Revue Inst. Int. de Stat.
25 : 1/3 (1957)

7

"INDUCTIVE BEHAVIOR" AS A BASIC CONCEPT OF PHILOSOPHY OF SCIENCE

by

J. Neyman

Research Professor at the Institute for Basic Research in Science at the University of California, Berkeley, California

1. *Philosophy of science as an empirical scientific discipline.* The term philosophy of science, as used in the present article, is understood to mean an empirical discipline concerned with the nature of scientific research. In a way, it is analogous to anatomy or physiology. When children grow up a little, they are taught how to walk and how to make noises of a particular kind, called speech, associated with their mental processes. Within a relatively short time this training brings excellent success and the children learn to walk, to run, and to express their thoughts verbally. However, while performing all these functions to everyone's satisfaction, the children, and also most adults, are entirely unaware of what is going on in detail. For a human being to start walking, certain muscles must begin to contract and to relax, in a certain order. Which muscles, in what order? Can any of these contractions be omitted without precluding the process of walking? All these questions cannot be answered by the majority of hikers. The process of walking goes on without thinking and the muscles contract and relax automatically in an established order. Then an anatomist appears on the scene, studies the nature of muscles, classifies them according to their role in walking and eventually may be capable of answering the question as to which of these muscles are indispensable for this function (basic) and which, if destroyed by a mishap or by an operation, would be, so to speak, expendable. At this stage of study, the term "basic" would be applied to selected muscles the use of which is most difficult to avoid in the process of walking. However, and this is an essential point, this assignment of the term "basic" can hardly remain definitive. For, in the study of human physique, the anatomist is followed by a neurologist. The studies of the latter reveal that success in walking depends very much on the general organization of successive contractions and relaxations of the available muscles, and that this organization is determined by the functions of the nervous system and the brain. If some of the muscles are affected and have to remain inactive, but the nervous system is in order, the individual concerned may still be capable of walking, perhaps using a cane. On the other hand, a serious defect in the nervous system may totally immobilize a human, even though all his muscles be in relative order. After this finding, we may decide to reassign the term "basic". As far as walking is concerned, the nervous system may be recognized as the basic element of our anatomy, and even the most important muscles may be assigned only secondary importance.

The general development of science proceeds very much as the process of children's learning to walk. With a tremendous expenditure of time, energy, and funds, a great number of organized sets of observations and of experiments are performed daily. The results of these observations and experiments are recorded, and some of them are systematized and collated with material accumulated earlier. Then theoreticians ponder on the records and work out theories. This is a familiar collective process, developing more or less spontaneously, with the scientists' attention centering on the subjects of their study, rather than on their own mental processes. The situation is very similar to that of a child intent on making its first trip across the nursery without falling and entirely unaware of his muscles and of the order in which these muscles should contract and relax. It is the problem of philosophy of science to anatomize scientific research, to bring out distinctly the separate mental processes involved and to arrange their classification. Thus, in relation to science, the philosophy of science plays the same role as anatomy and neurology play in relation to the process of walking.

In the present paper I intend to sketch an anatomization of a particular phase of scientific research, namely the concluding phase, frequently labeled *inductive reasoning*. I hope to show that the mental processes concerned involve an element that appears to have been overlooked by most philosophers of science and that, furthermore, these same mental processes are very different from those, commonly labeled by the term *reasoning*, that are involved in proving a theorem, say in Euclidean geometry. In the light of these findings, it will appear that the term "inductive reasoning" is a misnomer, contributing to the confusion regarding the nature of scientific research, and that a better term would be something like *inductive behavior*. Furthermore, I hope to show that the understanding of the true nature of the concluding phase of research in the sense of inductive behavior, notably by Laplace and Gauss, resulted in the development of a very powerful scientific tool, interesting by itself and important in substantive studies. It appears likely that the present revival of the concept of inductive behavior will bring about similarly important results.

In order to reach this final goal, it will be necessary to make a digression.

2. *Scientific theories are models of natural phenomena.* One of my favorite ideas, learned from Mach via Karl Pearson's "Grammar of Science", is that scientific theories are no more than models of natural phenomena, frequently inadequate models. A model is a set of invented assumptions regarding invented entities such that, if one treats these invented entities as representations of appropriate elements of the phenomena studied, the consequences of the hypotheses constituting the model are expected to agree with observations. If, in all relevant trials, the degree of conformity appears to us satisfactory, then we consider the model an adequate model. Otherwise, the model is inadequate but still may be useful for limited purposes.

Surveyors and astronomers engage in measurements of what I shall generally describe as spatial relations. The object of such studies is a class of phenomena that may be termed spatial phenomena. Euclid constructed the first model of these phenomena. This model deals with various invented entities like the straight line, plane, sphere, triangle, etc.

These invented entities were invested by Euclid with especially invented properties enumerated in his "axioms". Together, the invented entities with their invented properties constitute Euclidean geometry, the Euclidean model of the familiar spatial phenomena. For a long time this model was commonly considered adequate to the point that Euclid's axioms (including the one on parallel lines) have been and still are described in textbooks and dictionaries as "self evident truths". More recently, certain astronomical measurements appear to indicate that the Euclidean model of space is not entirely adequate, though useable within the limits of daily experience. While the hypothetical character of Euclidean axioms appears to have been recognized for some time, it took the genius of Riemann to put the idea in unmistakable terms. This occurred in the title of Riemann's famous trial lecture *Über die Hypothesen welche der Geometrie zu Grunde legen.*

3. *Indeterministic approach to natural phenomena.* Up to about the beginning of the

twentieth century, an overwhelming majority of the attempts at scientific theories were deterministic in character. For a succession of elements of each complex phenomenon, efforts were made to establish formulae by which the value of one element Z could be calculated from known values of some other elements X and Y. There prevailed the belief that, provided such prior elements of the phenomena X, Y, . . ., sufficient in number, are properly selected, their values would determine uniquely the value of Z. Hence the term "determinism".

