SCIENCE

VOL. LXIV JULY 16, 1926 No. 1646

CONTENTS

Empiricism and Rationalism: PROFESSOR EDWIN B. WILSON 47

Sidney Irving Smith: PROFESSOR A. E. VERRILL 57

Scientific Events:
The Conversazione of the Royal Society; The Russian Mining Congress; An Ophthalmological Laboratory for Harvard University; Research by Foreign Scholars at Yale University; Biological Abstracts; The Fourth American Association Prize 58

Scientific Notes and News 61

University and Educational Notes 63

Discussion:
The Structure and Formation of Bast Fibers in Flax: DONALD B. ANDERSON. *Should New Fossil Species be described from Wells?* HUBERT G. SCHENCK. *The Price of Honor:* MAYNARD SHIPLEY 64

Quotations:
Chemistry and Disease 66

Scientific Books:
Davison's History of British Earthquakes: PROFESSOR BAILEY WILLIS 66

Scientific Apparatus and Laboratory Methods:
Methods for the Determination of Replaceable Bases in Soils: P. S. BURGESS and J. F. BREAZEALE 69

Special Articles:
Analysis of the Color of the Skin and its Significance: DR. CHARLES SHEARD. *Ciliary Activity of the Oyster:* PROFESSOR T. C. NELSON 70

Science News viii

SCIENCE: A Weekly Journal devoted to the Advancement of Science, edited by J. McKeen Cattell and published every Friday by

THE SCIENCE PRESS

Lancaster, Pa. Garrison, N. Y.

New York City: Grand Central Terminal.

Annual Subscription, $6.00. Single Copies, 15 Cts.

SCIENCE is the official organ of the American Association for the Advancement of Science. Information regarding membership in the Association may be secured from the office of the permanent secretary, in the Smithsonian Institution Building, Washington, D. C.

Entered as second-class matter July 18, 1923, at the Post Office at Lancaster, Pa., under the Act of March 8, 1879.

EMPIRICISM AND RATIONALISM[1]

MEMBERS of the Harvey Society, ladies and gentlemen, I have been asked to discuss the proper manner of treating data from a statistical and mathematical viewpoint and I have chosen as the precise wording of my topic the more general formulation "Empiricism and Rationalism," to the end that I might emphasize a distinction in point of view between methods, and more generally between aims, in the treatment of data by statistical or mathematical analysis. For I believe that without a keen appreciation of the distinction between empiricism and rationalism it is impossible properly to understand the problem of the treatment of observational material.

When we seek for definitions of empiricism or rationalism we may well turn to the Century dictionary in which the philosophical definitions were formulated by Charles S. Peirce, an expert in making refined physical observations and in reducing them, and a great logician and philosopher. I understand that in the medical sense empiricism is quackery, so at any rate the Century dictionary states, but this part of the definition may not be due to Peirce. We find the following:

> Empiricism—3. The metaphysical theory that all ideas are derived from sensuous experience—that is, that there are no innate or *a priori* conceptions.

And again:

> Rationalism—3. In metaphysics the doctrine of *a priori* cognitions, the doctrine that knowledge is not all produced by the action of outward things upon the senses but partly arises from the natural adaptation of the mind to think things that are true.

You will notice the difference between these definitions. It isn't that empiricism emphasizes the importance of sensuous experience. It is that it states that all ideas are so derived and that there are no innate or *a priori* conceptions. This notion is not unfamiliar; one finds it expressed by a good many writers, and particularly by writers in the biologic fields. Some seem to hesitate a little bit at the extreme form of the statement and to qualify it by some sort of assumption that there may be an inheritance of ideas, so that empiricism should be stretched to

[1] Lecture delivered before the Harvey Society on February 6, 1926.

include not only the sensuous experience of the individual but the sensuous experience of the race as transmitted to the individual. It seems to me that if one so stretches the notion one might almost as well give it up; because it is hard to see wherein sensuous experience derived through evolution of the race should differ from innate or *a priori* conceptions. In fact, one might almost maintain that innate and *a priori* conceptions are precisely the quintessence of the sensuous experience of the race. We shall therefore cleave to the original extreme form of the statement that all ideas are derived from sensuous experience and that there are no innate or *a priori* conceptions.

