

GIORA HON*

TOWARDS A TYPOLOGY OF EXPERIMENTAL ERRORS: AN EPISTEMOLOGICAL VIEW

There *is* no such thing as a classification of the ways in which men may arrive at an error; it is much to be doubted whether there ever *can be*.¹

A. de Morgan (1847)

Abstract — This paper is concerned with the problem of experimental error. The prevalent view that experimental errors can be dismissed as a tiresome but trivial blemish on the method of experimentation is criticized. It is stressed that the occurrence of errors in experiments constitutes a permanent feature of the attempt to test theories in the physical world, and this feature deserves proper attention. It is suggested that a classification of types of experimental error may be useful as a heuristic device in studying the nature of these errors. However, the standard classification of systematic and random errors is mathematically based and as such does not focus on the causes of the errors, their origins, or the contexts in which they arise. A new typology of experimental errors is therefore proposed whose criterion is epistemological. This typology reflects the various stages that can be discerned in the execution of an experiment, each stage constituting a category of a certain type of experimental error. The proposed classification consists of four categories which are illustrated by historical cases.

1. General Discussion

IN HIS BOOK, *Mathematical Elements or Natural Philosophy Confirm'd by Experiments*, which he intended as an introduction to Newton's natural philosophy, Gravesande (1688–1742) — the propounder of Newtonian physics on the Continent — demands that nature should be examined 'attentively and incessantly . . . with indefatigable Pains. That Way', he maintains, 'our Progress will be but slow, but then our discoveries will be certain'; indeed, in his view, even the limits of human understanding could thus be determined.² 'What has led most People into Errors', he observes, 'is an immoderate Desire of Knowledge, and the Shame of confessing our Ignorance. But Reason should get the better of that ill-grounded Shame; since', as he writes — perhaps under the influence of Cusanus — 'there is a learned Ignorance that is the Fruit of Knowledge, and which is much preferable to an ignorant Learning'.³ Although

*Department of Philosophy, Haifa University, Mt Carmel, Haifa 31999, Israel.
Received 10 May 1988

¹ De Morgan (1847), p. 237.

² Gravesande (1726), p. viii (Preface).

³ *Ibid.* See Cusanus (1981).

Gravesande admits that many things in nature are hidden from us,⁴ he nevertheless holds that ‘what is set down in Physics, as a Science, is undoubted’.⁵ Thus, notwithstanding his explicit awareness of the limits and faults of human understanding, Gravesande does hold that the results of physics are certain.

According to Gravesande it is the method of mathematical demonstration — a method which the physicist of the scientific revolution employs to deduce, and thus to explain, ‘from a few general Principles numberless particular Phenomena or Effects’⁶ — that furnishes physics with certainty. ‘Whoever will go about that Work any other Way, than by Mathematical Demonstrations, will be sure’, Gravesande warns, ‘to fall into Uncertainties at least, if not into Errors’.⁷ In his view, this way, the method of mathematical demonstration, is in fact the Newtonian method.

Gravesande is explicit about the trust he invests in the Newtonian method. ‘Philosophers’, he observes, ‘do not equally agree upon what is to pass for a Law of Nature, and what Method is to be followed in Quest of those Laws’.⁸ However, for Gravesande there is no doubt that one ought to follow the Newtonian method. And he maintains that ‘whoever will seriously examine, what Foundation this Method of Physics is built upon, will easily discover this to be the only true one, and that all Hypotheses are to be laid aside’.⁹

The great innovation of Gravesande is the attempt to illustrate, as he writes, ‘every Thing by Experiments, and to set the very Mathematical Conclusions before the Reader’s Eyes by this Method’.¹⁰ Although he admits that ‘Mathematicians think Experiments superfluous, where Mathematical Demonstrations will take Place’,¹¹ he has no doubt — since the mathematical conclusions are abstract — that it is easier to grasp these conclusions ‘when Experiments set [them] forth . . . before our Eyes’.¹²

This display of confidence should be contrasted with the cautious advice which Einstein gave the young Heisenberg in the spring of 1926. ‘You must appreciate’, Einstein instructed Heisenberg, ‘that observation is a very complicated process’.¹³ The phenomenon under observation, Einstein explained,

⁴ Gravesande, *ibid.*, pp. xii–xiii.

⁵ *Ibid.*, p. xiii.

⁶ *Ibid.*

⁷ *Ibid.*

⁸ *Ibid.*, p. ix.

⁹ *Ibid.*, p. xvii. Gravesande perceives this method as a two-stage procedure; first there is the deduction of the laws of nature from the phenomena and, second, there is a process of induction to prove the generality of the deduced laws. (*Ibid.*)

¹⁰ *Ibid.*

¹¹ *Ibid.*, p. xviii.

¹² *Ibid.* Gravesande seems to go here beyond Newton by introducing some Cartesian elements into Newton’s methodology. The notion that experiment is a didactic aid has many undertones of the rationalism of Descartes.

¹³ Heisenberg (1971), p. 63.

produces certain events in our measuring apparatus. As a result, further processes take place in the apparatus, which eventually and by complicated paths produce sense impressions and help us to fix the effects in our consciousness. Along this whole path — from the phenomenon to its fixation in our consciousness — we must be able to tell how nature functions, must know the natural laws at least in practical terms, before we can claim to have observed anything at all.¹⁴

In other words, even when we execute this basic requisite of science, namely observation, we ought to be explicitly aware of the assumptions involved. According to Einstein, we must be aware of the fact that while taking observations we assume, even if we are about to formulate a new law of nature which contradicts the old ones, that the current laws — believed to cover the whole path from the phenomenon to our consciousness — function in such a way that we can rely upon them.¹⁵ The possibility of one such law being incorrect or the ‘chain’ which leads from the phenomenon to the observer’s consciousness being faulty, should not of course be ruled out.

Although Einstein would presumably have agreed with Gravesande about the power of mathematical demonstrations in physics, he would not have considered the results of these demonstrations certain. Einstein succinctly formulated this view with the maxim that ‘as far as the proposition of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality’.¹⁶ Experimental error is a recurring feature of the attempt to test the results of these demonstrations in the physical world. As Morris Cohen remarks,

everyone who has ever worked in a laboratory or with instruments of precision knows that the simple laws of nature, so clearly formulated in elementary and popular treatises, are never verified with absolute accuracy. The results of actual measurements always differ.¹⁷

It is a standard practice to attribute this lasting discrepancy between the theoretical formulae and the actual results of measurements not to some faults in the theory but rather to the ‘error’ of the instruments. However, the refinement and improvement of the available instruments never eliminate the discrepancy between theory and experiment. ‘On the contrary’, Cohen concludes, ‘it often compels us to abandon the simple law in favour of a more complicated one’.¹⁸

Yet there is a prevalent belief that exact agreement can be attained. Jeffreys suggests that the use of the expression ‘exact science’ and the scarcity of explicit statements in popular works on physics concerning the imperfection of

¹⁴ *Ibid.*

¹⁵ The thrust of Einstein’s argument is directed against the view that none but observable magnitudes must go into a physical theory. In Einstein’s view, ‘it is quite wrong to try founding a theory on observable magnitudes alone. . . Only theory, that is, knowledge of natural laws, enables us to deduce the underlying phenomena from our sense impressions.’ (*Ibid.*)

¹⁶ Einstein (1921), p. 228.

¹⁷ M. R. Cohen (1949), p. 224.

¹⁸ *Ibid.*; cf., Duhem (1974), pp. 168–179.

agreement between physical laws and observation, have contributed to this misleading belief. Where errors of observation are mentioned at all they are dismissed, in Jeffreys' words, 'as a minor complication'.¹⁹ Sellars' casual remark that 'once [errors of measurement and other forms of experimental error]. . . have been discounted, our attention can turn to the logico-mathematical structure',²⁰ epitomizes this attitude towards the problem of error. Indeed, with a few exceptions,²¹ the problem of error has been treated as quite incidental to the pursuits of science: 'a tiresome but trivial excrescence on the neat deductive structure of science'.²²

¹⁹ Jeffreys (1973), p. 63. Jeffreys observes that a typical presentation in textbooks of physics starts by stating a set of laws, taken as fundamental, and proceeds to develop a series of consequences — the observational evidence being only incidental. This approach is unsatisfactory and indeed misleading since it takes no account of errors; if the results were compared with observation directly, almost every law would be rejected. As Jeffreys remarks, 'the observed value hardly ever agrees exactly with prediction, and the laws are taught as if they were exact. . . [this treatment] is essentially inapplicable unless it is supplemented by the notion of error'. (*Ibid.*, p. 80.) In Jeffreys' view 'the laws of physics, so far as they relate to observations, become statements of probability distributions. The quantities of physics arise fundamentally as parameters in these probability distributions.' (*Ibid.*) The principal objective of Jeffreys is to show that 'errors of observation have to be considered in the process of establishment of the laws: we can, if we like, say that there is a form of the law that expresses exact relations between true values, but the law that is verified is a modification of this that takes account of probabilities of errors of observation in different ranges'. (*Ibid.*, p. 212.) Jeffreys points out that there are three fundamental misconceptions which pervade most modern accounts of scientific principles and are the principal source of confusion. 'The first is that in some sense scientific laws are statements made with certainty. The second is that physical measures can be exact. The third is that there is a clearly marked boundary between science and ordinary thought.' (*Ibid.*, p. 183.) According to Jeffreys, the second fallacy arises from ignoring the occurrences of observational errors. Jeffreys' general objective is to show how the first two fallacies are avoided by the theory of probability.

²⁰ Sellars (1961), p. 73; quoted by Mellor (1965), p. 106.

²¹ Chwistek (1948). According to Chwistek, measurement is a crudely defined activity. Since slight differences in the results of a measurement are disregarded, a measurement can give no basis for the establishment of a one to one correspondence between its results and a real number. In other words, as he writes, 'many real numbers correspond to one measurement and the class of real numbers is not precisely determined. The only way to make this statement more precise is to fix the limits between which the number obtained by measurement can vary, i.e. to designate two numbers, between which the number obtained by measurement may be found'. (*Ibid.*, p. 256.) The thrust of Chwistek's criticism is that 'the concept of ideal length is meaningless and that experience furnishes not determinate numbers but classes of numbers which lie between certain limits'. (*Ibid.*, p. 257.) Presumably, Chwistek would have agreed with Born's claim that 'statements like "A quantity x has a completely definitive value"' (expressed by a real number and represented by a point in the mathematical continuum) seem. . . to have no physical meaning'. (Born, 1956, p. 167.)

²² Mellor (1967), p. 6. Scriven's view is a subtle example of this attitude; while acknowledging that laws of nature are virtually all known to be in error, he concludes that since 'they represent great truths. . . we forgive them their errors'. (Scriven, 1961, pp. 91, 101.) Mellor is in fact a critic of this attitude. In a series of papers published in *Philosophy of Science* (Mellor, 1965, 1966, 1967), he 'tried to show how the effects of experimental error or imprecision can be accommodated within a testable, deductive scientific structure'. (Mellor, 1965, p. 121.) There is now a growing interest in the critical examination of the *actual* practice of the scientific method in general and the method of experimentation in particular. Different perspectives are taken: from an internal analytical point of view to sociological reconstructions and anthropological studies. See for example the following works: Franklin (1987); Franklin and Howson (1984, 1988); Cartwright (1984); Hacking (1983); Pickering (1984); Gooding (1981); Gooding (1982); Galison (1985, 1987); Latour and Woolgar (1979). For a rich bibliography see the editorial in *Isis*: Sturchio (1988). As Pickering demands, 'experimental techniques, methods, and procedures — the very stuff of empirical science — should not be treated as unproblematic adjuncts of some higher theoretical exercise'. (Pickering 1981, p. 235.)