While the deterministic approach brought about colossal progress in science, it failed to produce desirable results in a number of domains. One such is the domain of hereditary phenomena. Sir Francis Galton and Karl Pearson entertained hopes that, if the characteristics of a long line of ancestors of a given individual could be ascertained, then it would be found that these ancestral characteristics determine uniquely the characteristics of the descendant, perhaps with an allowance for influences of environment. While the studies of the two scholars brought very interesting results (correlations between father and son, etc.), their general model of hereditary phenomena proved inadequate and is abandoned. In fact, it is abandoned in favor of an alternative model, due to Mendel, which has an entirely different character, being indeterministic.

Instead of attempting to establish a (deterministic) relation between the characteristics of the progeny and those of the ancestors (and perhaps of some other factors) Mendel's model attempts to establish a relation between the *frequency* of the progeny having specified characteristics and the given characteristics of the ancestors. The distinction between the two approaches to the same phenomena is very important and, therefore, I shall emphasize it by formulating the questions that the two different approaches are intended to answer. In the one case we consider ancestors of a given individual and ask exactly what this individual's characteristics, say the color of his eyes, must be. In the other case we consider *a category of individuals*, all having their ancestors with the same fixed set of characteristics, and ask *how frequently* the individuals of this category have blue eyes. This latter approach, characterized by the question "how frequently?" is the indeterministic approach. A somewhat naive but appealing version of the point of view behind this approach is that there are no sets of "causes" that *determine* future developments uniquely for some phenomena. A less naive version of the same point of view is that, while all the attempts made thus far at a deterministic model of a certain class of phenomena have failed, there appears to be a possibility of answering the question "how frequently?" and this answer is interesting in itself.

As is well known, Mendel's model of hereditary phenomena proved a great success and developed into what we now call genetics. Another class of phenomena in which the indeterministic approach proved singularly successful is atomic and nuclear physics. In addition, the same indeterministic approach yields excellent results in a number of other fields, astronomy, biology, experimentation, industrial research, medicine, etc., etc., without end.

The basic mathematical tool of the indeterministic approach to phenomena, that is, the general model of the phenomena of relative frequencies, is provided by the calculus of probability. Its origin lies in an attempt made in the seventeenth century to deal theoretically with series of games of chance. The modern term "probability" represents an idealization of the empirically observable relative frequency in the same sense as the Euclidean straight line serves as an idealization of the measuring chain strung between two posts by a surveyor.

I emphasized the fact that the development of science proceeds somewhat like the process of learning to walk by children, with considerable successes achieved by scholars, frequently without clear realization of exactly what has been done. The development of the theory of probability presents an excellent example of this. In fact, an entirely clear realization of the scientific role of the theory of probability came about only after centuries of studies on the theory itself. This clear cut analysis of the status of the theory of probability among other disciplines, namely as a general model of the phenomena of relative frequency, is due to the recently deceased German philosopher and probabilist Richard von Mises [9]. Von Mises attempted to complement his philosophical views by a fresh attempt at the foundations of the mathematical theory of probability, in which probability is conceived as a postulated limit of relative frequency. This attempt is frequently viewed as not entirely successful. At any rate, opinion seems to prevail that the most satisfactory system of axioms underlying the theory of probability is that of the Russian scholar, A. N. Kolmogorov [7].

4. *Alternative views on probability theory*. The above view on the role of the theory of probability in science is often called the "frequentist" view. Two other schools of thought should be mentioned briefly. According to one of them, which I shall symbolize by the name B. O. Koopman [8], the theory of probability is a mathematical model of mental processes concerned with intensities of beliefs. This particular theory, then, is essentially of the same kind as the frequentist theory. Both are mathematical models and the difference between the two is that they are models of two different classes of phenomena. Also, they have the important characteristics common to all faultless mathematical models, that their problems consist of calculating values of some unknown invented entities from the given values of other entities of the same kind, all included within the same model.

Another kind of theory of probability, symbolized by the name of Sir Harold Jeffreys [6], differs from the two mentioned above in that it does not appear to provide a model of any phenomena. However, it does deal with intensities of belief and one of its main problems appears to be the assignment of numerical probabilities to particular propositions concerned with real phenomena. As Sir Harold writes, p. 36, "Our main postulates are the existence of unique reasonable degrees of belief, which can be put in a definite order . . .".

In what follows, I shall be concerned solely with the frequentist theory of probability, intended to provide a mathematical model of relative frequencies. Within this theory, the assertion that the probability of A is equal to, say, one-half, is interpreted to mean that, in a specified hypothetical series of instances, the relative frequency of A is equal to one-half.

5. *Anatomization of the concluding phase of scientific research*. After this somewhat long digression, we now return to the main subject of this article and examine the concluding phase of scientific research with the object of determining the nature of its constituent mental processes. This is just the phase frequently described as induction. I hope to show that its constituent processes fall under three headings: (i) visualization of several possible sets of hypotheses relevant to the phenomena studied, (ii) deductions from these sets of hypotheses, and (iii) an act of will or a decision to take a particular action, perhaps to assume a particular attitude towards the various sets of hypotheses mentioned under (i).

The mental processes of category (i) are not reasoning, at least not reasoning in the sense this word is used with reference to proofs of theorems. They represent the work of creative imagination, combined with a careful scanning of information stored in the scholar's memory. The processes under (ii) are reasoning, but all this reasoning is, of course, deductive. The processes of the last category, (iii), constitute the element, mentioned in the first section of this paper, that appears to have been either overlooked or misinterpreted. These processes are certainly not any sort of "reasoning", at least not in the sense in which this word is used in other instances; they are acts of will.

In the present connection, it would be interesting to review and to analyze a considerable number of pieces of research, each time pointing out the elements of the three categories of mental processes enumerated. However, the framework of the present paper requires economy of time and space and, therefore, only three examples will be examined briefly. The reader will have no difficulty in performing similar analyses of as many samples of research as desired.

