equally sure that the other fellow has done no such thing . . . That fictional character, the impartial observer, would probably say that we are all nursing pet prejudices.

Perhaps the severest criticism of the concept of a scientific method is presented by Kuhn (1970) and Feyerabend (1975, 1978). They make a valid point that, depending upon a person's social status, economic background, and training, and the prevailing social and scientific climate, it is possible to view scientific problems in different ways (see Haines-Young and Petch, 1986). For example, if one's geomorphic training was in an area of a humid climate and low relief, a low-energy system, one would conclude that very little has happened since the Pleistocene, and for a long time, British geomorphologists did not involve themselves in studies of erosion process and rates of change because the senior people were conditioned to believe that very little had happened since the last glaciation in the British Isles. During this time American geomorphologists were carrying out numerous studies of hillslopes and rivers, which provided information on rates of change and the response of the landscape to climatic change.

Oliver Sacks (1987, pp. 92–101) provides a neurological example. Tourette's syndrome is characterized by an excess of nervous energy and by tics, jerks, grimaces, noises, etc. Tourette described this condition in 1885, but in this century the syndrome seemed to have disappeared, and it was rarely diagnosed in the first half of this century. According to Sacks, some physicians regarded it as a 'product of Tourette's colorful imagination', but Sacks diagnosed a case in 1971. The next day 'without especially looking' he saw two more in the streets of New York. He was prepared to see rather than to ignore.

We tend to see what we are trained to see. This was brought home to the author, a geomorphologist, in a very forceful way while travelling in a car with a soil scientist, a structural geologist, and a botanist. All four of us were viewing the same landscape. A discussion with regard to a particular location revealed that the botanist was observing the vegetation, but he did not look through the vegetation to see the variations of soil, which were obvious to the soil scientist. I was looking at erosion features on a hillslope, when the structural geologist pointed out that the hill was an anticline. I had looked through the vegetation and the soil to the erosion features but ignored the geology. The structural geologist on the other hand had ignored the vegetation, soil and geomorphology to observe the very obvious structural folding that the other three had ignored. An individual cannot be criticised for the bias of his profession, but one should be aware of it.

A fictitious example of interpretation gone awry as a result of bias is provided by David McCaulay (1979) in his amusing book, *Motel of the mysteries*. In the year 4000 AD a twentieth-century motel is excavated by an archaeologist who was clearly conditioned by his study of early Egyptian civilization. His religious interpretation of the objects found in an average motel room indicate how badly things can go wrong if one does not maintain objectivity. For example, his interpretation of a flush toilet was that it was a sacred urn. During a burial ceremony 'The ranking celebrant, kneeling before the urn, would chant into it while water from the sacred spring flowed in to mix with sheets of Sacred Parchment.' Everything in the bedroom was oriented toward the television, therefore it was the high altar. The purpose of the multiple hypotheses approach is to prevent this type of error.

A bias of training becomes even more obvious when the same situation is viewed by scientists and engineers. For example, there can be three approaches to incised channel (gully, arroyo, entrenched stream) stabilization. The geomorphic approach can be to 'let it go,' as it will eventually reach a new condition of relative stability. The engineering approach will normally be to use a variety of structures to control the problem. There can also be an intermediate 'rational' approach that incorporates geology, geomorphology, and engineering. This approach uses engineering structures only at sites selected by a careful study that has identified the different stages of incised-channel change and that can provide the basis for selecting where engineering structures will be effective and where they are redundant or useless (Schumm *et al.*, 1984). This is similar to Mackin's (1963) rational approach to scientific problems where understanding is stressed rather than facile acceptance of purely statistical results (Klemés, 1985).

The problem is, of course, that the same field situation or 'the same set of experimental data can often be interpreted in more than one way – which is why the history of science echoes with as many venomous controversies as the history of literary criticism'. 'The data may be "hard" . . . but what you read into them is another matter' (Koestler, 1978, p. 153).