Rationalism, on the other hand, does not say that all knowledge arises from the natural adaptation of the mind to think things that are true. It states that there are *a priori* cognitions, that knowledge is not all produced by the action of outward things upon the senses but partly arises from the natural adaptation of the mind to think things that are true. It is therefore not precisely the antithesis of empiricism. That antithesis would be found more nearly in the extremist interpretations of the idealism of Berkeley where the existence of external things is made to depend on their perception by the mind. Rationalism is a sort of middle ground and as such might readily be assumed to be nearer the truth than either extreme, empiricism or idealism.

We are very prone to extremes and I would not deny that very much advance in science and in philosophy and in art has been made by the struggles of the extremist of one sort or another to prove that a single point of view is adequate for the systematic formulation of a philosophy. As a matter of fact the extremists on both sides are apt somewhat to ridicule the moderate position of any one who occupies intermediate ground; he is, so to speak, between two fires. He has perhaps not the same initiative of attack, not the same uncontrolled zeal of the extremist and this constitutes for him a certain weakness or vulnerability. We are prone to follow special pleaders, whether in religion or in science or in ethics. I might liken empiricism to one end of the spectrum, let us say, the infra-red, and liken idealism to the other end, the ultra violet, and then I should characterize rationalism as constituting the visible light. And I have an idea that we can not see nature whole in any monochromatic light, whether visible or invisible. Our own interests may be important, but so are the other interests of other persons.

One aim of statistical and mathematical analysis in the treatment of observations is the empirical aim of describing our experience. If we have a large number of observations we may wish to describe them by certain characteristics of the whole group. This leads to using the mean or median to express the center of the group or rather some center of the group. We use other constants, for example, the standard deviation or the probable error or the interquartile range to express a measure of the scattering of the individuals of the group from their center. We can determine other characteristic constants of the group. This is purely descriptive statistics. Its value lies in enabling us to replace the great variety of the group of observations by a lesser variety of somewhat technical descriptive constants computed from the elements of the group.

In other types of problems we need the empirical equation. We have one variable which depends more or less upon another and we make a plot to show the values of one variable coordinated with those of the other. If the values run fairly smoothly we draw a curve threading among them in such a way as to satisfy our esthetic judgment as to the probable relation between the variables. For many purposes such a graphical delineation of the smoothing process may be adequate. But even when it is adequate and in many cases when it is not we have recourse to the empirical equation—which means that we select some type of mathematical expression which in a general way runs along the graphical curve and which contains a certain number of parameters that may be assigned, by one method or another, such values as to make the analytical expression lie extremely close to the observations.

In case there is a great deal of scattering among the observed relationships such as we should find, for example, if we undertook to plot the heights and weights of different individuals, we may have recourse to decidedly complicated methods of calculating what we consider to be the best curve to represent the relation between these variables when abstraction is made from the accidental variations of each variable. This field of effort may be generally subsumed under the title of correlation. We should not restrict this definition to imply that the regression equations need be linear.

In all these cases, whether we are content with representing the characteristics of a group by a few statistical constants, whether we describe the cogrediency of a pair of variables by a graphical or analytical smooth curve, or whether on account of the greater scattering we combine these two notions into the general notion of correlation we are still in the domain of description or of empiricism. We are in the domain which is represented, for example, in botany by the herbarium with the dried plants attached to the sheets with their appropriate descriptions and filed away for reference. We are in a museum.

There is, on the other hand, the rationalistic point of view in almost all science, namely, the effort to apply original thought to the explanation of the relationship between variables. In a certain sense an explanation means a search for causes, and in a certain sense one may maintain that there are no causes; that throughout nature there is only concomitancy; that those who speak in terms of forces and causes are merely using a different kind of description or a different extent of description from the frank empiricists; but certainly the aim of the person who undertakes to discover natural laws, so-called, is somewhat different from the aim of him who undertakes to describe. Their methods also differ. Ordinarily the empiricist multiplies description until it becomes more and more realistic. Ordinarily progress in the rationalistic direction is made by ignoring the lesser variations which may be assumed to be due to accident, or at any rate to lesser causes, and by focusing the attention upon an ideal situation where only a few major causes are working; that is, rationalism proceeds by idealization, whereas empiricism proceeds by realization. For the rationalist it may be a positive handicap to know too much in detail the relations which exist in nature. Often the great generalizations come early. Isaac Newton perhaps had a simpler problem before him when he had the observations of planets as reduced by Kepler and systematized into the three laws of Kepler than he would have had if he had been in possession of knowledge of all the multifarious perturbations introduced in the orbits of each planet by the influences of all the others. You can think of many such cases in the biologic field.