The first two examples are selected from the well-known book of Sir Ronald Fisher, *Statistical Methods for Research Workers*, Eighth edition, 1941 [4]. This choice is dictated by two considerations. The first is that, in his philosophical discussions, Sir Ronald is one of the most emphatic proponents of "inductive reasoning". Therefore, it is particularly interesting to examine his mental processes when he is engaged, not in philosophical discussions, but in actual substantive research. It would have been even better to analyze sections of Fisher's *Design of Experiments* or some work on genetics. Unfortunately, the considerations of brevity are strongly against such a choice because it would require somewhat lengthy quotations or a danger of taking certain sentences out of their context. On the other hand, the examples used by Fisher in his *Statistical Methods* ... are remarkable for their compactness and unmistakable terms, and can be quoted almost without omission. In quoting them, I omit all arithmetic and also some passages that are not relevant to the present discussion.For the sake of ease in reference, the passages quoted are marked with successive letters of the alphabet, (a), (b), etc. Some of the most relevant passages are italicized. On pages 86—88 we find the following:

> ex. 12. *Test of independence in a 4 × 4 classification.* – As an example of a more complex contingency table we may take the results of a series of back-crosses in mice, involving the two factors Black-Brown, Self-Piebald (Wachter's data):
>
> TABLE 18
>
> ...
>
> The back-crosses were made in four ways, ...
>
> (a) The simple Mendelian ratios may be disturbed by differential viability, by linkage, or by linked lethals. Linkage is not suspected in these data, ...
>
> (b) ... and if the only disturbance were due to differential viability of the four genotypes, these should always appear in the same proportion; to test if the data show significant departures we may apply the χ^2 test to the whole 4 × 4 table. The values expected on the hypothesis that the proportions are independent of the matings used, or that the four series are homogeneous, are given above in brackets. The contributions to χ^2 made by each cell are given below (Table 19).
>
> TABLE 19
>
> ...
>
> The value of χ^2 is therefore 21.832; ...
>
> (c) For $n = 9$, the value of χ^2 shows that *P is less than .01, and therefore the departures from proportionality are not fortuitous;* it is apparent that the discrepancy is due to the exceptional number of Brown Piebalds in the F_1 males Repulsion series. ...
>
> (d) *Provided the deviation is clearly significant, it is of no practical importance whether P is .01 or .000,001, and it is for this reason that we have not* tabulated the value of χ^2 beyond .01. ...

The passages of Fisher omitted are those dealing with arithmetic, those given to the tdetails of the experment, and those containing the discussion of the fact that the χ^2 est is used to establtish association but not to measure the degree of association. All these points are irrelevant for the present purposes.

The examination of the passages quoted shows that (a) falls under the heading (i) of my classification. In fact, in (a), Fisher reviews several sets of hypotheses regarding the chance mechanism that might have produced the observations. The passage marked by the letter (b) represents, in a compact manner, the *deductions* from the hypotheses that Fisher chose to consider. The most relevant passages are (c) and (d). Following as they do the empirical part of the study, these particular passages might have been expected to contain "inductive reasoning". Indeed, the italicized sentence in (c) does look like a deduction. The trouble is that the premise "*P* is less than .01" does not imply that "the departures are not fortuitous". In fact, even if the inheritance of the characteristics considered conformed exactly with the assumed model, the probability of observing χ^2 corresponding to the value of P less than 0.01 is positive and approximately equal to 0.01. Thus, the assertion "the departures are not fortuitous" cannot be *deduced* from "*P* is less than .01". Yet, this assertion is made, and is made in very definite terms. With reference to passage (a), one may presume that the assertion "the departures are not fortuitous" is interpreted by Fisher as equivalent to the adoption of the hypothesis of differential viability. Presumably, if some further experimentation is conducted, the experiments would be designed taking it for granted that the viability of mice does depend upon their genetical composition. Should it happen that, in actual fact, the viability of all the genetical types be the same, there would result some unpleasantness.

I submit that the contents of passage (c) have nothing in common with reasoning and that this passage amounts to taking a "calculated risk", to an act of will to behave in the future (perhaps until new experiments are performed) in a particular manner, conforming with the outcome of the experiment. It is this act of will adjusting our behavior to the results of observations that is the overlooked element of the final stages in scientific research and that is covered by the term "inductive behavior".

Passage (d) deserves special notice. It emphasizes the automatic character of Fisher's own rule of inductive behavior to reject the hypothesis tested whenever the selected test leads to the value of P that is less than the fixed number 0.01. There are weighty arguments against this automatism. In fact, it appears desirable to determine the level of significance in accordance with quite a few circumstances that vary from one particular problem to the next. One is the consequences of committing an error (so-called of the first kind) when the hypothesis tested is in fact true and is rejected. Depending upon the problem, these consequences may be grave and, in such a case, a more stringent level of significance, perhaps 0.001, will be indicated. Another circumstance to be considered when fixing the level of significance is the consequences of failing to reject the hypothesis tested when, in fact, it is false (error of the second kind), and a specified alternative hypothesis is true. Also, the probability of this error (or, equivalently, the power of the test) must be taken into account. On occasion it happens that if one insists on the level of significance 0.01, then the probability of failing to detect the falsehood of the hypothesis tested, when it is in fact false to an importance degree, is very small, substantially less than one-half. On the other hand, if one relaxes the level of significance a little, say taking it at 0.02 or 0.05, then the probability of detecting an important degree of error in the hypothesis tested may jump up to, say 0.8. If this be the case, then, depending upon the relative importance of the two kinds of error, there may be an advantage in working with 0.02 or 0.05 as the level of significance rather than with 0.01.

On p. 89 Fisher describes his next example.

> ex. 13. *Homogeneity of different families in respect of ratio black : red.* – The following data show in 33 families of *Gammarus* (Huxley's data) the numbers with black and red eyes respectively:
>
> TABLE 20
>
> . . .
>
> (a) The totals 2565 black and 772 red are distinctly not in the ratio 3 : 1; the discrepancy is ascribed to linkage. The question before us is whether or not all the families indicate the same ratio between black and red, or whether the discrepancy is due to a few families only.
>
> (b) For the whole table $\chi^2 = 35.620$, $n = 32$. This is beyond the range of the table, so we apply the method explained on p. 79:
>
> $$\sqrt{2\chi^2} = 8.44;$$
> $$\sqrt{2n-1} = 7.94;$$
> $$\text{Difference} = +\ .50 \pm 1.$$
>
> (c) The series is therefore not significantly heterogeneous; *effectively all the families agree and confirm each other* in indicating the black-red ratio observed in the total.