Jeffreys emphatically states that ‘exact agreement between physical laws and observation was never attained’,²³ and he criticizes the exaggerated attention which the uncertainty principle has attracted. Although ‘the uncertainties treated in the quantum theory are far smaller than any of the discrepancies between previous theories and observational results . . . these [discrepancies]’, Jeffreys remarks, ‘had attracted little attention from philosophers’.²⁴ J. O. Wisdom indeed claims that ‘all the standard philosophies of science that have been current in this century fail to solve the problem of the data-theory gap’.²⁵

The experimenter however constantly encounters these discrepancies between theories and observational results; indeed, part of his or her daily routine consists, as M. Polanyi intimates, in explaining away these discrepancies. ‘In my laboratory’, Polanyi writes,

I find the laws of nature formally contradicted at every hour, but I explain this away by the assumption of experimental error. I know that this may cause me one day to explain away a fundamentally new phenomenon and to miss a great discovery. Such things have often happened in the history of science. Yet I shall continue to explain away my odd results, for if every anomaly observed in my laboratory were taken at its face value, research would instantly degenerate into a wild-goose chase after imaginary fundamental novelties.²⁶

Focusing our attention on these discrepancies we see that their origins can be of different kinds: occurring in different contexts and arising from different

²³ Jeffreys (1973), p. 183.

²⁴ *Ibid.*, pp. 63–64. Evidently, the attraction of the uncertainty principle is due to the very fact that it is indeed a principle: a fundamental limitation on the accuracy with which one can execute a measurement independent of one’s experimental skill. There is, however, a phenomenon which is associated with classical physics that sets a definite limit to the ultimate sensitivity of measuring instruments beyond which one cannot advance, namely Brownian motion. (Ising, 1926; Barnes and Silverman, 1934.) To be sure, classical physics is based on idealization which can be valid only under the assumption that absolute accuracy is within reach; that is, that mathematical accuracy can be attained with respect to the initial state of the physical system involved. F. Waismann takes stock of this classical attitude; he remarks that ‘the accuracy of any measurement, it was supposed — rather light-heartedly as it would seem now — could be increased to any degree by improved technique. Irrespective of whether absolute precision will ever be attainable with our blunt instruments, we can at any rate go on refining our measuring methods, it was assumed, and proportionally our predictions concerning the future will become more and more reliable. That there is *no limit* to this approach, this was, ultimately, the tacit assumption underlying classical physics, and one so brilliantly vindicated by the successes in astronomy’. (Waismann, 1959, p. 107.) Waismann associates the decline of causality with the recognition of ‘how utterly Utopian the idea of absolute precision is’; (*Ibid.*, p. 112) and the fall of causality with the discovery of the uncertainty principle. (*Ibid.*)

²⁵ Wisdom (1971), p. 281. Wisdom argues that ‘the fundamental assumption that sets the problem for . . . all three approaches [instrumentalism, conventionalism and classical induction], is the givenness of observations; observations are simply there, as the empiricist tradition has it, waiting to be recorded by that classical *camera obscura*, the human mind. In other words, they share a philosophy of observationalism’. (*Ibid.*) On the gap between mathematical accuracy and physical approximation see also Duhem (1974), pp. 132–143; cf., Mellor (1965), pp. 110–111; Laymon (1985).

²⁶ Polanyi (1964), p. 31. Polanyi concludes that ‘just as there is no proof of a proposition in natural science which cannot conceivably turn out to be incomplete, so also there is no refutation which cannot conceivably turn out to have been unfounded. There is a residue of personal judgement required in deciding — as the scientist eventually must — what weight to attach to any particular set of evidence in regard to the validity of a particular proposition’. (*Ibid.*)

causes. To clarify this complex array of different causes it is useful to have, as a heuristic device, a system of classification of experimental error.

The most common classification of experimental error is the classification which distinguishes between two categories of error: systematic and random errors. As we shall see, this is not the kind of classification needed for our purposes.

Systematic errors, as their name indicates, systematically obstruct the measurement from reaching the intended actual value by shifting the result — either positively or negatively — by a magnitude which may be constant or may vary in some regular fashion. By contrast, random errors are disordered in their incidence and vary accidentally in their magnitude. A simple example of the former is an incorrect calibration of the measuring instrument; undue mechanical vibrations of the equipment — vibrations which interfere with the measurement — constitute a cause of the latter.²⁷

The distinction between the terms accuracy and precision — terms which are mistakenly used interchangeably — corresponds to this dichotomy between systematic and random errors. Accuracy refers to the closeness of the measurements to the ‘true’ value of the sought physical quantity, whereas precision indicates the closeness with which the measurements agree with one another independently of their relations to the ‘true’ value. Accuracy thus implies precision but the converse is not necessarily true.²⁸

In scientific and technical writings it is common to find different usages of the word error. One school of thought considers error the difference between the experimental result and the ‘true’ value; another usage is that error is the number placed in the statement of the result after the plus-or-minus sign, irrespective of the ‘true’ value.²⁹ It has been suggested — in a *Code of Practice* addressed to the National Physical Laboratory — to use the term uncertainty to cover this multiple usage of the word error.³⁰ In this *Code of Practice* the uncertainty of a measurement is divided into two categories: the random uncertainty and the systematic uncertainty.³¹ According to this *Code*, the estimation of random uncertainty is derived by a statistical analysis of repeated

²⁷ See for example Topping (1975), p. 10; Parratt (1961), pp. 64–69.

²⁸ Topping, *ibid.*, p. 14; Parratt, *ibid.* Cf., Franklin (1981), p. 367.

²⁹ Campion *et al.* (1980), p. 25. At the time of publication, Campion was the Deputy Director of NPL. E. R. Cohen and J. W. M. DuMond observe in their 1965 review of the fundamental constants that there seem to be two completely incompatible ways in which experimenters regard experimental error. ‘Some regard the number following the sign of ambiguity as expressing “limits of error” with the unstated implication that the true value lies anywhere within the gap and that there is something rather virtuous in overestimating the magnitude of this gap “for safety” or “to take care of possible but unknown sources of systematic error”. . . Such an error estimate. . . is not a quantitative estimate, but a statement of inequality: the error is less than or equal to so-and-so much’. Cohen and DuMond in fact deplore this approach. (Cohen and DuMond, 1965, pp. 540–541.)

³⁰ Campion *et al.*, *ibid.*, pp. 1, 30.

³¹ *Ibid.*, p. 1.

measurements while the estimation of systematic uncertainty is assessed by non-statistical methods and much depends on the judgement of the experimenter in allocating limits to this uncertainty.³²

From a different perspective, Jeffreys has defined systematic errors as 'errors that *could be* precisely calculated for each observation, given the values of certain parameters'.³³ Evidently, random errors do not have this feature. It thus appears that whatever the terms and the definitions, the methods used for estimating uncertainties constitute the underlying criterion of the dichotomy between systematic and random error.

Many textbooks and manuals on the design of experiments and the treatment of data can attest to the general acceptance of this classification which is sometimes even considered exhaustive. Typically, it is claimed that the skilled experimenter can eliminate all systematic errors. Upon accomplishing this task the experimenter, it is further claimed, finds that there is still a margin of error which requires further consideration. This margin of error, it is asserted, is due to random errors.³⁴ Thus, the experimenter is culpable when he or she commits so to speak systematic errors: the experimenter may be blamed for failing to eliminate systematic errors or to account for them, but he or she may not be blamed for the occurrence of random errors; the latter are inevitable and the experimenter is not liable for their occurrence. It is perhaps due to this clear-cut distinction between culpability and inevitability that this classification has become so attractive to experimenters, so much so that they have never set out to question its conventional foundation and accepted it without any reservation.

The advice offered to the skilled experimenter concerning the inevitable occurrence of random errors, is that

³² *Ibid.*, pp. 14–15, 23.

³³ Jeffreys (1973), p. 77 (my emphasis).

³⁴ Cramér expresses this attitude in these words: 'Obviously. . . [systematic errors] will not lend themselves to probabilistic treatment, and accordingly we shall assume in the sequel [i.e. Cramér's treatise] that we are dealing with observations which are free from systematic errors'. (Cramér, 1966, p. 229.) Cf., Worsnop and Flint (1951), p. 5; Young (1962), p. 3. Parratt describes faults such as mistakes in recording data and in calculation as outright mistakes and calls them blunders. (Parratt, 1961, p. 69.) Thomson and Tait assign these faults to a third class: the class of what they call avoidable mistakes. (Thomson and Tait, 1872, pp. 112–113.) Worsnop and Flint also call errors of this type mistakes. They use the term mistake to denote a fault of measurement or of observation which can be avoided by care on the part of the observer. Thus, they remark that 'inexperienced observers or observers not in a normal state [*sic*] make errors of varying magnitude which should strictly be described as mistakes'. (Worsnop and Flint, 1951, pp. 3–4.) Evidently the personal qualification of the experimenter is at the root of this usage of the term mistake. Worsnop and Flint distinguish further between constant and systematic errors. In their view, 'constant errors are those which affect the results of a series of experiments by the same amount. An example is the case of the faulty graduation of a scale'. (*Ibid.*, p. 4.) They maintain that these errors are distinct from systematic errors which 'occur according to some definite rule, such as would be the case in readings on a circular scale if the pointer were not pivoted at the centre'. (*Ibid.*) However, this is not a categorical distinction; in fact by these very definitions, a constant error is a particular instance of systematic error.

it is found empirically that. . . random errors are frequently distributed according to a simple law. This makes it possible to use statistical methods to deal with random errors.³⁵

In sum, the standard view is that, apart from random errors, all experimental errors can be eliminated, and that the distribution of random errors can be captured by a simple law, namely, the Normal law, which, it is claimed, has been established empirically.

However, in practice it is very rarely, if ever, the case that the experimenter can remove the systematic errors altogether. In Duhem's view the estimation of the degree of exactness of an experiment requires, among other things, that 'we evaluate the systematic errors that could not be corrected; but, after making as complete an enumeration as possible of the causes of these errors, we are sure to omit infinitely more than have been enumerated, for the complexity of concrete reality is beyond us'.³⁶

Furthermore, the distribution of errors follows the Normal law only approximately, even when the quantity to be measured is as steady as possible.³⁷ Although this approximation can be justified under much wider conditions, namely the central limit theorem, it is still the case, Jeffreys remarks, that these conditions 'are seldom *known* to be true in actual applications. Consequently', Jeffreys argues, 'the normal law as applied to actual observations can be justified, in the last resort, only by comparison with the observations themselves'.³⁸ Needless to say, this is a vicious circle; it is the result of justifying the treatment of observations by exclusively referring to the observations themselves.

This state of affairs has led to much confusion with respect to the validity of the Normal law; a confusion to which Lippmann wittily referred in his remark to Poincaré:

Everybody believes in the exponential law of errors: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation.³⁹

In the NPL's *Code of Practice*, Campion *et al.* state quite clearly that 'there is of course no reason for experimental observations to follow the Normal distribution exactly — it is a convenient mathematical expression which fits most of the experimental observations'.⁴⁰ They stress that 'it should be recognized that this is an assumption which may not always be justified'.⁴¹

³⁵ Young (1962), p. 3.

³⁶ Duhem (1974), p. 162.

³⁷ Jeffreys (1973), p. 67.

³⁸ *Ibid.*, emphasis in the original; cf., Cohen and DuMond (1965), pp. 540–541.

³⁹ Quoted by Whittaker and Robinson (1924), p. 179. (See Poincaré, *Calcul des Probabilités*, p. 149.)

⁴⁰ Campion *et al.* (1980), p. 11.

⁴¹ *Ibid.*, p. 5.

Indeed, as Margenau critically remarks

experience presents the scientist with innumerable *skew* distributions, differing perceptibly from the Normal law. These he often dismisses or corrects, because for some hitherto unstated reason he objects to them. He uses the normal distribution both as an inductive generalization from experience and as a criterion for the trustworthiness of that experience. Thus he is lifting himself by his bootstraps unless an independent argument can be given for the normalcy of that distribution.⁴²

The correct approach is 'to regard the number following the plus-or-minus sign as an estimate of the width parameter of some', and it should be stressed, *some* 'statistical distribution of observed values which would be obtained if the measurement were replicated a number of times'.⁴³ Clearly, the appeal to probability is an attempt to break the vicious circle.