This crucial notion of the rôle of idealization in the discovery of natural law may be exemplified by any number of instances. Consider, for example, the question of motion and of force. The fact of observation is that all moving bodies come to rest unless some effort is expended in maintaining the motion. Prior to the time of Newton this universal experience was interpreted as meaning that a forward force was acting on all uniformly moving bodies. Newton said, No, that which stops the body is in the nature of a resistance, bodies left quite alone must persist in uniform motion. Such an idealization requires insight. It may be doubted whether Newton got it from his sensuous experience. It is possible that he contributed this idea, and that we are here in the presence of a mind especially adapted to penetrate behind the deceptions of things as they seem and to think things as they are. Lavoisier's law of the indestructibility of matter or conservation of mass is another case of reversing the obvious to find an idealization. Fortunately for the advance of science the reversal of an accepted point of view is not necessary to the discovery of a law of nature, but a persistent intensity of original thought directed toward the formulation of an ideal situation undisturbed by accessory happenings does seem essential. Moreover, one must have the intuition to decide rightly what is accessory and what is fundamental in the problem considered. And further, he must have a feeling for what are the present problems that are worth while.

So long as persons merely observe nature, howsoever intently, and describe, howsoever accurately, that which they observe they experience real difficulty in discovering natural laws and in confirming their discoveries. This is due to nature's infinite variety. It is the experimental method which has so advanced science by leaps and bounds. The experimenter can somewhat control conditions, he can limit the accessory variations, he can repeat and vary his experiments until a general inference becomes possible.

I believe that Maxwell, contemplating the great complexity of the spectrum, once remarked that given a mathematician of sufficient ability a wonderful contribution to our understanding of the constitution of matter could be made by the mathematical analysis of the spectrum. Scientific history now tells us that better experiments, sharper eliminations of the complexities, closer attention to the simplest cases, proper and new coordinations of idealized physical concepts and relatively simple mathematics have set us on what we believe to be the right track. This is, I venture to think, the usual way of advance—idealization, a recombination, sometimes a reversal, of scientific concepts, new experiments, and a little mathematics. It is the breeders, Mendel with peas or Morgan with Drosophila, who urge genetics forward, not the sociologist or statistician. The place for complicated mathematics is in the follow-up, in the codification of the whole field.

What is mathematics? Every mathematician knows, but few others realize that mathematics is but the details of the tree of logic. Indeed, one may say that mathematics form the branches, the twigs, the leaves of the tree of deduction of which the trunk is our everyday logic and the roots are those dark intricacies of the professional logician. Mathematics is not science; it is not nature, unless it be in the nature of the mind; it is not concerned with the truth but only with the exactness of the deductive process. The confusion of many a scientist with respect to what mathematicians can do for his subject is due to the historic fact that in the past mathematicians have been for the most part interested in the application of their methods to natural phenomena, they have been astronomers, physicists, physiologists working with analysis as a tool. Mathematics is an affair not of empiricism nor of rationalism but of idealism. Once a scientific problem has been formulated in exact quantitative premises which may be converted into formulas, mathematics may go at its deductive processes and may or may not arrive at valuable conclusions from those premises. Its chief use is in the follow-up, in the codification, in the elaboration of scientific advance, not in the original discovery. And as far as its limited use in discovery goes, it must be mathematics in the head rather of the scientific discoverer than of the mathematician.