In this quotation, passage (a) contains process (i) of my classification, namely scanning of possible systems of relevant hypotheses. Passage (b) contains deductions in a very abbreviated form. Passage (c) summarizes the results and, as in the preceding example, does not involve any reasoning. Instead, it is an outspoken case of an automatic adherence to a preassigned rule. Much too automatic, to my mind. I need not emphasize to the reader that large values of P may be obtained when the hypothesis tested is false to an important degree. Thus, before making a strong statement like that italicized, it is advisable to investigate at least the two following points: (a) how important will be the consequences of accepting the hypothesis of homogeneity of the families considered when, in fact, they are heterogeneous to a specified degree, and (b) what is the probability (probability of an error of the second kind) of obtaining a large value of P in cases when the families are heterogeneous to the degree that is judged important? Should this latter probability be large, the result of the experiment would be summarized as follows: the scope of the experiment (that is, its size and/or design) is insufficient to allow the assumption of a definite attitude towards the possibility of heterogeneity of the families considered. No such considerations are present in Fisher's treatment of the problem.

One more quotation from Fisher's book, this time of an isolated sentence. In the concluding paragraph of p. 109, Sir Ronald writes: "The absence of significant values in the first three columns of Table 25.4 shows that no further modifications of the divisors are necessary . . .". Here Sir Ronald is quite explicit on the point that the calculated values of P lead him, automatically, *to a choice between two contemplated actions:* if the P were small, he would have made "further modifications of the divisors"; otherwise, no "further modifications" are necessary. In this example, Fisher's choice of the action seems again precipitous because it is made without any consideration of the probability of obtaining no "significant values" even though, in actual fact, the divisors need important modifications. Suppose it is found that this probability is something like 0.9. Would not then the decision "no further modifications of the divisors are necessary" appear somewhat risky?

My third example of scientific research to be briefly anatomized here is taken from my own experience. For the last seven years, under the influence of, and in cooperation with, Elizabeth L. Scott and C. D. Shane, my colleagues at the University of California, I have been engaged in the study of the distribution of galaxies in space and, in particular, of the problem of the expansion of the Universe. Currently there are a number of cosmological theories (that is, models of the phenomenon of the Universe, with its galaxies and clusters of galaxies, etc.). Some theories postulate that clusters of galaxies recede from each other with velocities increasing with distance. Astronomers are divided. Some act on the hypothesis that expansion is a reality and devote their efforts to the question of possible temporal changes in the velocities of expansion. Contrary to this, other astronomers act on the hypothesis that there is no expansion at all. Our own group at the University of California seeks new evidence to be treated by new statistical methods. Our evidence consists of the distribution of images of galaxies on the photographs of the sky taken by Professor Shane at the Lick Observatory. Our method consists in comparing certain numerical characteristics of this distribution with the consequences of our theoretical models, involving various "invented entities", such as "clusters of galaxies", "superclusters", "observational errors", and either the presence or absence of "universal expansion". As might be expected, the problem is extremely complicated, involves a heavy observational program, and depends on laborious computations. As a result, it has to be treated in steps. After each completed step there is a variety of further steps to choose from. In our last publication on the subject [13], we more or less abandoned (an act of will!) our original model involving the hypotheses of (a) stationary universe, (b) simple clustering, and (c) a particular mechanism of observational errors, because the agreement between the consequences of this model and the results of Shane's observations does not seem satisfactory to us. Thus, we had to look for new avenues; after weighing the consequences of this or that possible modification of the original model, we decided (an act of will!) to give it another chance, with a modification of the postulated mechanism of observational errors. Regretfully, in order to effect this modification in definite terms, a new observational program became necessary.

To sum up, whatever the scale of the study, and whether this study is conducted for purposes of some immediate practical action (should one use the Salk vaccine against polio?), or for the sake of scientific curiosity (does our Universe expand?), the examination of experimental or observational data is invariably followed by a set of mental processes that involve the three categories enumerated above: (i) scanning of memory and a review of the various sets of relevant hypotheses, (ii) deductions of consequences of these hypotheses and the comparison of these consequences with empirical data, (iii) an act of will, a decision to take a particular action. None of these processes fits the term "inductive reasoning". On the other hand, because process (iii) results in an adjustment of our future actions to the results of observations, it may appropriately be labeled inductive behavior.

6. *Role of the concept of inductive behavior in modern research.* As a result of the above anatomization of the concluding phase of scientific research, one might perhaps admit that the term inductive behavior is a better term to describe the phase than the term inductive reasoning. However, this, by itself, can hardly be considered adequate to elevate inductive behavior to the dignity of a basic concept of the philosophy of science. With reference to the analogy with the role of muscles or the nervous system in the process of walking (see section 1), in order to treat inductive behavior as an important concept, it is necessary to bring clearly to light the difference that the adoption of the inductive behavior approach makes in scientific research.

As we have seen, whether a piece of serious scientific research is conducted by a proponent of "inductive reasoning" or of "inductive behavior", it invariably ends with the application of some adopted rule of behavior. On occasion, the rule used by a behaviorist will coincide with that applied by a non-behaviorist. In other cases the two rules will differ. The question arises as to what is the origin of these rules and whether there is any reason to prefer one category of rules to another.

Naturally, the sources of rules of making decisions at the concluding stage of research must be sought within the philosophical background of each school of thought. This background of the inductive reasoning approach is discussed in detail by its proponents in their many writings. All of them appear to have a particular point in common.

The delicacy of the questions discussed and the arguments adduced in these works tend to blur the real issue. However, if one tries to present it briefly, if somewhat bluntly, the common element of all writings on the inductive reasoning approach appears to indicate the conviction of the authors that it is possible to devise a formula of universal validity which can serve as a normative regulator of our beliefs. Furthermore, the absence or, at least, the rarity of discussions of consequences of this or that choice of action suggests another conviction, namely that human actions are motivated predominantly, if not exclusively, by beliefs. Given a set of hypotheses, given the information stored in the memory of the scientist, and given a set of data, the scientist is expected to evaluate the normative formula, to adjust his beliefs to the value obtained, and to act accordingly.