Resistance to change

There is also an apparent reluctance to accept new ideas in science, which suggests that scientists may be less objective than generally assumed (Barber, 1961) and in fact some unpublished notes of T. H. Huxley reflect this fact (Bibby, 1960, p. 77; Gregory, 1985, p. 10; Stoddart, 1986, p. 16). In the following I have substituted the word theory for Huxley's word novelty:

1. Just after publication – The theory is absurd and subversive of religion and morality. The propounder both fool and knave.
2. 20 years later – The theory is absolute Truth and will yield a full and satisfactory explanation of things in general. The propounder a man of sublime genius and perfect virtue.
3. 40 years later – The theory will not explain things in general after all and therefore is a wretched failure. The propounder a very ordinary person advertised by a clique.
4. 100 years later – The theory is a mixture of truth and error. It explains as much as could reasonably be expected. The propounder worthy of all honor in spite of his share of human frailties, as one who has added to the permanent possessions of science.

The long delay in the acceptance of the concepts of plate tectonics (continental drift) and the rejection of the clear evidence of the great Pleistocene floods that so modified the Scabland landscape of eastern Washington should not, however, be used as evidence that scientists are not objective, as it is difficult to accept new hypotheses when they do not have what appears at the time to be a physical basis. What was the mechanism that moved continents? Where did the water originate that caused the floods? In the latter case, when glacial lake Missoula in western Montana was discovered and linked to the megafloods, the origins of the Channeled Scablands by regional flooding was accepted (Bretz, 1969; Baker, 1973). Hence, not all resistance to change is irrational. The explanation must be related to physical reality.

However, Giere (1988, p. 292) makes the suggestion that the 'virtues of multiple working hypotheses tend to be advanced by scientists attempting to gain a hearing for a new view', and these virtues are minimized by scientists who are defending an established position. This suggests that older scientists are less open to new ideas. Grinnell (1987, p. 292) however, expresses a more charitable view that scientists who are deeply involved in a problem know a great deal about it, and they can legitimately raise objections to new hypotheses. This indeed, makes the younger generation of investigators seem to be more open-minded, but the end result of such a controversy may be that 'older hypotheses are not disproved so much as they are replaced by newer ones' (Grinnell, 1987, p. 39).

Falsification

In order to overcome prejudices and conditioning the 'critical rationalist approach' of Popper (1968) produces a conscious attempt to disprove or falsify a hypothesis. That is, we should attempt to disprove rather than verify our hypotheses. One may argue that we do falsify hypotheses when the data fail to support them, but it is the negative and critical state of mind that is cultivated in the falsification approach that is encouraged. Although philosophically this may be advantageous and a more objective approach, it is difficult to imagine most scientists struggling to disprove their ideas, where, as is the case in earth science, supporting or disproving data may be so limited.

Popper argues that it is impossible to prove the correctness of a hypothesis, but one can disprove it by one failure. We can take hope in the assumption that science is 'self-correcting'. That is, error or fraud will be revealed with time. Popper's approach is certainly different from what has been described as the scientific method, but it could lead to identification of new data that are needed or additional observations that need to be made that will either support a hypothesis or destroy it.

An example of the falsification approach is provided by a medical situation described by Rouché (1984) in which the disease rikettsial pox was diagnosed and the treatment established by Doctor Benjamin Shankman. In February, 1946 a young boy who had a spotty rash, high fever and considerable joint pains was examined by Dr Shankman. The doctor was unable to diagnose this disease. The boy was sufficiently ill to be placed in the hospital and he was treated with penicillin for infection, codeine for pain and aspirin for fever. After approximately five days the boy recovered. Shortly afterwards a woman with the same symptoms was admitted to the hospital by Dr Shankman. He wondered whether penicillin really resulted in the cure of the boy, and he modified his treatment by prescribing sulfonamide instead of penicillin. The woman recovered. A third patient arrived with the same symptoms, and Dr Shankman questioned the need for antibiotics.

He treated the woman only with codeine and aspirin, and the woman recovered. Dr Shankman's conclusion was that the disease would run its course in a few days and that the patient would recover without any special treatment other than that required to reduce joint pain and fever (codeine and aspirin). Although in the first instance the treatment was successful, Dr Shankman was not convinced that the treatment was necessary, and he falsified his hypothesis that an antibiotic was required. It was later determined that the disease was transmitted by mites living on mice in a large apartment complex in New York. In this case, the diagnosis of the disease was correctly established long after the 'cure' was found by Dr Shankman and, therefore, his falsification procedure was applied to treatment rather than to the cause of the disease.