If you rightly seize my point you will know the importance of having any young man who may contemplate a life devoted to rationalistic science acquire in his college course a knowledge of mathematics through the elements of the differential and integral calculus. If he is fortunate he will need that knowledge sooner or later. He may meanwhile have forgotten most of the detail he once learned, but under the stress of necessity and the stimulus of his problem he can recall as much of the general principles as he is likely to need. Happily that type of mind from which spring the rationalistic advances of science is usually so constituted that it can acquire even in maturity those simple mathematical notions which may have become indispensable; it can almost invent them for the concrete instance at hand. Ronald Ross and Galton are examples; but the process of invention is harder than that of recollection. And again, if you have seized my point you will understand why in acceding to your request that I discuss the statistical and mathematical methods of treating data, I have come to you with general ideas, with points of view, with distinctions in aims, instead of with formulas. All formulas are technical details of problems already formulated, and it is the formulation which is at once the more difficult and the more important part. Runaway mathematics is like a runaway horse in doing nothing but harm to itself and others; it is irrational.

To return to our subject. We have mentioned empirical equations and laws of nature which may be converted into formulas. What is the distinction between them? If we were interested in smart dialectic we might launch forth on a demonstration that there is no such distinction, that from the most empiric of equations one may proceed by imperceptible gradations to the most universal of natural laws as from darkness to light. But this tour de force would not be a useful contribution to our present discussion. Better is it to contrast the more empiric with the more rationalistic. An empirical equation is a mathematical expression containing in addition to the variables certain parameters or adjustable constants which may be so chosen as best to represent the data. For example, suppose you plot the complete expectation of life at birth at different times, say every decade, from 1850 to 1920. You will naturally take the time as the abscissa or horizontal variable and the expectation of life as the ordinate or vertical variable. You all know what you will see, namely, a series of points rising fairly steadily from a value between 40 and 45 to a value between 55 and 60, representing an increase of about fifteen years in expectation in seventy years. You may smooth this series of points graphically. You may fit some equation to them, for example, a linear equation, as "expectation = a + bt," where the time t is measured in decades from 1850, and where a and b are the two parameters to be chosen to fit the data, as expectation = 43 + 2t. Such an equation is empirical. It describes roughly the variation of expectation that has been experienced during the time in question. If the graph appears to show a general curvature with an upward acceleration and if you are a little more adept in curve-fitting you may venture to try a parabolic or quadratic form of relationship such as "expectation = a + bt + ct²" and on determining the constants a, b, c by any method, say by a simple trial until you have as satisfactory a shape for the curve as seems possible, you will again have an empirical equation to represent the facts.

Nobody, however, should now assume that he has discovered a veritable law of nature. It is not a law of any generality that the increase of expectation is linear, is proceeding universally and always has proceeded at a uniform rate, or with a uniformly accelerated rate. It may be interesting, but it is not scientifically valid, to produce or extrapolate the empirical

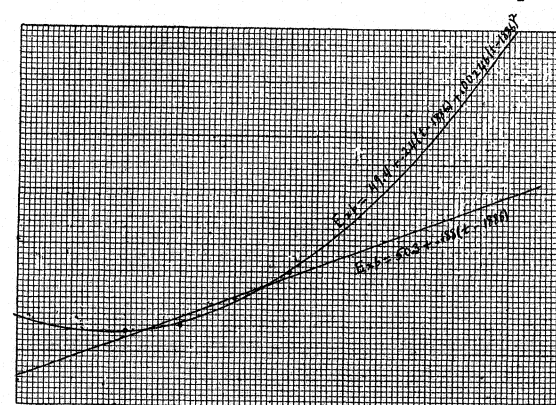

FIG. I

equation forward to 2000 A. D. or backward to 1800 to ascertain what are the expectations of life for babies born as of those dates. No analysis has been made of causes, no detailed analysis has been made of the ways in which the change of expectation has come about. There has been merely a gross description of a total phenomenon; it is graphic, but not the basis for an induction. Figure I represents such a treatment of some Swedish data which are at hand.

EXPECTATION OF LIFE AT BIRTH IN SWEDEN AT SPECIFIED DATES

Fitted by I. $\text{Exp} = a + bt = 50.3 + .188\ (t - 1886)$
and by II. $\text{Exp} = a + bt + ct^2 = 49.4 + .24\ (t - 1886) + .00246\ (t - 1886)^2$

Year	Expectation	Calc. by I	Diff.	Calc. by II	Diff.
1846	44 years	42.8	−1.2	43.8	−0.2
1866	45 "	46.5	+1.5	45.6	+0.6
1886	50 "	50.3	+0.3	49.4	−0.6
1896	52 "	52.2	+0.2	52.0	0.0
1906	55 "	54.1	−0.9	55.2	+0.2

You see that the linear fit depending on two parameters is not bad, that the quadratic description depending on three parameters is excellent and that either, when produced backward or forward, gives within relatively short lapses of time results which themselves challenge belief in the generality of the description.