While the conviction of the possibility of a universal normative regulator of beliefs is common to the writers on inductive reasoning and may serve as a definition of this particular school of thought, these writers differ in their views on the nature of the appropriate regulator. For Sir Harold Jeffreys, the normative regulator of beliefs is Bayes' formula, giving the inverse probability of hypotheses. However, the validity of this formula is limited to its evaluation using the particular prior probabilities devised by Sir Harold for this purpose. This view is contrasted by that of Sir RonaldFisher who is convinced that the inverse probability approach is founded on an error. A brief relevant statement of Fisher's opinions is contained in the following passage, quoted from his *Statistical Methods*..., pp. 9—10.

> For many years, extending over a century and a half, attempts were made to extend the domain of the idea of probability to the deduction of inferences respecting populations from assumptions (or observations) respecting samples. Such inferences are usually distinguished under the heading of Inverse Probability, and have at times gained wide acceptance. This is not the place to enter into the subtleties of a prolonged controversy; it will be sufficient in this general outline of the scope of Statistical Science to reaffirm my personal conviction, which I have sustained elsewhere, that the theory of inverse probability is founded upon an error, and must be wholly rejected. Inferences respecting populations, from which known samples have been drawn, cannot by this method be expressed in terms of probability, except in the trivial case when the population is itself a sample of a super-population the specification of which is known with accuracy.
>
> The probabilities established by those tests of significance, which we shall later designate by t and z, are, however, entirely distinct from statements of inverse probability, and are free from the objections which apply to these latter. Their interpretation as probability statements respecting populations constitute an application unknown to the classical writers on probability. To distinguish such statements as to the probability of causes from the earlier attempts now discarded, they are known as statements of Fiducial Probability.
>
> The rejection of the theory of inverse probability was for a time wrongly taken to imply that we cannot draw, from knowledge of a sample, inferences respecting the corresponding population. Such a view would entirely deny validity to all experimental

science. What has now appeared is that the mathematical concept of probability is, in most cases, inadequate to express our mental confidence or diffidence in making such inferences, and that the mathematical quantity which appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability. To distinguish it from probability, I have used the term "Likelihood" to designate this quantity; since both the words "likelihood" and "probability" are loosely used in common speech to cover both kinds of relationship.

It is seen that the inverse probability approach is "discarded" and that, instead, Fisher introduces two new measures "of our mental confidence or difference". Thus, if a scientist inquires why should he reject or accept hypotheses in accordance with the calculated values of P (see the examples quoted from Fisher, section 5) the unequivocal answer is: because these values of P are the ultimate measures of beliefs especially designed for the scientist to adjust his attitudes to. If one inquires why should one use the normative formulae of one school rather than those of some other, one becomes involved in a fruitless argument.

It must be obvious that, with the above essential contents of the inductive reasoning approach, its use as a basic principle underlying research is unsatisfactory. The beliefs of particular scientists are a very personal matter and it is useless to attempt to norm them by any dogmatic formula. Furthermore, our actions, while influenced by beliefs, are also motivated by considerations of consequences, that is to say, by considerations of what is desired to be achieved. In fact, on occasion, we act against our beliefs. When I embark on a pleasure trip by airplane, I am firmly convinced that this trip will not end in an accident. Yet, considerations of consequences of an accident lead me to buy an insurance policy.

While rejecting the inductive reasoning approach because of its dogmatism, lack of clarity, and because of the absence of considerations of consequences of the various actions contemplated, I am appreciative of the several sections of the literature that this approach generated. For example, I am highly appreciative of the literature concerned with the properties of maximum likelihood estimates. However, it will be noticed that the point of view behind this literature is that of inductive behavior. In fact, the question studied is "what will the consequences of using maximum likelihood estimates be?"

The content of the concept of inductive behavior is the recognition that the purpose of every piece of serious research is to provide grounds for the selection of one of several contemplated courses of action. Also, the recognition that the desirability of this or that course of action depends upon the circumstances and, of course, on the subjective preferences and beliefs of the individual concerned. In trivial cases, such as that of buying or not buying an accident insurance policy, the situation is perfectly clear and no calculations are necessary. However, if the problem is slightly more complicated, the scientist may be perfectly clear about what he would like to achieve, but the best method of adjusting his actions is not easy to determine. Therefore, it is practically expedient and theoretically interesting to investigate the totality of possible methods of adjusting our future actions to the results of observations so that it becomes possible for everyone to use the method of adjustment that fits his own case best, in accordance with his own preferences and beliefs.

In the following pages I present a brief sketch of studies of this kind, limiting myself to the indeterministic approach to phenomena. However, I believe that the inductive behavior point of view is equally important in deterministic work. One reason for this belief is that, because of the necessity of considering unavoidable experimental errors, every deterministic study must have some indeterministic elements.

Consider, then, a class, say C, of phenomena for which a tentative indeterministic model has been constructed. Ordinarily, at least at the beginning stages, this model is not quite specific. It involves hypotheses regarding the general and, so to speak, basic chance mechanism behind the phenomena C, but fails to specify certain details, perhaps fails to specify the values of certain parameters. Thus, for example, in the problems quoted from Fisher, the basic chance mechanism underlying the observable variables is the Mendelian law of heredity, with the possibility of linkage. The unspecified details of the mechanism are the degree of the possible linkages and the possible differential viability of the various genetical types. Again, in our astronomical studies, the basic mechanism behind the observations consists in the particular postulated chance machinery behind the distribution of galaxies in space, say in simple clusters, and either expansion in time or lack of this expansion. The unspecified details include the variability in the number of galaxies from one cluster to the next, the distribution of galaxies within a cluster and certain details of the observational errors. The possible subjects of research may include the question of the presence of linkage in one case and of the presence of expansion in the other.

In general terms, then, for the phenomena C, we visualize not just one but a family of models, say of models $M(C, \Theta)$ where Θ is a variable (parameter) symbolizing the details of the general model that are left unspecified and may vary within a certain range (parameter space) Ω. In some cases, Θ may stand for just one number, characterizing a single unspecified detail of the model. In other cases, Θ may be a set of several numbers, each referring to a particular wheel in the chance machinery contemplated. Thus, for example, in genetical studies the only unspecified detail of the model may be the linkage between two pairs of genes. Then Θ would represent a measure of this linkage. If differential viability is also considered, then in addition to the measure of linkage we shall need one or more numbers characterizing the differences in viability of the several genetical types. In this case the single letter Θ stands for a set of parameters $(\Theta_1, \Theta_2, \ldots, \Theta_s)$, called parameter point.