Two objections which are closely connected may be raised against this classification of systematic and random errors. First, as the protagonists of this classification are interested in the resultant error and not so much in its source, they are bound to classify together phenomena which may indeed perpetrate as it were the same kind of error, e.g. random error, but are nevertheless distinct as to their causes. Thus, for example, 'small errors in judgement on the part of the observer, such as in estimating tenths of the smallest scale division',⁴⁴ are classified together with 'unpredictable fluctuations in conditions, such as temperature. . . or mechanical vibrations of the equipment'.⁴⁵ The implication of this arrangement is that errors which are peculiar to an individual observer are conflated with errors that have originated in the instrument.

Grouping together such diverse sources of error obscures the nature of experimental error. To put it differently, the dichotomy between systematic and random errors does not focus on the source of the error; rather, it examines the nature of the error by applying a mathematical criterion. This criterion judges whether the estimation of the error is derived by a statistical analysis of repeated measurements or by non-statistical methods in which much depends on the judgement of the experimenter in allocating limits to the accuracy of the measurement. As we have seen, the former error will be random, the latter systematic.

The fact that the criterion which underlies this dichotomy is mathematical constitutes the second objection. Most writers in this field satisfy themselves with a few remarks concerning the origin and treatment of systematic errors. Having briefly dealt with systematic errors, these authors proceed to a detailed

⁴² Margenau (1950), p. 114.

⁴³ Cohen and DuMond (1965), p. 540.

⁴⁴ Young (1962), pp. 2-3.

⁴⁵ *Ibid.*, p. 3. For another example see Worsnop and Flint (1951), p. 4. They place errors which arise from instruments and external conditions such as temperature variation on an equal footing with those arising from personal idiosyncrasies, since all of these errors are in their view systematic.

analysis of the theory of error which provides the mathematical tool for treating random errors. However, under this treatment the error which we know to originate in a certain conceptual framework and under some physical conditions, is transformed into a technical term which is defined mathematically as either the difference between the measurement results and the 'true' value, or the departure from the mean — the mean being another mathematical concept designed to capture the plurality of the results. In this sense, errors are in effect residuals, as indeed Jeffreys calls them.⁴⁶

The upshot of such analysis is that the experimenter gets a mathematical insight into his or her collection of data. Indeed, the object of the mathematical theory of error 'is to work out methods for estimating the numerical values of the required magnitudes by means of a given set of observations, and also to make it possible for the observer to arrive at some conclusion with respect to the degree of precision of the estimates obtained'.⁴⁷ In accordance with that theory the experimenter identifies the unknown 'true' value of the observed magnitude with the mean of the corresponding Normal distribution, and the degree of precision with the standard deviation of the distribution.⁴⁸ Commenting on this theory of error, Margenau remarks that

the philosopher of science is obliged to take note of this remarkable fact: both 'truth' and 'tolerance' must be fished out of the uncertainties of the immediately given by more or less arbitrary rules not immediately presented in Nature.⁴⁹

Be that as it may, the experimenter remains however in the dark as to the conceptual and physical circumstances in which the errors have originated.⁵⁰

To illuminate these aspects we shall treat experimental error as a twofold

⁴⁶ Jeffreys (1973), p. 64.

⁴⁷ Cramér (1966), p. 229.

⁴⁸ *Ibid.*, p. 230. In view of the fact that the mathematical theory is somewhat arbitrary, it is no surprise that there is no general agreement among scientists as to the adequacy of the standard deviation as a measure of precision, and that some other measures are being used, e.g. probable error. In their 1965 review of the fundamental constants, Cohen and DuMond state their preference for expressing all their error measures in terms of standard deviations: the root-mean-square deviations from the mean, rather than in terms of probable errors. In their view, 'it is a grave mistake... to regard the error expressed in standard deviations as simply differing by a multiplying factor (1.48) from the probable error. The probable error is defined as that measure of dispersion such that the odds are even that the observations may lie either inside or outside the given limits. Only for the Gaussian [i.e. Normal] distribution are the standard deviation and the probable error related to each other by the numerical factor just cited. They are two entirely different kinds of dispersion measures, and the standard deviation is far more general in its applicability under the theory of least squares'. (Cohen and DuMond, 1965, p. 541.) Campion *et al.* note in their *Code of Practice* that 'the standard deviation depends only on the precision of the technique and apparatus used and, so long as these are not changed, it will not change significantly however many observations are taken; on the other hand, the standard error of the mean depends on the number of observations in the sample as well as on the precision of the technique. The two quantities... present different information', conclude Campion *et al.* and recommend that 'the statement of both in a result may sometimes be appropriate'. (Campion *et al.*, 1980, p. 7.)

⁴⁹ Margenau (1950), p. 113.

⁵⁰ Cf. Duhem (1974), p. 162.

phenomenon consisting of conceptual and physical elements, and not as a mathematical abstraction. Indeed, it appears that these two elements are invariably interwoven: there is always a certain physical condition and the experimenter's conceptual understanding of it. Most scientists and philosophers of science regard error as an essentially probabilistic phenomenon resulting from some stochastic process.⁵¹ Although the probabilistic approach is of considerable importance in coping with inaccuracy and imprecision, we are concerned here with error as an epistemological concept, and this view calls for a different classification of experimental error than the dichotomy of systematic and random error.⁵²

To arrive at a classification which focuses on the epistemological nature of experimental error, one has to distinguish between the different contexts in which error may arise, and within each context to determine the kind of possible reasons for an error to occur.

I discern four distinct stages in the execution of an experiment:

- (1) laying down the theoretical framework of the experiment;
- (2) constructing the apparatus and making it work;
- (3) taking observations or readings; and,
- (4) processing the recorded data and interpreting them.

Corresponding to these four stages we may classify experimental errors as arising in:

- (1) background theory;
- (2) assumptions concerning the actual set-up and its working;
- (3) observational reports; and,
- (4) theoretical conclusions.

2.1. Background Theory

Any attempt to create artificially certain physical conditions with a view to examining experimentally a particular phenomenon, indeed, even the making of the simplest observations, takes place within a theoretical framework which underpins the experimental inquiry and thus determines and directs it. This theoretical framework — the 'background theory' — is tentatively taken for granted and considered correct. The experiment, or for that matter the observation, is not designed to test or demonstrate the background theory; rather, the experimenter relies upon this theory to advance the physical

⁵¹ Salmon, for example, holds that the concept of error is basically probabilistic. (Salmon, 1967, p. 65.)

⁵² Historical studies of experimental error as a probabilistic concept have been carried out by L. Tilling (1973) and O. B. Sheynin. Sheynin published a series of papers on this subject in the *Archive for History of Exact Sciences* (see in particular his general historical essay, 1973). For a complete bibliography of his study see his 1983; cf., 1966. For an historical study of error as an epistemological phenomenon see Hon (1987b, 1989).

Table 1. A classification of types of experimental error

Categories	Background theory	Assumptions concerning the actual set-up and its working	Observational reports	Theoretical conclusions
The stages of an experiment	Laying down the theoretical framework	Constructing the apparatus and making it work	Taking observations or readings	Processing the recorded data and interpreting them
Possible sources of mistake and error	<p>(1) False general theory</p> <p>(2) False instrumental theories</p> <p>(3) False theory of the set-up</p>	<p>(1) Parts which are not up to the standard</p> <p>(2) The dirt factor: impure samples; poor vacuum; convection current; thermal expansion; diffusion</p> <p>(3) Incorrect calibration; faulty graduation of a scale; pointer is not pivoted at the centre</p>	<p>(1) Observational error: parallax; aberration; refraction; effective resolution</p> <p>(2) Personal equation: colour blindness; slow reaction time; expectation; anticipation; hallucination</p>	<p>(1) Mistakes in applying the theory of error to the collection of data and errors in assessing systematic errors</p> <p>(2) Misunderstanding the result: error of interpretation; fallacious reasoning in either the induction or deduction method</p>
Examples	(1) Millikan vs. Ehrenhaft: measurements of the charge of the electron. (Inappropriate use of Stokes' law led Ehrenhaft to believe in the existence of subelectrons)	(1) Hertz's experiment on the deflection of cathode rays in an electric field. (Poor vacuum allowed an ionization process to take place which in turn interfered with the electric field)	<p>(1) The canals of Mars. (Physical and physiological limitations coupled with psychological expectation formed a condition for imagining canals)</p> <p>(2) Blondlot's N rays. ('The observer should look at the screen just as an impressionist painter would look at landscape')</p>	<p>(1) Franck's and Hertz's experiment on the quantized spectrum of the atom's energy levels. (Excitation potential was taken to be an ionization potential)</p> <p>(2) Herschel's discovery of the planet Uranus. (The observation was initially misinterpreted as a discovery of a new comet)</p>

argument which he or she then puts to test. As Wittgenstein remarked,

all testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments: no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life.⁵³

The background theory is indeed ‘the element in which arguments have their life’; it thus constitutes the first stage of the method of experimentation and therefore the first category in the present scheme of types of experimental error.

One may clarify this point by making the simple observation of looking at a family photograph. Having examined such a photograph, one instantly recognizes the images which comprise it and is able to relate them in a confident fashion to the relatives one has come to know. It seems that such an exercise does not require the slightest effort: there is no need for the observer to consider the mechanical, optical and chemical theories which underlie the production of such an artefact. However, one can take this exercise to its extremes and consider an electron-micrograph and a radio-photograph transmitted, say, by a spacecraft such as Mariner; the theory in the background which was previously set aside must come now to the fore. Since one cannot be sure about the relation between the unfamiliar composition of images to the so-called ‘reality’ this composition is supposed to represent, one has to rely upon the background theory to ensure that the composition is indeed a faithful representation.⁵⁴

Clearly, we can further distinguish within this class between the general theory and the instrumental theories which are assumed to govern the performance of the instruments intended for use in the experiment. Amongst these instrumental theories there is one theory which deserves a special attention, namely, the theory which underlies the set-up itself. This theory occupies, therefore, a subclass of its own in this category of background theory. It is in this primary stage of the experiment that the experimenter develops the theory of the set-up, plans its design and specifies the initial conditions.

The experimental method can be generally characterized as a method which inquires into the nature of phenomena by *varying* a certain group of elements or a single one, and recording the change, if any, in some other elements. In fact, as Mach remarks, ‘what we can learn from an experiment resides wholly and solely in the dependence or independence of the elements or conditions of

⁵³ Wittgenstein (1977), p. 16e (par. 105).

⁵⁴ Many difficulties may be raised with regard to the application of photographic techniques in scientific inquiries; prominent amongst them is the possibility of creating artefacts which do not represent faithfully the objects under study (e.g. problems in enhancement techniques). See also Hillman and Sartory (1977). For a positive view concerning electron-micrographs see Hacking (1983), pp. 186–209, Chap. 11: Microscopes.

a phenomenon'.⁵⁵ In studying a particular agent or cause, the experimenter should seek to isolate the object of study. Thus, the design must be such as to increase the effects due to the object of study until these effects exceed considerably the unavoidable concomitants; the latter can then be considered as only disturbing these effects, and not essentially modifying them.⁵⁶ In other words, the experimenter has to make the set-up most sensitive to the object of study and as insensitive as possible to all other elements that may play a part. As Weyl remarks, 'this accounts, among other things, for the tedious efforts involved in screening off all kinds of "sources of error"'.⁵⁷ Evidently, the attainment of isolation is essential for achieving this goal. Indeed, the mark of a great experimenter is the ability to simplify the set-up in such a way that only the factor in question remains in evidence while all other influences become negligible.⁵⁸

In the background-theory stage of the experiment error may arise as a consequence of using false theory. Corresponding to the distinctions between the general theory, the instrumental theories and the theory of the set-up, we may determine one possible source of error as the falsity of one or more of these theories. However, it should be stressed that a false theory does not necessarily lead to an error in the final experimental result. One may develop the argument of an experiment within the framework of a false general theory and yet produce a correct result; one may use an instrument or even the apparatus itself without a proper theoretical understanding of its working, indeed its theory may be wrong or even missing altogether, and yet the overall result still be correct. As C. Bernard summed up this point, 'even mistaken hypotheses and theories are of use in leading to discoveries. . . It seems, indeed, a necessary weakness of our mind to be able to reach truth only across a multitude of errors and obstacles'.⁵⁹

In addition to falsity as a source of error, error may occur in this stage when the theory is incomplete, or when it is correct in the main but false in some isolated aspect — a falsity which does not undermine the theory itself. Furthermore, in cases where the theory makes use of some external — either calculated or measured — constants, error may arise if these auxiliary elements are in error.