Probability and quality of a hypothesis

Strahler (1987), in his excellent book on the evolution–creation controversy takes issue with Popper's approach and suggests that if it is not possible to prove a hypothesis then it is not possible to disprove or falsify it. Rather we should think of the probability of the hypothesis being false or true.

He states his case as follows:

> To summarize the concept of quality of a scientific hypothesis and to give it graphic expression that may make it easier to grasp, I have drawn a . . . 'ladder of excellence' [Fig. 2.5] on which each 'rung' denotes a tenfold increase in quality. We must take into account two complimentary statements of probability:
>
> P_T: Probability that the hypothesis is true.
>
> P_F: Probability that the hypothesis is false.
>
> The sum of these two probabilities is unity; therefore, as P_T increases, P_F decreases. However, we must keep in mind that neither value can actually reach either unity or zero – those limits can only be approached, so our ladder has no upper or lower end. Gambling odds on the hypothesis actually being a true statement of nature are given as the ratio of P_T to P_F. I have suggested adjectives to describe the quality of the hypothesis, but these are quite subjective because, as with people in general, scientists differ among themselves as to the risks they are willing to take in a given situation. At some point high on the ladder, the hypothesis may take on the status of a law of science.

Certainly in the earth sciences all except the simplest hypotheses will have a degree of uncertainty, and Stahler's approach makes a good deal of sense. The problem, of course, is to determine the probabilities of an accurate extrapolation.

Beveridge (1980, p. 58) is also very critical of the falsification approach. He claims that many important scientific advances have been made in spite of apparent falsification, and it was only the investigator's persistence in the face of negative results that led to a solution. Most of the time a scientist is 'trying to achieve some objective, or to understand some phenomenon, not to refute some hypothesis. Antibiotics and the helical structure of DNA were not discovered by trying to disprove a theory'. He would agree with Strahler that it is difficult to design an experiment that gives a clearly negative result, and therefore, most results are expressed, as Strahler suggests, in statistical terms. It seems that falsification of a hypothesis is as uncertain as its confirmation (Grinnell, 1987).

Popper would agree with these criticisms because he is aware that 'no conclusive disproof of a theory can ever be produced' (Popper, 1968, p. 50) and 'The game of science is, in principle, without end. He who decides one day that scientific statements do not call for further test, and that they can be regarded as finally verified, retires from the game' (Popper, 1968, p. 53). Indeed, Chapter 6 of his book deals with degrees of falsifiability. Furthermore, he makes a modern statement of multiple working hypotheses as follows: 'According to my proposal, what characterizes the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested. Its aim is not to save the lives of vulnerable systems but, on the contrary, to select the one which is by comparison