The contemplated research may consist, then, in an attempt to fill some or all of the gaps in the original model. Another possibility is to study whether the original model is adequate to represent the phenomena C. That is to say, whether there is an imaginable way of filling the gaps in the model so that it appears to be adequate. Whatever the case may be, the only bases on which research can proceed are the empirical data. Generally, they will be represented by a set of numbers $X_1, X_2, \ldots, X_n$, to be described as the sample point and to be denoted by a single letter X. The totality of all positions that the sample point may occupy in a given problem is described as the sample space and is denoted by W. Since, by our assumption, the contemplated model is indeterministic, the observations X must be regarded as particular values of random variables. Each contemplated model $M(C, \Theta)$ ascribes to X a probability distribution depending upon Θ, say $P(\Theta)$. It follows, then, that any research concerned with the family of models $M(C, \Theta)$ is equivalent to research concerned with the family of distributions $P(\Theta)$.

If the proposed research is concerned with filling up some or all of the gaps in the original model which by itself is not questioned (this is the more general case), then it is assumed that the parameter space Ω contains a point Θ^*, termed the true parameter point, such that $P(\Theta^*)$ represents the actual distribution function of X. However, the identity of Θ^* is unknown.

Following the point of view of inductive behavior, we consider that the proposed research is motivated by the necessity of taking an action, one of several possible actions with their relative desirability depending on Θ^*. The problem before us is how to perform the choice of action in conformity with the observations X and taking into account the relative desirability of the different actions.

We begin by visualizing the complete set, say A, of actions that may be taken. (a_1 -- Arrange a broad vaccination program with Salk vaccine for all children under sixteen. a_2 -- Limit this program to children from six to sixteen years of age. a_3 -- Do not use Salk vaccine at all, etc. Alternatively: a_1 -- Assume the attitude that the observational errors in the study of the distribution of galaxies in space are adequately represented by the original model [12] and concentrate on whether the galaxies are "superclustered" and/or the universe is expanding. a_2 -- Try a more realistic model of observational errors and obtain more data to test its adequacy, etc., etc.) The set A so defined is called the decision space. It will be seen that any *method* of choosing between actions of the set A in accordance with the result X of the observations is equivalent to a definition on the sample space W of a function, $a = f(X)$ with its values a in the decision space A. In fact, any (nonrandomized) method of the kind described must have the form: if the observations give $X = x_1$, then the action to be taken is a_1; if the observations give $X = x_2$, then the action to be taken is a_2, etc. Following Wald [16] the function $a = f(X)$ so defined is called the statistical decision function. Alternative, or at least associated, terms are: rule of inductive behavior, or simply, decision rule.

In order to be able to make a choice between all possible decision rules, it is necessary to devise a non-dogmatic method of characterizing the properties of any one of them in particular. Provided the decision function selected is "measurable" (which is postulated), the corresponding decision rule will prescribe any of the actions a with calculable frequencies. Thus, to every parameter point Θ in Ω and to every contemplated decision function f, there will correspond probabilities, say $P\{a|f, \Theta\}$ that, in the case when $\Theta^* = \Theta$, the application of the decision function f will lead to the specified action a. The set of these probabilities, corresponding to all possible values of Θ, describes completely the properties of the contemplated rule f. In fact, granting the adequacy of the basic model, these probabilities tell us how frequently the given rule will prescribe any of the actions contemplated when the true parameter Θ^* has this or that possible value. [How frequently will the contemplated rule prescribe mass application of a given vaccine when, in fact, this vaccine is dangerously toxic? How frequently will the same rule prescribe the abandonment of the vaccine when, in fact, this vaccine is capable of protecting 95 per cent of the inoculated children against polio?] For this reason, the set of probabilities $P\{a|f, \Theta\}$, with a ranging over the set A and with Θ ranging over the whole parameter space, is called the performance characteristic of the decision function f.

Whatever the scientist's beliefs and preferences may be, the knowledge of the performance characteristics of all possible decision rules will allow him to choose the one that fits his case best. The approach described is, thus, non-dogmatic and the most general approach to the problem of the concluding stage of every research, that is to say, to the problem of inductive behavior.

The complexity involved in practical applications of the procedure indicated depends upon the circumstances of the problem. If both the sample space W and the decision space A contain a finite number of elements, then there is only a finite number of possible decision rules and, at least in principle, it is possible to examine the performance characteristics of every one of these rules. However, even in this case, there are likely to be practical difficulties. These difficulties multiply when the sample space W is infinite and, particularly, when the decision space is infinite also.

When the one by one examination of performance characteristics becomes difficult, we give it up and, instead, try to solve a problem which, in a sense, is the inverse problem. We specify in advance the properties of the desirable performance characteristic and try to determine the decision function, if such exists, whose performance characteristics possesses the indicated properties. This particular approach resulted in the development of many now familiar concepts of mathematical statistics, such as uniformly most powerful tests of statistical hypotheses, shortest systems of confidence intervals, etc. On an elementary level, the details of these developments are presented in [11]. Excellent comprehensive treatments of theories of testin hypotheses and of estimation have been recently published by Schmetterer [15], and by Dunin-Barkovski and Smirnov [2]. The general theory of statistical decision functions has been developed by Wald [16]. Further interesting developments in the same direction are due to Blackwell and Girshick [1]. Finally, the attention of the reader is called to the forthcoming book by Erich L. Lehmann.

7. *Remark on the use of Bayes' formula.* Perhaps because of lack of clarity in some of my papers, certain authors appear to be under the impression that, for some reason, I condemn the use of Bayes' formula and that I am opposed to any consideration of probabilities *a priori*. This is a misunderstanding. What I am opposed to is the *dogmatism* which is occasionally apparent in the application of Bayes' formula when the probabilities *a priori* are not implied by the problem treated and an author attempts to impose on the consumer of statistical methods the particular *a priori* probabilities invented by himself for this particular purpose. Another unsatisfactory element of this kind of approach is the already discussed lack of consideration of what is desired to be achieved.