A case in which use was made of a falsely calculated constant is related by Wood. Mach reported that in a treatise entitled *Homes Without Hands*, published in 1865, Wood relates the following episode:

⁵⁵ Mach (1976), p. 149; Chap. 12: Physical Experiment and Its Leading Features.

⁵⁶ Thomson and Tait (1872), p. 108.

⁵⁷ Weyl (1963), p. 153.

⁵⁸ Mach (1976), p. 151. For further analyses of experiment see Franklin (1986) and Galison (1987).

⁵⁹ Bernard (1957), p. 170.

Maraldi had been struck with the great regularity of the cells of the honeycomb. He measured the angles of the . . . rhombs, that form the terminal walls of the cells, and found them to be respectively $109^{\circ}28'$ and $70^{\circ}32'$. Réaumur, convinced that these angles were in some way connected with the economy of the cells, requested the mathematician König to calculate the form of a hexagonal prism terminated by a pyramid composed of three equal . . . rhombs, which would give the greatest amount of space with a given amount of material. The answer was, that the angles should be $109^{\circ}26'$ and $70^{\circ}34'$. The difference, accordingly, was two minutes. Maclaurin, dissatisfied with this agreement, repeated Maraldi's measurements, found them correct, and discovered, in going over the calculation, an error in the logarithmic table employed by König. Not the bees, but the mathematicians were wrong, and the bees had helped to detect the error!⁶⁰

Mach went on to comment that

any one who is acquainted with the method of measuring crystals and has seen the cell of a honeycomb, with its rough and non-reflective surfaces, will question whether the measurement of such cells can be executed with a probable error of only two minutes.⁶¹

A clear example of an auxiliary measured-constant which introduced an error into the final experimental result is E. Harrington's value for the viscosity of air, of which R. A. Millikan made use in his evaluation of e , the charge of the electron. Millikan gave much weight to Harrington's result — Harrington was a colleague of Millikan — but the result was vitiated by several serious systematic errors.⁶² Harrington's value was too low by about 0.4% and consequently Millikan's value of e was too low by about 0.6%.⁶³ As E. R. Cohen *et al.* observe, this error 'remained completely unsuspected for a period of about 15 years [1916–1931]'.⁶⁴

An illustration of an incomplete theory which resulted in a discrepancy between theory and observation is the discrepancy between Newton's theoretical value for the speed of sound in air and actual measurement of that speed. T. S. Kuhn points out that 'this discrepancy, about twenty per cent, had been one of the scandals of physical science for more than a century and had repeatedly though fruitlessly drawn the attention of Europe's outstanding theoretical scientists, including Euler and Lagrange'.⁶⁵ When Laplace sug-

⁶⁰ Mach (1974), p. 548; cf. Fleck (1979), pp. 32–33.

⁶¹ Mach, *ibid.*, p. 549.

⁶² E. R. Cohen *et al.* (1957), p. 116. In Millikan's view, Harrington succeeded in making a determination of the viscosity of air which is 'altogether unique in its reliability and precision. I give to it alone', Millikan writes, 'greater weight than to all the other work of the past fifty years in this field taken together. For the individual determinations, though made with different suspensions and in such a way as to eliminate all constant sources of error save the dimensions of the cylinders, never differ among themselves by as much as .1 per cent, and the error in the final mean can scarcely be more than one part in 2000'. (Millikan, 1917, p. 9.)

⁶³ In the evaluation of the charge e , the coefficient of viscosity enters to the $3/2$ power.

⁶⁴ E. R. Cohen *et al.* (1957), p. 116.

⁶⁵ Kuhn (1958), pp. 136–137.

gested around 1802 that heating by compression might account for this discrepancy, direct confirmation was not possible. The existing thermometers had such a large heat capacity that their slow response did not permit direct measurements. Competing some ten years later for the French Academy prize concerning this problem, Delaroché and Bérard did measure the heat capacity of air at two quite different pressures. From these measurements Laplace was able to derive the first good theoretical value for the speed of sound in air. By 1825, Laplace succeeded in reconciling experiment and theory by using a new value for the speed of sound in addition to more direct data of adiabatic compression.⁶⁶ However, as Kuhn remarks, 'in retrospect, the agreement was artificial. . . In this case, as in remarkably many others throughout this period, errors of theory and experiment compensated more than well enough to satisfy expectation'.⁶⁷

Michelson's measurement of the velocity of light is an example of a correct set-up theory which contains an isolated mistake. In his 1941 review of the various determinations of c , the velocity of light in vacuum, Birge reported that

Michelson's published result of 299796 km/s is. . . in need of revision. To obtain it he used 67 km/s for the correction to vacuum from air at an assumed average temperature of 20°C and at an average barometric pressure of 625 mm Hg. It is easy to show that his correction results from the use of the *wave* index of refraction, instead of the correct *group* index of refraction.⁶⁸

Birge remarked that 'this is one of the most inexplicable errors that I have ever come across in the literature'.⁶⁹ According to Birge, Michelson never applied the concept of group velocity in his work on the absolute value of c ; indeed, 'Michelson did not even apply group velocity to his observed index for water. If he had, it would have spoiled the apparent agreement'.⁷⁰ Although Michel-

⁶⁶ *Ibid.*, pp. 136–139. Specifically, Laplace showed that if γ is defined as the ratio of the heat capacity of a gas at constant pressure to its capacity at constant volume, then the speed of sound must have Newton's value multiplied by the square root of γ .

⁶⁷ *Ibid.*, pp. 137–138. Kuhn remarks further that in the first half of the nineteenth century 'the heat capacity of calorimeter and thermometer was usually far larger than that of the gas they contained. As a result, few empirical bench marks were available to those who developed the caloric theory of gases, and it was difficult to tell which of the available and by no means consistent experimental measurements was reliable. . . The selection and evaluation of empirical tests was as much a matter of taste and judgement as the selection and evaluation of theory'. (*Ibid.*, p. 140). C. Truesdell disputes this historical analysis of Kuhn. According to Truesdell 'Lagrange found a suitable fudge factor in the 1760s, which would probably satisfy a modern physicist but was not acceptable then because it had no basis in principle. Failure to find a theory that explains from a general standpoint a datum of experiment is not a scandal; rather, it is an impetus to progress. Many aspects of acoustics are dimensionless (nodal patterns and ratios, for example); they were developed with great skill in the eighteenth century by Taylor, Daniel Bernoulli, Euler, and others'. (I am grateful to C. Truesdell for this remark; cf., Truesdell, 1980, p. 30; Fox 1971.)

⁶⁸ Birge (1941), p. 93.

⁶⁹ *Ibid.*

⁷⁰ *Ibid.*, p. 94.

son was aware of the distinction between group and wave velocity,⁷¹ he did not apply it correctly.

Ehrenhaft's measurements of the charge of the electron which 'demonstrated' the existence of subelectrons, provide an illustration for a false set-up theory that resulted in an error. Like Millikan, Ehrenhaft presupposed Stokes' law to govern the fall of the particles. However, he used metal particles which were much smaller than the oil drops of Millikan, not to mention their irregular surface in contrast to the smooth surface of the droplets. Consequently, Stokes' law was not applicable to the physical conditions Ehrenhaft had created, and his final result was in error.⁷²

Ehrenhaft declined to accept this objection and saw no reason why he should abandon Stokes' law in its original form. He claimed that if

the validity of the resistance law for small moving spheres of mercury and other material is to be doubted, then it cannot hold, either, for other substances, e.g. oil. The withdrawal of such a premiss would of course cut the ground away from the elementary quantum idea... as deduced from observations of electric charges on individual oil drops.⁷³

Ehrenhaft's objections to Millikan's work were logical and experimental; he argued that Millikan presupposed atomicity and invariably begged the question, and that Millikan's experimental technique had not been sufficiently sensitive to detect fractions of what was considered the fundamental unit of electric charge. Ehrenhaft claimed that he had always found charges smaller than the charge of the electron, and that 'this result does not depend upon any assumptions whatever as to atomicity of matter or electricity'.⁷⁴

Millikan, on his part, strongly advocated the view that Ehrenhaft's results 'mean simply that he has assumed an incorrect law of movement of his minute charged particles through a gas'.⁷⁵

An interesting case of a false background theory which actually prevented the execution of what may be considered a correct idea for an experiment, occurs in the history of the experimental research into the phenomenon of light-pressure. In a paper delivered to the French Academy in 1731, J. J. Dortous de Mairan rightly remarked that all direct experiments which had been designed to detect this phenomenon were bedevilled by effects which the heat produced by the light had caused.⁷⁶ Nevertheless, he did construct a sensitive pivoted mill of low friction which was extremely mobile, and trained on it the focus of a lens; but he still could not establish unambiguously the

⁷¹ *Ibid.*, pp. 93–94.

⁷² For a detailed study and bibliography see Hon (1985), pp. 191–209; Chap. 5: Case-Studies of Experiments.

⁷³ Ehrenhaft (1925), p. 639.

⁷⁴ Ehrenhaft (1940), p. 385.

⁷⁵ Millikan (1916b), p. 625.

⁷⁶ Worrall (1982), p. 142.

phenomenon of light-pressure. Indeed, as the light was focused on one of the vanes, the ‘machine turned sometimes to one side, sometimes to the other’.⁷⁷ Mairan concluded correctly that these results were due to the heating of the air around the machine. Naturally, the next step would have been to perform the same experiment in a vacuum. Mairan was explicitly aware of this possibility, but refrained from carrying it out, deciding that he ‘need not give [himself] the trouble’.⁷⁸ He was of the opinion that,

there is in our atmosphere amongst that more gross air which we breathe and which does not at all penetrate glass, another more subtle air of some other fluid which penetrates glass.⁷⁹

In Mairan’s view, this more subtle air would lead, in these light-pressure experiments, to the same uncertain results. Being misled by this theory, Mairan did not pursue further the correct idea of executing this experiment *in vacuo*.⁸⁰

The theory of the set-up may mislead in giving the impression that it governs different set-ups of the same type. A case in point is the experiments on the double scattering of electrons in the 1920s and thereafter. As Franklin and Howson point out,

the electrons used in these experiments came from both β -decay and thermionic sources and physicists of the day believed that the type of source used made no essential difference. Later work, however, showed that β -decay electrons were longitudinally polarized while thermionic electrons were unpolarized, which resulted in significantly different experimental results.⁸¹

However, theories can be irrelevant to the success of the experiment. Galileo, for example, did not have any theory to explicate the working of his telescope.⁸²

2.2. Assumptions Concerning the Actual Set-up and Its Working

Whereas the background-theory class focuses on the theoretical basis of an experiment, the second class is concerned with the actual process of setting up and operating the apparatus involved, that is, the hardware. One may characterize this second stage as the materialization of the theoretical requirements stipulated in the first stage; in other words, the nuts and bolts stage.

In the process of constructing and setting up the various hardware required by the experiment, the experimenter makes numerous assumptions. The most

⁷⁷ Quoted by Worrall, *ibid.*, p. 143.

⁷⁸ *Ibid.*

⁷⁹ *Ibid.*

⁸⁰ Worrall remarks that ‘Mairan seems here illicitly to be assuming, independently of experiment, that light does indeed exert a pressure — a fact in which, as a corpuscularist, he firmly believed. For, had he reperformed the experiment *in vacuo* and found the erstwhile movement *quelled* then it would follow that this “more subtle air” could play no role in the phenomenon’. (*Ibid.*)

⁸¹ Franklin and Howson (1984), pp. 60–61; cf. Franklin (1979), pp. 247–248.