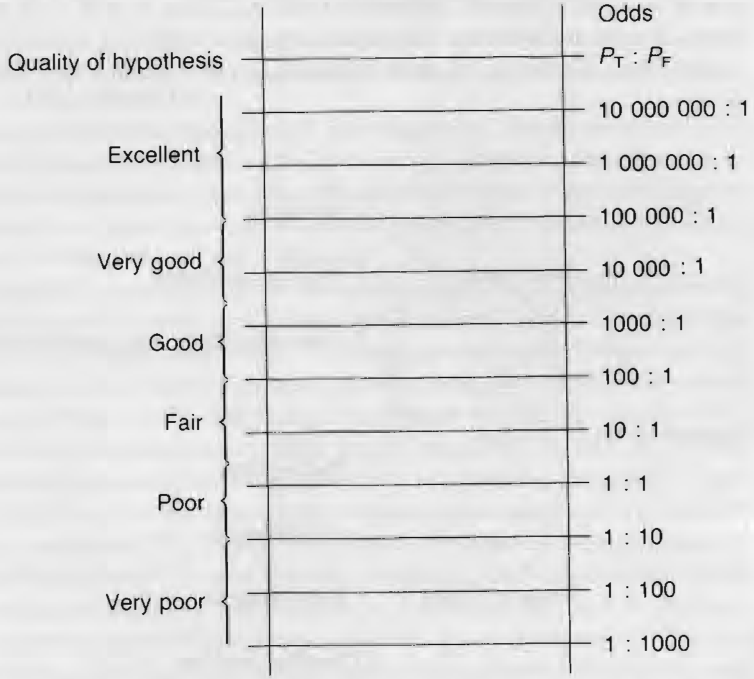

Figure 2.5. The quality or probability of a scientific hypothesis. P_T is probability of being true or of enjoying long survival. P_F is probability of being false or of being short lived (from Strahler, 1987).

the fittest, by exposing them all to the fiercest struggle for survival' (Popper, 1968, p. 42). Substitute in the above the word scientific for empirical and the word hypotheses for system.

Pseudoscience and the scientific approach

Obviously there is no single 'scientific method', but all scientists attempt to approach a problem in such a way that their conclusions are tested. Bunge (1984) contrasted the characteristics of science and pseudoscience, and this may also help to identify a general approach that is scientific. He describes both science and pseudoscience as cognitive fields which are 'a sector of human activity aiming at gaining, diffusing, or utilizing knowledge of some kind, whether this knowledge be true or false.' The cognitive fields can be divided into research fields (science) and belief fields (Fig. 2.6). A research field is continually changing, as a result of research results, but a belief field changes only as a result of controversy, brute force or revelation. Bunge then compares the attitudes and activities of the scientist and pseudoscientist (Table 2.3). The scientist attempts to function with an open mind and uses certain scientific methods (Table 2.3, items 3, 5, 6, 9–14) and a scientific state of mind or a scientific approach (Table 2.3, items 1, 2, 4, 7, 8, 15). The approach need not include a determined attempt to falsify, but it must include a consideration and testing of other explanations and a rational approach to the problem.

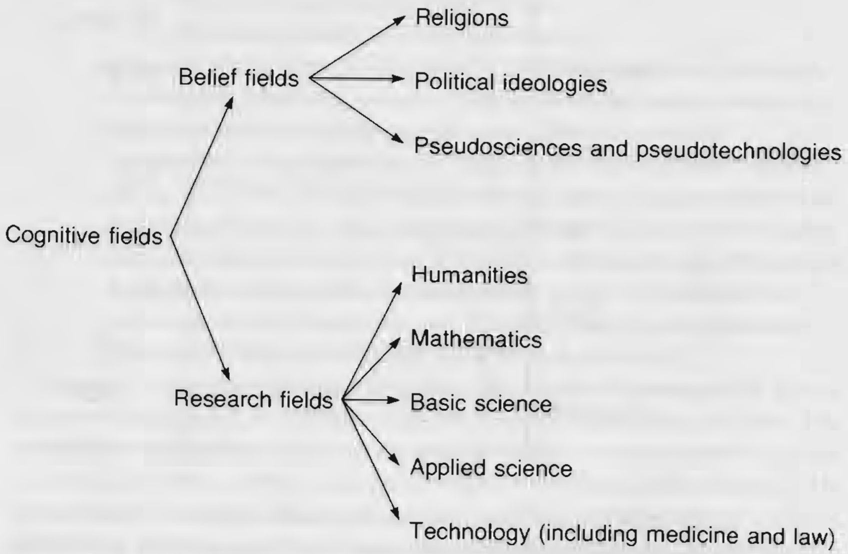

Figure 2.6. Cognitive fields (from Bunge, 1984).

Table 2.3. *Attitudes and activities of scientists (from Bunge, 1984)*

1 Admits own ignorance, hence need for more research
2 Finds own field difficult and full of holes
3 Advances by posing and solving new problems
4 Welcomes new hypotheses and methods
5 Proposes and tries out new hypotheses
6 Attempts to find or apply laws
7 Cherishes the unity of science
8 Relies on logic
9 Uses mathematics
10 Gathers or uses data, particularly quantitative ones
11 Looks for counter-examples
12 Invents or applies objective checking procedures
13 Settles disputes by experiment or computation
14 Updates own information
15 Seeks critical comments from others

DISCUSSION

One may conclude that although there are variations in scientific methods, there is a general scientific approach (Table 2.3) that should lead us on the path to explanation and understanding. Part of the difficulty is the different approaches used in applied and basic research and, of course, the great variety of scientific specialties each with its own problems and techniques.