The Bayes' approach to the problem of decisions may be of two kinds, the "inductive reasoning" kind and the "inductive behavior" kind. An outstanding example of the latter is presented by the recent paper of Robbins [14]. This brief paper is quite remarkable and may well be the forerunner of extensive important developments.

8. *Historical note.* The term inductive behavior was first introduced [10] in 1937. It was conceived as a didactic means to help in the understanding of the theory of confidence intervals. At that time science was completely dominated by "inductive reasoning" and all sorts of routines of thought were hard to overcome. I felt pleased with the invention of the suggestive term pointing in the right direction and was under the impression that not only the term but also the concept behind it were new. During the subsequent twenty years I have learned differently. The point is that, if one digs a little into the writings of Laplace and, particularly, of Gauss, one finds that the idea of inductive behavior is there.

In fact, it is this idea of inductive behavior, combined with the concept of loss functions for faulty judgments that led Gauss to his brilliant foundation of the method of least squares, independent of the normal distribution, that played and still plays an outstanding role in research. Here is the relevant passage of Gauss [5].

> Die Bestimmung einer Grösse durch eine einem grösseren oder kleineren Fehler unterworfene Beobachtung wird nicht unpassend mit einem Glücksspiel verglichen, in welchem man nur verlieren, aber nicht gewinnen kann, wobei also jeder zu befürchtende Fehler einem Verluste entspricht. Das Risiko eines solchen Spieles wird nach dem wahrscheinlichen Verlust geschätzt, d.h. nach der Summe der Produkte der einzelnen möglichen Verluste in die zugehörigen Wahrscheinlichkeiten. Welchem Verluste man aber jeden einzelnen Beobachtungsfehler gleichsetzen soll, ist keineswegs an sich klar; hängt doch vielmehr diese Bestimmung zum Theil von unserem Ermessen ab. Den Verlust dem Fehler selbst gleichzusetzen, is offenbar nicht erlaubt; würden nämlich

positive Fehler wie Verluste behandelt, so müssten negative als Gewinne gelten. Die Grösse des Verlustes muss vielmehr durch eine solche Funktion des Fehlers ausgedruckt werden, die ihrer Natur nach immer positiv ist. Bei der unendlichen Mannigfaltigkeit derartiger Funktionen scheint die einfachste, welche diese Eigenschaft besitzt, vor den übrigen den Vorzug zu verdienen, und diese ist unstreitig das Quadrat. Somit ergiebt sich das oben aufgestellte Princip.

Laplace hat die Sache zwar auf eine ähnliche Weise betrachtet, er hat aber den immer positiv genommenen Fehler selbst als Mass des Verlustes gewählt. Wenn wir jedoch nicht irren, so ist diese Festsetzung sicherlich nicht weniger willkürlich, als die unsrige . . .

It will be seen that the passage quoted contains all the elements of the inductive behavior approach to the problem of point estimation. There are no references to beliefs. In particular there are no references to what a "rational mind" *should* believe. No dogmatism. Instead there is the recognition that any solution to the problem of estimation must be equivalent to a rule (statistical decision function) ascribing to the quantity to be estimated a value computed from the conditions of the problem and from the results of the observations. Since, by assumption, these observations are random variables, the results of the computations are also random. Further, there is the understanding that the desirability of coming out with any given value of the estimate depends upon the true value of the parameter to be estimated. There is the understanding that the importance of any given error in the estimate is subjective and must depend upon the circumstances of the problem. With this reservation, Gauss proposes to consider that the value of the loss function for faulty judgments is proportional to the square of the eror. This is motivated not by the presumption that it must appeal to everyone, but because of its simplicity. Finally, under the influence of the contemporary preoccupation with games of chance, Gauss chose to look for such methods of estimation (such decision functions) that, whatever the true parameter point might be, the expectation of the estimate equals the true value of the parameter (unbiasedness) and the expectation of the adopted loss function is a minimum. Obviously, these two requirements refer to the performance characteristic of the decision rule sought. The result of Gauss' study is that, when the variables considered are independent, have finite variances and their expectations are linear combinations of the estimated parameters, then the solution of the problem stated exists and is obtainable by the method of least squares.

The reader will see that this is the same kind of problem of inductive behavior indicated at the end of section 6.

9. *Concluding remarks*. Now we may return to the main subject of the present article, as indicated in its title.

Under the influence of certain early writings on probability, including some of Laplace, there became established a conviction that mathematical theory may devise universally valid normative measures of the intensity of belief. This dominating conviction initiated an era of "inductive reasoning" marked by efforts to devise normative measures of belief and characterized by lack of understanding of behavioristic elements in the concluding phases of research. The proposed solutions of problems involved in these phases were based on dogma.

However, while Laplace contributed to the development of the era of inductive reasoning by some of his philosophical writings, in his substantive studies he followed the path of inductive behavior. In fact, as suggested by the above quotation, Laplace's departures from the point of view of measure of belief may have influenced Gauss. Whatever the case may have been, Gauss' clear-cut foundation of the method of least squares could not have been formulated without the adoption of the point of view of inductive behavior. This is the basis of my claim that the concept of inductive behavior has an outstanding place in the philosophy of science.

Remark. After the above article had been completed, I read a paper by B. de Finetti [3] representing the contents of his lecture delivered at the "III Entretiens de Zürich" in 1951. Seeing that Professor de Finetti represents the "subjectivist" school of thought on probability, it is a pleasure to find myself in full agreement with most of what he writes. In particular, I welcome de Finetti's concluding paragraph directed against formulae intended to impose on thinking public normative measures of intensity of beliefs that this public *should* experience in the given circumstances. This paragraph reads:

> Comme on le voit, tout est simple et peut demeurer simple si l'on se contente d'étudier nos raisonnements tels qu'ils sont.

(I interpret this to mean that the situation is satisfactory so long as the intensity-of-belief theory of probability concerns itself with building a model of our mental processes related to beliefs.)

> Tout devient inutilement compliqué et obscur si l'on prétend l'absurde, c'est-à-dire de posséder la faculté du roi Midas et de transformer une phrase quelconque dans l'or d'une "vérité absolue" toute fois que l'on dédaigne d'avouer que nos opinions ne sont et ne peuvent être autre chose que des opinions.