⁸² Feyerabend (1979), Chaps. 9–11.

common of these is the assumption that the parts one has procured and their arrangement stand up to the required specifications and conditions. To be sure, it is possible to ascertain physically some of these assumptions by putting them to the test; however, since such tests involve further measurements, one has here a regressive sequence of measurements that for practical reasons must be truncated. Moreover, the experimenter has to ensure that the specifications are maintained while the experiment is carried out, or to allow for any change that may occur. Arising therefore in this class is the general type of error which originates in the belief that the set-up has met all the requirements, including the specified initial conditions, which, however, in point of fact it has not: the vacuum may be poor, the electric and magnetic fields may not be as uniform as expected, the insulation may not suffice, convection current or diffusion may occur, and so forth; the experimental result may thus be misleading.

The determination of the Hall effect which is of great interest in the theory of liquid metals abounds with such difficulties. An electric current is passed through a liquid metal (at a temperature of, say, 1000°C) in a strong transverse magnetic field. A very small voltage has then to be measured across the specimen. One can easily imagine the problems of containing the sample, maintaining it at a uniform temperature, avoiding convection currents and effects of magneto-hydrodynamic voltages, damping down vibrations, amplifying the signal, and so forth. As Ziman remarks,

it is scarcely surprising that a decade of experimental work by several very accomplished research workers has not produced an agreed set of data for this (in principle) basic and elementary physical parameter.⁸³

The study of the photoelectric effect presents another case. In 1899, J. J. Thomson successfully showed that the entities emitted in interactions of light with metal — including neutral metal plates — were the very same entities which made up the cathode rays.⁸⁴ This result firmly confirmed the assumption that the photoelectric effect was due to interaction of light with the atomic or subatomic constituents of the metal surface. It thus provided a definite framework within which work on this effect could be carried on. Further experimental work gave J. J. Thomson the impression that increase in temperature intensified the photoelectric action; the emission velocities of the ejected electrons appeared to be dependent on temperature. He therefore argued that the photoelectrons must be the free electrons of the metal as these particles partake in the energies of thermal agitation.⁸⁵

However, in 1907 Millikan and Winchester demonstrated conclusively that this result of J. J. Thomson, namely, that temperature affects the emission velocities of photoelectrons, was based on misleading experimental data: data

⁸³ Ziman (1978), p. 59.

⁸⁴ J. J. Thomson (1899); Humphreys (1968), p. 46.

⁸⁵ J. J. Thomson (1902); Humphreys, *ibid.*, pp. 47–48.

drawn from experiments conducted in air, rather than in a vacuum. Millikan and Winchester could not find in a careful survey of 11 metals any dependence of the photoelectric effect on temperature. This finding, coupled with Lenard's discovery that the emission velocities of photoelectrons are independent of the intensity of the incident light,⁸⁶ led Millikan and Winchester to conclude that their result 'constitutes very conclusive evidence that, if free electrons exist at all within metals, it is not these electrons which escape under the influence of ultra-violet light'.⁸⁷ In other words, J. J. Thomson's result was in error.

Another case in point is Hertz's experimental study of cathode rays. In 1883, H. Hertz conducted experiments on cathode rays and failed to realize that due to poor vacuum the electric field which he had applied to a beam of cathode rays did not attain sufficient intensity. As a result, a distinct deflection of the cathode rays could not be detected. Hertz concluded his pioneering experiment by stating that

the electrostatic and electromagnetic properties of the cathode rays are either *nil* or very feeble.⁸⁸

It is ironic that the prototype of the oscilloscope — for that is what Hertz's apparatus amounted to — should be instrumental in demonstrating that cathode rays have no closer relation to electricity than has light produced by an electric lamp. Indeed, Hertz argued that since 'cathode rays are electrically indifferent, . . . the phenomenon most nearly allied to them is light'.⁸⁹

It took almost 15 years to show that Hertz was in error. When J. J. Thomson evacuated a cathode-ray tube more carefully and efficiently, he could see quite clearly — as the pressure was being reduced — how an electric field can deflect a beam of cathode rays; a deflection which earlier had escaped Hertz. The discovery of the electron, as is well known, took place with the advent of this experiment of Thomson.⁹⁰

N. de Lacaille (1713–1762), the meticulous French astronomer who mapped the southern skies, was troubled by the fact that the result of his geodetic survey in the vicinity of Cape Town supported the hypothesis that the earth is a prolate, not an oblate, spheroid. Although he partially rechecked the result, he could find no error and it remained a puzzle for some years. Apparently the result was due to the deviation of the plumb line at his southern station caused by the large mass of Table Mountain.⁹¹

In their 1965 review of the fundamental constants, Cohen and DuMond point out that there is a 17 km/s difference between the general region of values

⁸⁶ Lenard (1900, 1902).

⁸⁷ Millikan and Winchester (1907), p. 198; Humphreys (1968), p. 48.

⁸⁸ Hertz (1896), p. 254. For a detailed study see Hon (1987a).

⁸⁹ Hertz, *ibid.*

⁹⁰ J. J. Thomson (1897).

⁹¹ Gingerich (1973), p. 544. I am grateful to B. R. Goldstein for drawing my attention to this case.

obtained for c and the older result adopted in 1941 by R. T. Birge. The case of the 17 km/s discrepancy between the old Birge-recommended value of c and the newer values obtained by more modern methods of measurement has been described by Birge himself as one of the most astonishing systematic errors in the history of physics. Cohen and DuMond remark that

the chief source of error in the older estimate came from a systematic error in the experimental result of Michelson, Pease and Pearson, performed in a mile-long evacuated tube laid on unstable soil near Santa Ana, California. Although some 2885 replicated observations of the time of flight were made, there were only a few (2 or 3) determinations of the distance. All the time measurements were made at night and the distances in the daylight. The site was near the ocean and variations in the results which seem correlated with the tides were observed.⁹²

Yet, no one knows for sure what caused the systematic error in this work.

2.3. Observational Reports

The preceding two stages constitute the foundation — in the abstract and the concrete sense — upon which the empirical programme can be brought to its completion, that is, knowledge through sense perception. The central feature of the third stage is thus the process of observation. It is at this stage that the contact, as it were, between the senses and the allegedly isolated ‘world’ of the experiment is explicitly made. This contact may be at fault due to either misinterpretation of what has been otherwise correctly observed, or the limitation of the senses and their subjective nature. I call the former possible fault observational error, and the latter, for an historical reason — personal equation.

An observational error occurs when the observer makes a correct observation but fails to bring into consideration the external circumstances in which the observation took place; such a failure arises from misinterpreting the correct observation. These errors usually occur when the observer encounters a misleading situation of which he or she is not aware, and proceeds to interpret the otherwise correct observation without allowing for the misleading circumstances.

For example, were the experimenter to make a pointer reading, and should he or she ignore the gap between the pointer and the scale, then the reading might be in error due to the phenomenon of parallax. Another case in point is refraction: if one were to disregard this phenomenon, especially in astronomical observations, one would not succeed in locating the actual position of the object one is observing. However accurate the observation may be, one would inevitably determine the *apparent* position of the object concerned. In my terminology it would be an observational error if the observer were to consider the result of this observation the *true* position of the object.

⁹² E. R. Cohen and J. W. M. DuMond (1965), p. 550.

In the subclass of personal equation I subsume, in contrast to observational errors, those errors which originate in the limitation of the senses and their subjective nature. It was F. W. Bessel (1784–1846) — the famous mathematician and astronomer, the pioneer in the more exact measurements of modern astronomy — who conceived the notion of personal equation. Working in a period of great interest in errors of astronomical observation and their mathematical treatment,⁹³ Bessel perceived, in 1816, the importance of the Kinnebrooke incident, and published an analysis of the case in 1822 — a publication which attracted immediate attention.

The Kinnebrooke incident took place at Greenwich Observatory where, in January, 1796, the Astronomer Royal, N. Maskelyne (1732–1811), dismissed his assistant, Kinnebrooke, for observing the times of stellar transits almost a second later than he himself had done. It should be stressed that the error was considered serious, since the calibration of the Greenwich clock depended upon such observations.⁹⁴

This event was recorded in the pages of the journal *Astronomical Observations at Greenwich*,⁹⁵ and might have passed into oblivion had it not been for Bessel who saw its significance. Having studied the case, Bessel set himself the task of determining whether or not such a difference could be found amongst observers with more experience than Kinnebrooke. To his astonishment he found great differences which he presented in the form of the equation $A - B = x$ sec. The difference between observer A and observer B was called the 'personal equation'.⁹⁶ It was thus realized that even the most expert observers must allow themselves to be corrected by statistical averages from other observers. As R. L. Gregory points out,

⁹³ By 1816, both Laplace and Gauss had already published their seminal works on the mathematical treatment of errors of observation.

⁹⁴ Maskelyne followed the Bradley 'eye and ear' method which combines the audible beats of the pendulum clock with the perceptible transit of a given star across the hair-line marking the meridian. In this method, as E. G. Boring reports, 'the observer looked at the clock, noted the time to a second, began counting seconds with the heard beats of the clock, watched the star cross the field of the telescope, noted and "fixed in mind" its position at the beat of the clock just before it came to the critical wire, noted its position at the next beat after it had crossed the wire, estimated the place of the wire between the two positions in tenths of the total distance between the positions, and added these tenths of a second to the time in seconds that he had counted for the beat before the wire was reached. It is obviously a complex judgement', Boring remarks. 'Not only does it involve a coordination between the eye and the ear, but it requires a spatial judgement dependent upon a fixed position (the wire), an actual but instantaneous position of a moving object, and a remembered position no longer actual. Nevertheless, "the excellent method of Bradley" was accepted and regarded as accurate to one or at least two tenths of a second. In the face of this belief', Boring concludes, 'Kinnebrooke's error of eight tenths of a second was a gross error and justified Maskelyne's conclusion that he had fallen "into some irregular and confused method of his own" and his consequent dismissal'. (Boring, 1950, p. 135; cf., Gregory, 1984, p. 212).

⁹⁵ Maskelyne (1799).

⁹⁶ For Bessel's discovery, see Bessel (1823, 1826, 1836, 1876); cf. Boring (1950), pp. 136–137, 150 (Chap. 8: The Personal Equation); Gregory (1984), pp. 210–216.

for the first time in science, the average took over to correct the individual testimony of expert observers.⁹⁷

By comparing directly and indirectly his observations with those of Struve, Bessel went further and showed that one could not reliably 'calibrate the observer' — the 'personal equation' being itself variable.⁹⁸ Yet, the variability of the 'personal equation' was not so great as to render correction entirely useless. Bessel, for example, always observed in advance of Struve by an amount varying between 0.770 and 1.021 seconds.⁹⁹

During the 50 years since Bessel's discovery of the phenomenon of 'personal equation', astronomers were much concerned with this problem and sought methods of correction and elimination. As it turned out, the emerging new discipline of experimental psychology gradually took hold of this issue, focusing its attention on explanation rather than on correction or elimination. Hermann von Helmholtz's (1821–1894) measurement of the velocity of the nervous impulse in 1850, which demonstrated that nervous conduction is considerably slower than the speed of sound in air,¹⁰⁰ and J. Hartmann's (1814–1876) experiment in 1858, which showed expectation and anticipation to be decisive elements of the 'personal equation',¹⁰¹ gave much impetus to the study of this problem from physiological and psychological standpoints respectively. Furthermore, the development of the electromagnetic circuit made at that time the construction of chronographs practicable, and it became possible to determine what is called the 'absolute personal equation': the reaction time in making a movement as rapidly as possible after perceiving a signal with respect to a standard time measure, and not relative to the reaction time of another observer.¹⁰²

The sources of error of the personal-equation type reside therefore either in the physiological condition of the observer's senses, or in the observer's mental state, or both. Examples of the former sources of error are colour blindness and slow reaction time, and of the latter — anticipation and hallucination.

The distinction between observational errors and errors of the personal-equation type reflects the categorical differences between causes from without

⁹⁷ Gregory, *ibid.*, p. 213.

⁹⁸ Boring (1950), p. 137.

⁹⁹ *Ibid.*, p. 138. Duhem remarked that 'astronomers try to determine... [the acuteness of the observer's senses] in the mathematical form of a personal equation, but this equation partakes very little of the serene constancy of geometry, for it is at the mercy of a splitting headache or painful indigestion'. (Duhem, 1974, p. 162.)