Although I have attempted to illustrate some approaches to explanation in Fig. 2.2, as based upon the writings of Gilbert and Chamberlin, Fig. 2.7 contains what I perceive to be the probable course of a scientific investigation from recognition of a problem to its solution. This sequence is a combination of all of the paths to explanation of Fig. 2.2. For me, at least, it is difficult to think of more than three or four working hypotheses to explain a problem such as P_1 (H_1, H_2, H_3). As data are obtained, hypotheses can be rejected (H_1, H_3) as being improbable explanations of P_1 but as data are collected, new hypotheses are generated (H_4) and new problems recognized (P_2). Depending upon the importance of the problem, P_1 can be abandoned and P_2 can be followed. Similarly, as the data are analysed, the result will usually cause the abandonment of hypotheses (H_4) and the development of new hypotheses (H_5) and the recognition of new problems (P_3). The final solution will probably be a combination of two or more hypotheses (H_2, H_5) that lead to a composite explanation of the phenomenon of concern (ES) and perhaps recognition of new problems.

The procedure outlined in Fig. 2.7 is obviously a combination of the working hypotheses concepts of Gilbert and Chamberlin and the falsification procedure of Popper. The end result is a most probable composite hypothesis that is the result of a combination of hypotheses.

The warning by Gilbert, Chamberlin and Popper against a ruling hypothesis cannot be too strongly supported. An impressive example is provided by the autobiographical account of Susan Blackmore's (1986, p. 8) efforts to demonstrate the validity of the paranormal, extrasensory perception; she was 'hooked' by parapsychology because 'Parapsychology has everything a hook needs. It is mysterious and alluring. It has just enough "scientific" evidence to provide bait, while at the same time it is rejected by most orthodox scientists, the inspiration for a crusading spirit to shout "I'll show them". And that is, I suppose, what I wanted to do.' And she did for years until finally convinced by continuing negative results to question the ruling hypothesis. Obviously Blackmore was operating in a 'belief field'. Of greater consequence is the ruling hypothesis in politics, political science and law where individuals and societies are damaged by it.

The scientist is somewhat in the situation of a parent. Each child and each scientific problem is different, and clearly a better job would be done if in both cases we could do it again. In both cases methods will vary, but a consistent approach is mandatory (Table 2.3, Fig. 2.7).

In addition to the basic scientific problems being different there are specific procedural problems that may be encountered in the development of explanations of phenomena and in the extrapolation of research findings to analogous and homologous situations. Ten of these problems will be discussed in Chapter 3.

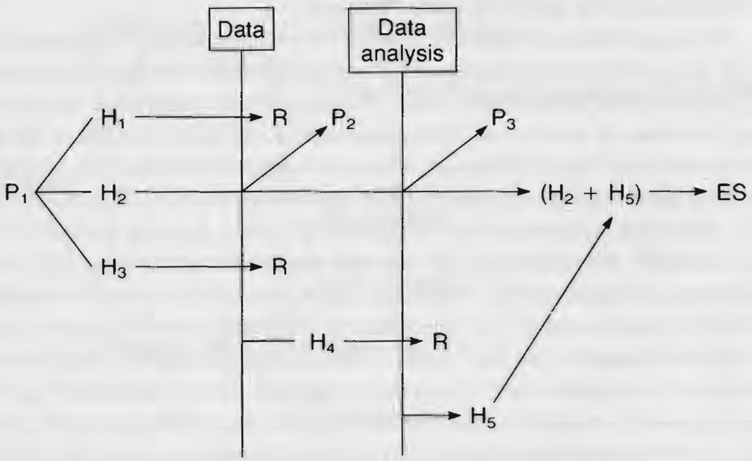

Figure 2.7. The path to explanation (P = problem; H = hypothesis; ES = explanation and solution; R = rejection).