The difference between the standpoint of de Finetti and the one expressed here is, perhaps, that for me philosophy of science is a little more of an empirical discipline than it is for de Finetti. As a result, my own attention is more intensely directed towards the acts of will involved in the concluding phase of any research. As a result, whenever I have previously decided to act as if I were assured that a given stochastic model M adequately represents a certain class of phenomena (here de Finetti would probably say "whenever *I believe* that, for example, hereditary phenomena are governed by Mendel's laws") and I have to choose between several possible actions the desirability of which depends on the details of the phenomena that are unknown, I consider it advisable to make the choice *after* examining (a) the relative undesirability of consequences of the various possible errors and (b) the frequencies, implied by that same model M, with which a given rule of behavior would lead to these different errors. As far as I can see, de Finetti does not discuss these questions and I hope that, if and when they attract his attention, we shall find ourselves in further perfect agreement. Among other things, I hope that we shall agree that the awareness of the problem of inductive behavior leads to practically important and theoretically interesting new mathematical problems.

REFERENCES

[1] Blackwell, D., and M. A. Girshick. *Theory of games and statistical decisions*, New York, J. Wiley and Sons, 1954.

[2] Dunin-Barkovski, I. V., and N. V. Smirnov. *Theory of probabilities and mathematical statistics in industry*, Moscow, Gosisdat, 1955. [Russian].

[3] Finetti, B. de. Expérience et théorie dans l'élaboration et dans l'application d'une doctrime scientifique. "Rev. de métaphysique et de morale", 60 : 3, p. 264-286, 1955.

[4] Fisher, R. A. *Statistical methods for research workers*, 8th ed., London, Oliver and Boyd, 1941.

[5] Gauss, C. F. *Abhandlungen zur Methode der kleinsten Quadrate*, Berlin, 1887. Transl. from the Latin by A. Borsch and P. Simon.

[6] Jeffreys, H. *Theory of probability*, London, Clarendon Press, 1939.

[7] Kolmogorov, A. N. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, Springer, 1933.
[8] Koopman, B. O. The axioms of algebra of intuitive probability, "Ann. of mathem.", 41, p. 269-292, 1940.
[9] Mises, R. von. *Wahrscheinlichkeit, Statistik und Wahrheit*, Berlin, Springer, 1936.
[10] Neyman, J. L'estimation statistique traitée comme un problème classique de probabilité. "Actualités scientifiques et industrielles", No. 739, p. 25-57, 1938.
[11] Neyman, J. *First course in probability and statistics*, New York, H. Holt and Co., 1950.
[12] Neyman, J., E. L. Scott and C. D. Shane. On the spatial distribution of galaxies; a specific model. "Astrophysical Jl.", 117, p. 92-133, 1953.
[13] Neyman, J., E. L. Scott and C. D. Shane. Statistics of images of galaxies with particular reference to clustering. In: "Proceedings of the third Berkeley symposium on mathematical statistics and probability, 1955," Vol. III, Berkeley (Cal.), Univ. press, 1956. p. 75-111.
[14] Robbins, H. An empirical Bayes approach to statistics. In: "Proceedings of the third Berkeley symposium on mathematical statistics and probability, 1955," Vol. I, Berkeley (Cal.), Univ. press, 1956. p. 157-164.
[15] Schmetterer, L. *Einführung in die mathematische Statistik*, Berlin, Springer, 1956.
[16] Wald, A. *Statistical decision functions*, New York, J. Wiley and Sons, 1950.

Résumé: Le sujet traité se rapporte à l'ultime phase de la recherche scientifique, la phase qu'on désigne souvent par le terme "raisonnement inductif". Une analyse des processus mentaux qui accompagnent cette phase indique qu'ils se réduisent à une combinaison des éléments suivants: (i) un effort d'imagination et de mémoire pour passer en revue les hypothèses possibles qui puissent expliquer les phénomènes observés; (ii) les déductions de ces hypothèses et (iii) un acte de volonté d'agir conformément à une des hypothèses considérées. Par conséquent, le terme 'raisonnement inductif" semble déplacé et le terme "comportement inductif" paraît préférable pour désigner la phase (iii).

Les efforts faits pour préciser le concept de raisonnement inductif conduisent aux formules dogmatiques désignées pour prescrire universellement les degrés de confiance qu'un chercheur devrait attribuer aux différentes hypothèses considérées. A cause du caractère subjectif des degrés de confiance, les tentatives d'établir ces formules comme des normes universelles des attitudes, conduisent à des controverses animées. D'autre part, si l'on se rend compte du fait que la suite nécessaire d'une expérience donnée est une action de la part du chercheur, on voit immédiatement que ces disputes sont inutiles. L'avantage de chaque action considérée dépend des attitudes individuelles du chercheur et est une fonction des circonstances inconnues, qu'on désigne par le terme "état de l'univers". Si le choix de l'action se fait par une méthode déterminée utilisant les valeurs observées de quelques variables aléatoires dont la loi dépend de l'état de l'univers, l'action choisie devient aléatoire elle-même, ainsi que l'avantage gagné par cette action. Le problème de la statistique mathématique, considérée comme la théorie mathématique du comportement inductif, est de déterminer les règles d'ajuster notre comportement aux valeurs des variables aléatoires observées de telle manière que la loi de probabilité de l'avantage à gagner soit "optimum". La définition de cet "optimum", ainsi que la définition de l'avantage d'une action en fonction de l'état de l'univers, sont les circonstances non-statistiques de la situation et restent à la disposition du chercheur, affecté par ses propres attitudes et croyances individuelles.

La notion (mais pas le terme) du comportement inductif a été connue à Laplace. Suivant Laplace, elle a été adoptée par Gauss. L'adoption du carré de l'erreur d'une estimation statistique comme la mesure du désavantage de cette estimation, et l'adoption de l'espérance mathématique du désavantage d'une estimation comme la mesure du désavantage d'une méthode d'estimation sans biais, ont amené Gauss au développement de la méthode des moindres carrés. Il semble probable que le traitement moderne des problèmes statistiques du point de vue du comportement inductif conduira à des développements futurs d'une importance comparable.