¹⁰⁰ Helmholtz's values are 42.9 and 25.0 m/s; they lie well within the range as determined by modern methods. Recent determinations give values for the velocity of the nervous impulse as high as 120 m/s for the largest nerve fibres and as low as 1 m/s for the smallest. For comparison, the speed of sound in air is about 330 m/s. (Boring, 1950, pp. 48, 144.)

¹⁰¹ Hartmann's contribution was very important since it demonstrated that the 'personal equation' is not simply a variable delay which can be explained physiologically, but that it also depends upon psychological elements. (*Ibid.*, p. 145; Gregory, 1984, pp. 215–216.)

¹⁰² Boring, *ibid.*, pp. 140–141, 147; Gregory, *ibid.*, p. 215.

and causes from within. External circumstances can give rise to the possibility of committing an observational error, whereas one's own physiological condition and psychological make-up determine one's personal equation.

A historical case which clearly demonstrates the problem of psychological expectation and the limitation of the senses — both possible sources of error pertain to the subclass of personal equation — is the case of the canals of Mars. There is no item in the entire history of the telescopic observations of Mars which has been more widely proclaimed, more vehemently debated, and more abruptly forgotten than the canals.

The first astronomer to observe this phenomenon, Schiaparelli of Milan, noted in 1877 that 'there are on this planet, traversing the continents, long dark lines which may be designated as *canali*. . . Their arrangement appears to be invariable and permanent'.¹⁰³

The possibility that the intersecting dark lines which Schiaparelli had observed were literally canals, built by intelligent beings, was immediately taken up; it gave weight to the belief that there exists a highly developed civilization on Mars. One of the most energetic proponents of this theory was the American P. Lowell (1855–1916), who argued that the 'intelligent creatures' of Mars were living on an ageing, desert-like, planet whose water was trapped in the polar caps; hence the canals: an artificial device to transport water into the equatorial zones.¹⁰⁴

Lowell admitted that it is very difficult to see the canals. 'Success', he wrote, depends on the acuteness of the observer's eye and upon the persistence with which he watches for the best moments in the steadier air.¹⁰⁵

Thus, he continued,

not everybody can see these delicate features at first sight, even when pointed out to them; and to perceive their more minute details takes a trained as well as an acute eye, observing under the best conditions. When so viewed, however, the disk of the planet takes on a most singular appearance. It looks as if it had been cobwebbed all over.¹⁰⁶

Since it appeared that the whole planet was encompassed by the canals, Lowell suggested that the Martian community could act as a unit throughout its globe; he therefore argued that the community on Mars had an intelligent and non-belligose character.¹⁰⁷

However, the critics argued that even an unbiased human eye would connect, under conditions of poor resolution, discontinuous blotches and

¹⁰³ Quoted by Mutch *et al.* (1976), p. 21.

¹⁰⁴ *Ibid.*

¹⁰⁵ Quoted by Mutch *et al.*, p. 23.

¹⁰⁶ *Ibid.* A typical observational report on the canals of Mars can be found in: *Nature* **84**, (1910), 172–173.

¹⁰⁷ Mutch *et al.*, *ibid.*, pp. 23–24.

streaks into regular straight lines; and, secondly, since the effective resolution of the best telescopes is about 100 km on the Martian surface, Lowell's claim to identifying canals of 50 km and less in width is simply not compatible with optical principles. The critics did not deny the visual effect of dark markings on the image of Mars; they objected, however, to identifying them, as narrow continuous lines, with actual Martian surface markings.¹⁰⁸

As late as 1962, Lowell Observatory published a photographic atlas of Mars, claiming to provide 'each reader with the best possible opportunity to distinguish the fine lines and to judge their reality [*sic*] for himself'.¹⁰⁹

When the Mariner pictures arrived, it became clear that the critics' point of view has been vindicated. The pictures obtained by the high resolution cameras aboard the spacecraft made it amply clear that no regular, linear elements of the size required by the canals theory were observable. It appears that poor resolution initiated a condition which facilitated a visual synthesis of discontinuous elements; the interpretation of these elements as canals was supported in turn by the belief that there is a Martian community. Thus, the observation of the canals was in error due to physical-physiological limitation as well as psychological expectation.¹¹⁰

Another interesting case which belongs to this category is the 'discovery' of the N rays by R. Blondlot. Blondlot, a well-known French physicist working at Nancy University, claimed to have discovered in 1903 a new form of radiation which he called N rays, after his university. The original method of detecting the N rays was to observe an electric spark which allegedly got brighter when subject to this radiation. Blondlot was confident that he could distinguish visually between the brightness of the sparks. Later on, he developed a special calcium sulphide screen whose phosphorescent glow increased markedly, so it was claimed, when it was exposed to N rays. Although the production of this effect and its cessation were not instantaneous, Blondlot claimed that amongst all the actions produced by N rays, this was the most easily observed.¹¹¹

Blondlot and his co-workers at Nancy established that there were many sources of N rays: artificial and natural, and they diligently proceeded in the most efficient manner to examine the properties of this new radiation.

By the summer of 1903 the discovery of this new form of radiation became a major issue and many attempts were made to reproduce it. However, except in France, all the attempts to reproduce Blondlot's results failed.

Blondlot argued that the N-ray phenomena

lie almost at the limit of what... [the observers] are able to discern, and it is only after a certain amount of practice that they succeed in catching them easily, and in

¹⁰⁸ *Ibid.*, p. 24.

¹⁰⁹ Quoted by Mutch *et al.*, *ibid.*

¹¹⁰ *Ibid.*, p. 25.

¹¹¹ Blondlot (1905), pp. 22-23. For a detailed study and bibliography see Hon (1985), pp. 226-252; Chap. 5: Case-Studies of Experiments; cf., M. Jo Nye (1980).

observing them with complete certainty. The smallness of the effects and the delicacy of their observation must not deter us from a study which puts us in possession of radiations hitherto unknown.¹¹²

As more criticism mounted, the *Revue Scientifique* polled leading scientists on their judgements concerning the existence of N rays. The editorial board announced that 'it is time for French science to settle this question definitely'.¹¹³

The strongest, and as it happened fatal, objection came from the American physicist, R. W. Wood. He paid a visit to Blondlot at his laboratory in Nancy, where he exposed Blondlot and his co-workers as the victims of autosuggestion.¹¹⁴

In reply, Blondlot affirmed that 'the phenomena of N rays have for me the same certainty that other physical phenomena have. Several of my colleagues and a number of other persons say the same'.¹¹⁵ In his view,

the observer should accustom himself to look at the screen just as a painter, and in particular an 'impressionist' painter, would look at landscape. To attain this requires some practice, and is not an easy task.¹¹⁶

In fact, according to Blondlot some people never succeed in attaining these requirements.

It is misleading to say with the critics of Blondlot that he was wrong in that the ultimate test was subjective. By definition, an observation is a subjective process; the problem lay rather in the nature of the observation: it was simply unamenable to objective, or rather inter-subjective criteria such as pointer reading or measurement of the length of a spark. To observe the N-ray effects one had to look for a small change in the brightness of a spark which needless to say required remembering brightness — a very imaginative task indeed.

2.4. Theoretical Conclusions

Having carried out the observations and the concomitant measurements, the experimenter arrives at the final stage of the study. To conclude the enquiry, that is, to obtain the final result, the experimenter analyses the data, and then incorporates the results of the analysis with an external theory so that the result can be put to use in either substantiating or refuting a theoretical claim, or laying an altogether new foundation. This is the goal of the experiment.

Two processes are therefore involved in this final stage: firstly, there is the process of reduction, of analysing the data with a view to obtaining a coherent

¹¹² Blondlot, *ibid.*, p. 37.

¹¹³ *Revue Scientifique* 2 ser. 5 (1904), 552.

¹¹⁴ Wood (1904). For another case of autosuggestion see Chadwick's exposé of Pettersson in the controversy over artificial disintegration. (Stuewer, 1985, especially pp. 284–289.)

¹¹⁵ Blondlot (1904), p. 621.

¹¹⁶ Blondlot (1905), pp. 82–83.

result; and, secondly, the process of spelling out the physical meaning of this result.

It is in the first process that the mathematical theory of error is employed; indeed, it is *only* in the process of reduction that the mathematical procedures for treating errors come to the fore. And since the distinction between systematic and random errors is central to this analysis of error, it is here that this dichotomy is directly applicable. In this stage error may arise due to the inapplicability of the assigned distribution of random errors which is presupposed in the mathematical analysis. Error may further occur in the process of assessing systematic errors — a process for which no clear set of rules exists.

A fault in the second process, that is, the process of rendering the final result physically meaningful, may be due to two distinct causes. Such a fault may originate either in misunderstanding the physical result the experimenter has obtained, or in fallacious reasoning. The former fault is an error of interpretation, whereas the latter is a logical mistake. In this subclass there are therefore sources of error and mistake which are located in the faculties of understanding and reasoning respectively.

The discovery of the planet Uranus presents a clear illustration of this category of experimental error. On the night of the 13 March 1781, William Herschel noticed, in the course of a search for star pairs, a curious object which appeared to be in motion relative to nearby stars. He recorded in his notes that he had seen 'a curious either nebulous star or perhaps a comet'. A few nights later he stated that what he had found 'is a comet, for it has changed its place'. The possibility that the object was a planet — an interpretation which was compatible with the visual evidence — did not occur to Herschel initially. On the 26 April, Herschel reported his discovery to the Royal Society in a paper entitled 'Account of a Comet'.¹¹⁷

Notified of this discovery, astronomers attempted to determine the motion of the new body. Several attempts were made to compute a suitable parabolic (cometary) orbit as soon as a few observations had been accumulated. However, it was quickly realized that each set of elements coincided with the observed positions for a few days and then rapidly diverged. The wasted efforts were due to one single reason: an uncritical acceptance of Herschel's interpretation that the object was a comet. It was in the summer of that year that Herschel's interpretation was rejected and the object he had discovered was reinterpreted as a planet: the planet Uranus.¹¹⁸

The few observations of Uranus did not suffice for accurate computations of its orbit. Seeking more observations, it occurred to Bode that Uranus might have been considered a star in possible earlier sightings. He therefore scrutinized old astronomical catalogues with a view to finding Uranus recorded as a

¹¹⁷ Quoted by Grosser (1979), p. 19.

¹¹⁸ *Ibid.*, pp. 20–21.

star. He was indeed rewarded; he found two such sightings: in 1690 by Flamsteed and in 1756 by Mayer. In the following years 16 more 'preplanetary' observations of Uranus were discovered.¹¹⁹

The planet Neptune was sighted for the first time as a planet by Galle who, on the basis of the predicted elements computed by Leverrier, searched for it in the right region in September 1846. However, Neptune had in fact been sighted by Lalande some 50 years earlier, but he had considered it a star. An examination of Lalande's manuscript showed that he had made two observations of the planet: on the 8 and 10 May 1795. Finding these two observations discordant, he had rejected one as probably in error, and marked the other as questionable. It is not unlikely that a re-examination of the region to see which observation was in error would have led Lalande to the discovery of the planet Neptune half a century before it was actually recognized as a planet. Indeed, the difference between the positions of these two sightings is given accurately by the now-known elements of Neptune's orbit.¹²⁰

A subtle case of error of interpretation occurred in the experimental confirmation of Einstein's equation of the photoelectric effect. In his presentation speech of Einstein's Nobel Prize, Arrhenius made it clear that the Nobel Committee for Physics had chosen Einstein especially for his contributions to the quantum theory: his studies of specific heat and photoelectric effect.¹²¹ Arrhenius declared in his speech that

Einstein's law of the photo-electrical effect has been extremely rigorously tested by the American Millikan and his pupils and passed the test brilliantly. Owing to these studies by Einstein the quantum theory has been perfected to a high degree.¹²²

A year later, in 1923, when Millikan was awarded the Nobel Prize 'for his work on the elementary charge of electricity and on the photoelectric effect',¹²³ the chairman of the Nobel Committee expressed the view that if Millikan's studies of the photoelectric effect had given a different result, the photoelectric law of Einstein would have been without value; the chairman pointed out that Einstein received the prize after Millikan had confirmed the law experimentally.¹²⁴

The results of Millikan's painstaking experimental work on the photoelectric effect¹²⁵ did indeed establish the validity of Einstein's equation, and, moreover, provided an accurate determination of Planck's constant h . However, Millikan categorically rejected Einstein's hypothesis of a light corpuscle of energy $h\nu$. In

¹¹⁹ *Ibid.*, pp. 24–26, 40–41.

¹²⁰ *Ibid.*, pp. 115–117, 130–140; cf., Polanyi (1964), pp. 90–91, Humphreys (1968), pp. 34–42.

¹²¹ *Nobel Lectures* (1967), p. 479.

¹²² *Ibid.*, p. 480. On the deliberation of the Nobel Committee see Pais (1983), pp. 502–512.

¹²³ *Nobel Lectures* (1965), p. 49.

¹²⁴ *Ibid.*, p. 53.

¹²⁵ Millikan described the experimental arrangement as 'a machine shop in vacuo'. (Millikan, 1916a, p. 361.)

his concluding paper on this work, Millikan opined that Einstein had put forward 'the bold, not to say the reckless, hypothesis of an electro-magnetic light corpuscle of energy $h\nu$ '.¹²⁶ Millikan considered this hypothesis 'reckless' since, as he explained,

an electromagnetic disturbance which remains localized in space seems a violation of the very conception of an electromagnetic disturbance, and second because it flies in the face of the thoroughly established facts of interference.¹²⁷

Although his experiments confirmed Einstein's equation for the photoelectric effect, Millikan felt strongly that 'the semi-corpuscular theory by which Einstein arrived at his equation seems at present to be wholly untenable'.¹²⁸ Indeed, according to Millikan, despite the

complete success of the Einstein equation, the physical theory of which it was designed to be the symbolic expression is found so untenable that Einstein himself, I believe, no longer holds to it.¹²⁹

By putting a different interpretation on Einstein's confirmed equation,¹³⁰ Millikan found himself in a peculiar situation; as he remarked,

we are in the position of having built a very perfect structure and then knocked out entirely the underpinning without causing the building to fall. It stands complete and apparently well tested, but without any visible means of support. These supports must obviously exist, and the most fascinating problem of modern physics is to find them. Experiment has outrun theory, or, better, guided by erroneous theory, it has discovered relationships which seem to be of the greatest interest and importance, but the reasons for them are as yet not at all understood.¹³¹

The scepticism concerning Einstein's light quantum hypothesis prevailed till about 1924. It was the discovery of the Compton effect that provided, together with the photoelectric effect, that 'interlocking theoretical and experimental matrix',¹³² from which a concept such as the photon derives its validity.

¹²⁶ *Ibid.*, p. 355.

¹²⁷ *Ibid.*

¹²⁸ *Ibid.*, p. 383.

¹²⁹ *Ibid.*, p. 384. Pais traced this belief of Millikan to a remark Einstein had made in 1911 at the first Solvay congress. 'I insist on the provisional character of this concept [light-quanta]', Einstein is reported to have said, 'which does not seem reconcilable with the experimentally verified consequences of the wave theory.' (Quoted by Pais, 1983, p. 383.)

¹³⁰ Millikan conceded that 'the photoelectric effect... however it is interpreted, if only it is correctly described by Einstein's equation, furnishes a proof which is quite independent of the facts of black-body radiation of the correctness of the fundamental assumption of the quantum theory, namely, the assumption of a discontinuous or explosive emission of the energy absorbed by the electronic constituents of atoms from ether waves. It materializes, so to speak, the quantity " h " discovered by Planck through the study of black-body radiation and gives us a confidence inspired by no other type of phenomenon that the primary physical conception underlying Planck's work corresponds to reality'. (Millikan, 1916a, p. 385.) However, Millikan went on to develop a substitute for Einstein's theory based largely on Planck's theory. (*Ibid.*, pp. 385-388; Millikan, 1922, pp. 231-238.) An overview of the various interpretations of this effect can be found in Humphreys (1968), pp. 43-59; cf., Stuewer (1970).

¹³¹ Millikan (1922), p. 230.

¹³² Stuewer (1970), p. 263.

In these cases the experimental and, for that matter, the observational data were correct and indeed significant. However, according to current theories the final results were in error due to erroneous interpretations of the data.¹³³

3. Concluding Remarks

In this typology of experimental error I have distinguished and ordered according to an *ideal* chronology four different stages which are involved in executing an experiment. I have identified in each stage — a stage being in itself a category in the classification — possible sources of mistake and error that may impair the final result. The heuristic value of this typology, that is, the potentiality of this classification to elucidate experimental errors, has been illustrated by applying the classification to actual experiments in the history of physics.

It is important to stress that underlying the distinctions embedded in the proposed classification there exist some outstanding philosophical issues. For example, the distinction between the category of background theory and the observational-reports category, is based on the assumption that a report of observation can be dissociated from any theoretical consideration: that, in principle, it is possible to have a description of a state of affairs which is not bound by any theory. This assumption is very problematic; indeed, it has been challenged by many students of philosophy of science. Hanson, for example, writes that 'there is a sense. . . in which seeing is a "theory-laden" undertaking. Observation of x is shaped by prior knowledge of x '.¹³⁴ According to this view, an observational report necessarily contains an interpretation of what is seen; it is thus laden with a theory about the nature and explanation of what is seen.

Richard L. Gregory, who studies perception, subscribes to the views of Hanson, I. Lakatos and T. S. Kuhn in so far as they hold that observations and experimental tests are highly theory-laden. He is of the opinion that 'a great deal has to be accepted before observations can be made at all'.¹³⁵ According to Gregory,

the fact that vision is usually sufficient for immediate object identification distracts us from realizing the immense importance of contextual knowledge for reading data from signals. Scientific data from instruments are almost always presented with explicit collateral information on how the instrument was used, what source it was directed to, its calibration corrections and scale settings.¹³⁶

P. Feyerabend has argued from a different angle that the distinction between observational terms and theoretical terms is defunct. Although experience, as Feyerabend wants us to believe, has been discovered to arise '*together with*

¹³³ For a detailed case-study of an error of interpretation see Hon, forthcoming.

¹³⁴ Hanson (1979), p. 19.

¹³⁵ Gregory (1984), p. 255.

¹³⁶ *Ibid.*, p. 397.

theoretical assumptions *not* before them', and that 'an experience without theory is just as incomprehensible as is (allegedly) a theory without experience',¹³⁷ still, Feyerabend remarks, 'the inference that the distinction between theory and observation has now ceased to be relevant, is either not drawn or is explicitly rejected'.¹³⁸ Feyerabend does not deny that such a distinction can be made; he simply does not see its purpose. As he is of the opinion that the distinction between theory and observation has 'definitely lost its point',¹³⁹ he encourages us to take a step forward and 'abandon this last trace of dogmatism in science'.¹⁴⁰

However, in I. Hacking's view, all that Feyerabend has done is to show us 'how not to talk about observation, speech, theory, habits, or reporting'.¹⁴¹ Hacking in fact dismisses the claim that observational reports always contain or assert theoretical assumptions, as 'hardly worth debating because it is obviously false, unless one attaches a quite attenuated sense to the words, in which case the assertion is true but trivial'.¹⁴² He then puts forward a few examples from the history of science which he believes give substance to the view that observations, indeed important observations, have been made without any theoretical assumptions.¹⁴³

We should not be drawn here into this controversy;¹⁴⁴ suffice it to note that there is a very great difference between asserting that all observational reports are 'convention-laden' and asserting that all observational reports are 'theory-laden'. Following P. Alexander, I claim that in contrast to the former, the latter can be denied; consequently, it is possible to have observational reports which are neutral as between theories in general, and as between scientific theories in particular. All uses of language are necessarily 'convention-laden', but we can distinguish, within any set of conventions, between what is merely 'convention-laden' and what is also 'theory-laden', so that descriptions which fall into the former class are, so to speak, pure, and that the distinction between description and interpretation, in at least one clear sense of those words, can be retained.¹⁴⁵

The critic of the method of experimentation — the philosopher of science who looks for possible faults in this method — must attain an external vantage point from which a critical outlook can be exercised. Distinguishing categories

¹³⁷ Feyerabend (1979), p. 168 (emphases in the original).

¹³⁸ *Ibid.*, p. 169.

¹³⁹ *Ibid.*, p. 168.

¹⁴⁰ *Ibid.*, p. 169.

¹⁴¹ Hacking (1983), p. 175.

¹⁴² *Ibid.*, p. 174.

¹⁴³ *Ibid.*, pp. 176–179. Hacking discusses this issue in his analysis of observation. (*Ibid.*, Chap. 10: Observation; see also pp. 155–156.)

¹⁴⁴ For a detailed philosophical discussion see Hesse (1974), Chap. 1: Theory and Observation; cf., Gooding (1982), especially pp. 46–48, and footnote no. 45.

¹⁴⁵ Alexander (1963), pp. 88–89.

in the process of executing an experiment and identifying in each of these categories typical sources of mistake and error is one way of carrying out a critique of the method of experimentation.

Acknowledgements — I wish to thank Heinz Post and Allan Franklin for helpful critical comments. An earlier draft of this paper was delivered at Oxford in 1982 to a BSPS conference on experimentation.

References

- Achinstein, P. and Hannaway, O. (eds) (1985), *Observation, Experiment, and Hypothesis in Modern Physical Science* (Cambridge, Mass.).
- Alexander, P. (1963), *Sensationalism and Scientific Explanation* (London: Routledge & Kegan Paul).
- Barnes, R. Bowling and Silverman, S. (1934), 'Brownian Motion as a Natural Limit to all Measuring Processes', *Review of Modern Physics* **6**, 162–192.
- Bernard, C. (1957), *An Introduction to the Study of Experimental Medicine* (New York: Dover).
- Bessel, F. W. (1823), *Astronomische Beobachtungen in Königsberg* **8**, iii–viii (for the year 1822); **11**, iv (for the year 1825); 1836, **18**, iii (for the year 1832); 1876, *Abhandlungen* **III**, 300–304.
- Birge, R. T. (1941), 'The General Physical Constants as of August 1941 with Details on the Velocity of Light Only', *Reports on Progress in Physics* **8**, 90–134.
- Blondlot, R. (1904), 'Les Rayons N existent-ils? Déclarations de M. Blondlot', *Revue Scientifique* **2**, 621.
- Blondlot, R. (1905), '*N* Rays, A Collection of Papers Communicated to the Academy of Sciences' (London: Longmans).
- Boring, E. G. (1950), *A History of Experimental Psychology* (2nd edn) (New York: Appleton-Century-Crofts).
- Born, M. (1956), 'Is Classical Mechanics in Fact Deterministic?', in: *Physics in My Generation* (London: Pergamon Press), pp. 164–170.
- Campion, P. J., Burns, J. E. and Williams, A. (1980), *A Code of Practice for the Detailed Statement of Accuracy* (London: National Physical Laboratory).
- Cartwright, N. (1984), *How the Laws of Physics Lie* (Oxford).
- Chwistek, L. (1948), *The Limits of Science* (London: Kegan Paul, Trench & Trubner).
- Cohen, E. R., Crowe, K. M. and DuMond, J. W. M. (1957), *The Fundamental Constants of Physics* (New York: Interscience Publishers).
- Cohen, E. R. and DuMond, J. W. M. (1965), 'Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965', *Review of Modern Physics* **37**, 537–594.
- Cohen, M. R. (1949), *Studies in Philosophy of Science* (New York; Frederick Ungar).
- Cramér, H. (1966), *The Elements of Probability Theory* (New York: Wiley, Science Edition).
- Crombie, A. C. (ed.) (1959), *Turning Points in Physics* (Amsterdam: North-Holland).
- Cusanus, N. (1981), *On Learned Ignorance*, a translation and an appraisal of *De docta ignorantia*, by J. Hopkins (Minneapolis: Arthur Banning Press).
- Duhem, P. (1974), *The Aim and Structure of Physical Theory* (2nd edn) (New York: Atheneum).

- Ehrenhaft, F. (1925), 'The Electrical Behaviour of Radioactive Colloidal Particles of the Order of 10^{-5} cm as Observed Separately in a Gas', *Philosophical Magazine* **49**, 633–648.
- Ehrenhaft, F. (1940), 'Physical and Astronomical Information Concerning Particles of the Order of Magnitude of the Wave-length of Light', *Journal of the Franklin Institute* **230**, 381–393.
- Einstein, A. (1921), 'Geometry and Experience', in: (1981), *Ideas and Opinions* (2nd edn) (Dell, New York: Laurel-leaf Library), pp. 227–240.
- Feigl, H. and Maxwell, G. (eds) (1961), *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart & Winston).
- Feyerabend, P. (1979), *Against Method* (2nd edn) (London: Verso).
- Fleck, L. (1979), *Genesis and Development of a Scientific Fact* (Chicago: University of Chicago Press).
- Fox, R. (1971), *The Caloric Theory of Gases: from Lavoisier to Regnault* (Oxford: Clarendon Press).
- Franklin, A. (1979), 'The Discovery and Nondiscovery of Parity Non-conservation', *Studies in History and Philosophy of Science* **10**, 201–257.
- Franklin, A. (1981), 'What Makes a "Good" Experiment?', *British Journal for the Philosophy of Science* **32**, 367–374.
- Franklin, A. (1986), *The Neglect of Experiment* (Cambridge: Cambridge University Press).
- Franklin, A. and Howson, C. (1984), 'Why do Scientists Prefer to Vary their Experiments?', *Studies in History and Philosophy of Science* **15**, 51–62.
- Franklin, A. and Howson, C. (1988), 'It Probably Is A Valid Experimental Result: A Bayesian Approach to the Epistemology of Experiment', *Studies in History and Philosophy of Science* **19**, 419–427.
- Galison, P. (1985), 'Bubble Chambers and the Experimental Workplace', in: Achinstein and Hannaway (1985), pp. 309–373.
- Galison, P. (1987), *How Experiments End* (Chicago: University of Chicago Press).
- Gingerich, O. (1973), 'Lacaille, Nicolas-Louis De', in: *Dictionary of Scientific Biography*, C. C. Gillispie (ed.) (New York: Scribner), Vol. VII, pp. 542–545.
- Gooding, D. (1981), 'Final Steps to the Field Theory: Faraday's Study of Magnetic Phenomena, 1845–1850', *Historical Studies in the Physical Sciences* **11**, 231–275.
- Gooding, D. (1982), 'Empiricism in Practice: Teleology, Economy, and Observation in Faraday's Physics', *Isis* **73**, 46–47.
- 'sGravesande, W. J. (1726), *Mathematical Elements or Natural Philosophy Confirm'd by Experiments; or, an Introduction to Sir Isaac Newton's Philosophy* (translated by J. T. Desaguliers) (2nd edn) (London) Vol. I.
- Gregory, R. L. (1984), *Mind in Science* (Penguin Books).
- Grosser, M. (1979), *The Discovery of Neptune* (New York: Dover).
- Hacking, I. (1983), *Representing and Intervening* (Cambridge: Cambridge University Press).
- Hanson, R. N. (1979), *Patterns of Discovery* (reprint) (Cambridge: Cambridge University Press).
- Heisenberg, W. (1971), *Physics and Beyond* (London: Allen & Unwin).
- Hertz, H. (1896), *Miscellaneous Papers*, with an introduction by P. Lenard (translated by D. E. Jones and G. A. Schott) (London: Macmillan).
- Hesse, M. (1974), *The Structure of Scientific Inference* (London: Macmillan).
- Hillman, H. and Sartory, P. (1977), 'The Unit Membrane, the Endoplasmic Reticulum, and the Nuclear Pores are Artefacts', *Perception* **6**, 667–673.

- Hon, G. (1985), *On the Concept of Experimental Error*, Ph.D. thesis, London University.
- Hon, G. (1987a), 'H. Hertz: "The Electrostatic and Electromagnetic Properties of the Cathode Rays are Either *Nil* or Very Feeble."' (1883) A Case-study of an Experimental Error', *Studies in History and Philosophy of Science* **18**, 367–382.
- Hon, G. (1987b), 'On Kepler's Awareness of the Problem of Experimental Error', *Annals of Science* **44**, 545–591.
- Hon, G. (1989) 'Is There a Concept of Experimental Error in Greek Astronomy', *The British Journal for the History of Science* **22**, No. 2.
- Hon, G. (forthcoming), 'Ionization by Collisions in Gases: J. Franck and G. Hertz versus J. S. Townsend; A Study of Two Types of Experimental Error', *Historical Studies in the Physical and Biological Sciences*.
- Humphreys, W. C. (1968), *Anomalies and Scientific Theories* (San Francisco: Freeman, Cooper & Co.).
- Ising, G. (1926), 'A Natural Limit for the Sensibility of Galvanometers', *Philosophical Magazine* **1**, 827–834.
- Jeffreys, H. (1973), *Scientific Inference* (3rd edn) (Cambridge: Cambridge University Press).
- Kuhn, T. S. (1958), 'The Caloric Theory of Adiabatic Compression', *Isis* **49**, 132–140.
- Latour, B. and Woolgar, S. (1979), *Laboratory Life* (Beverly Hills/London).
- Laymon, R. (1985), 'Idealization and the Testing of Theories by Experimentation', in: Achinstein and Hannaway (1985), pp. 147–173.
- Lenard, P. (1900), 'Erzeugung von Kathodenstrahlen durch ultra-violettes Licht', *Annalen der Physik und Chemie* **2**, 359–375.
- Lenard, P. (1902), 'Ueber die lichtelektrische Wirkung', *Annalen der Physik und Chemie* **8**, 149–198.
- Mach, E. (1974), *The Science of Mechanics* (translated by T. J. McCormack) with an introduction by K. Menger (6th edn) (Lasalle, Ill.: The Open Court).
- Mach, E. (1976), *Knowledge and Error: Sketches on the Psychology of Enquiry* (1905), with an introduction by E. N. Hiebert (Dordrecht, Holland: Reidel).
- Margenau, H. (1950), *The Nature of Physical Reality, A Philosophy of Modern Physics* (New York: McGraw-Hill).
- Maskelyne, N. (1799), 'Astronomical Observations at Greenwich, section for 1795, pp. 3, 319 and 339 et seq.
- Mellor, D. H. (1965), 'Experimental Error and Deducibility', *Philosophy of Science* **32**, 105–122.
- Mellor, D. H. (1966), 'Inexactness and Explanation', *Philosophy of Science* **33**, 345–359.
- Mellor, D. H. (1967), 'Imprecision and Explanation', *Philosophy of Science* **34**, 1–9.
- Millikan, R. A. and Winchester, G. (1907), 'The Influence of Temperature upon Photoelectric Effects in a very High Vacuum, and the Order of Photo-electric Sensitiveness of the Metals', *Philosophical Magazine* **14**, 188–210.
- Millikan, R. A. (1916a), 'A Direct Photoelectric Determination of Planck's "*h*"', *Physical Review* **7**, 355–388.
- Millikan, R. A. (1916b), 'The Existence of Subelectrons', *Physical Review* **8**, 595–625.
- Millikan, R. A. (1917), 'A New Determination of *e*, *N*, and Related Constants', *Philosophical Magazine* **34**, 1–30.
- Millikan, R. A. (1922), *The Electron* (7th edn) (Chicago: University of Chicago Press).
- Morgan, A. de (1847), *Formal Logic: or, The Calculus of Inference, Necessary and Probable* (London: Taylor & Walton).

- Mutch, T. A., Arvidson, R. E., Head, J. W. III, Jones, K. L. and Saunders R. S. (1976), *The Geology of Mars* (Princeton, N.J.: Princeton University Press).
- Nobel Lectures, Physics 1901–1921 (Amsterdam: Elsevier) (1967).
- Nobel Lectures, Physics 1922–1941 (Amsterdam: Elsevier) (1965).
- Nye, M. Jo (1980), 'N-Rays: An Episode in the History and Psychology of Science', *Historical Studies in the Physical Sciences* **11**, 125–156.
- Pais, A. (1983), '*Subtle is the Lord...*' *The Science and the Life of Albert Einstein* (Oxford: Oxford University Press).
- Parratt, L. G. (1961), *Probability and Experimental Errors in Science* (New York: Wiley).
- Pickering, A. (1981), 'The Hunting of the Quark', *Isis* **72**, 216–236.
- Pickering, A. (1984), *Constructing Quarks* (Oxford: Oxford University Press).
- Polanyi, M. (1964), *Science, Faith and Society* (Chicago: Phoenix Books, University of Chicago Press).
- Salmon, W. C. (1967), *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press).
- Scriven, M. (1961), 'The Key Property of Physical Laws – Inaccuracy', in: Feigl and Maxwell (1961), pp. 91–101.
- Sellars, W. (1961), 'The Language of Theories', in: Feigl and Maxwell (1961), pp. 57–77.
- Sheynin, O. B. (1966), 'Origin of the Theory of Error', *Nature* **211**, 1003–1004.
- Sheynin, O. B. (1973), 'Mathematical Treatment of Astronomical Observations', *Archive for History of Exact Sciences* **11**, 97–126.
- Sheynin, O. B. (1983), 'Corrections and Short Notes on My Papers', *Archive for History of Exact Sciences* **28**, 171–195.
- Stuewer, R. H. (1970), 'Non-Einsteinian Interpretations of the Photoelectric Effect', in: R. H. Stuewer (ed.), *Historical and Philosophical Perspectives of Science* (Minneapolis: Minnesota Studies in the Philosophy of Science, University of Minnesota Press), Vol. V, pp. 246–263.
- Stuewer, R. H. (1985), 'Artificial Disintegration and the Cambridge–Vienna Controversy', in: Achinstein and Hannaway (1985), pp. 239–307.
- Sturchio, J. L. (1988), 'Artifact and Experiment', *Isis* **79**, 369–372.
- Thomson, J. J. (1897), 'Cathode Rays', *Philosophical Magazine* **44**, 293–316.
- Thomson, J. J. (1899), 'On the Masses of the Ions in Gases at Low Pressures', *Philosophical Magazine* **48**, 547–567.
- Thomson, J. J. (1902), *Conduction of Electricity through Gases* (Cambridge: Cambridge University Press).
- Thomson, W. and Tait, G. P. (1872), *Elements of Natural Philosophy* (Cambridge: Cambridge University Press), Pt. I.
- Tilling, L. (1973) *The Interpretation of Observational Errors in the 18th and Early 19th Centuries*, Ph.D. thesis, London University.
- Topping, J. (1975), *Errors of Observation and Their Treatment* (4th edn) (London: Chapman & Hall, Science Paperbacks 62).
- Truesdell, C. (1980), *The Tragicomic History of Thermodynamics 1822–1854* (New York: Studies in the History of Mathematics and Physical Sciences, No. 4, Springer Verlag).
- Waismann, F. (1959), 'The Decline and Fall of Causality', in: Crombie (1959), pp. 84–154.
- Weyl, H. (1963), *Philosophy of Mathematics and Natural Science* (New York: Atheneum).

- Whittaker, E. T. and Robinson, G. (1924), *The Calculus of Observations* (London: Blackie & Son).
- Wisdom, J. O. (1971), 'Four Contemporary Interpretations of the Nature of Science', *Foundations of Physics* **1**, 269–284.
- Wittgenstein, L. (1977), *On Certainty*, G. E. M. Anscombe and G. H. von Wright (eds) (translated by D. Paul and G. E. M. Anscombe) (Oxford: Blackwell).
- Wood, R. W. (1904), 'The n-Rays', *Nature* **70**, 530–531.
- Worrall, J. (1982), 'The Pressure of Light: The Strange Case of the Vacillating "Crucial Experiment"', *Studies in History and Philosophy of Science* **13**, 133–171.
- Worsnop, B. L. and Flint, H. T. (1951), *Advanced Practical Physics for Students* (9th edn) (London: Methuen).
- Young, H. D. (1962), *Statistical Treatment of Experimental Data* (New York: McGraw-Hill).
- Ziman, J. (1978), *Reliable Knowledge* (Cambridge: Cambridge University Press).