There are in the literature many instances of this sort of treatment of data, and I am tempted at times to fear that they may do more harm than good—not that in themselves they are bad, but because they allure readers toward illegitimate generalizations and forecasts. There is a physician of my acquaintance who makes it a rule when reporting his cases in print never to express his results in percentages. His reason is that figures in percentage imply a generality in his experience which he neither posits nor feels. The emotional reaction to a mathematical, even to a simply arithmetic statement, is often extreme. Many persons like to tease others into the spinning of yarns; under such provocation figures and formulas make first-rate liars; but why one then takes them so seriously I can not imagine.

As another illustration of an empirical equation we may take from the field of epidemiology Farr's Law, with which you are all familiar. The law states that, other things being equal, the death-rate increases with the density of population, and that in particular it increases as a root, say the tenth root, of that density. Other things never are equal, and the law is of relatively little use in estimating the death-rate from the density of population. Moreover, it is obviously a law of restricted application because it gives impossible results when extended to practicable differences of density. For example, there are rural districts in which the number of persons per acre is certainly not one thousandth of that in this great metropolis. The tenth root of 1000 is 2. Yet we should not expect to find a death-rate here, even when carefully adjusted to the age distribution of the population, twice as high as in those rural districts. Nevertheless, with reasonable allowances, there is a considerable degree of truthful generality covered by Farr's Law. Insofar as the law is sound it should have, even though empirically derived, some rational basis. Rationalistic science advances not merely, perhaps not mainly, by the discovery of laws through reasoning, but rather more by the invention of reasons for those uniformities or generalities which may have been observed. One may attempt to rationalize Farr's Law. He may point out that persons can not die of diseases they do not have, that many of our illnesses are infections and that the chance of contracting such diseases is greater in a denser population, so that the death-rate, other things being equal, might have been expected to be higher with the increase of contacts and the faster and further spread of infections. One might show that pneumonia is an urban more than a rural disease.

Such arguments and illustrations abound in the literature and indeed in our everyday thinking; they are evidences of a well-nigh universal tendency, an admirable tendency, of the mind to rationalize our experiences; they are enticing but must be critically examined. To avoid the pitfalls of plausibility, persistent thinking and varied experience must be applied. Rationalism is not thought *per se*, it is thought applied to observations. In an examination of the situation with reference to cancer in Massachusetts conducted during the past summer, it was found that in the main the mortality from cancer obeyed Farr's Law. It nearly doubled from the districts of lowest to those of reasonably high density of population, but did not seem further to increase at the highest densities. This may not play into the hand of a rationalization of the law upon the basis of contacts and infection; it may point to general living conditions with reference to light and air and outdoor work as contrasted in rural and urban living and dying. Premature or immature rationalization of empirical formulas is to be rated with the extrapolation of such formulas as an error of empiricism, and as one which approaches to quackery.

The statement of a law of nature as a formula ordinarily contains certain constants or parameters. Thus the gas law $pv = Rt$ uses the gas constant R. Some constant is necessary because of the arbitrariness of the units in which pressure, volume and absolute temperature may be measured. We know that with the units customary in physical chemistry the gas constant R is about two calories; its value when any other system of measures is used can be obtained by simple arithmetic. Such a constant is called a constant of nature or a universal constant. Its value has of course to be obtained from experiments and in this sense it is a fitted constant; but once determined it is fixed and must be assumed whenever the gas equation is used; it is not an adjustable parameter of the sort found in purely empirical equations. Again, the law of gravitation states that the attraction varies as the product of the masses and inversely as the square of the distance. Converted into a formula, $F=\gamma m_1 m_2/r^2$. Here the multiplier γ is a universal constant. The values of the masses may be determined by experiment or, in celestial mechanics, by observation, and in this sense those values are fitted constants, but they are not general empirical parameters; you may not use one value of the mass of the earth in the solution of our motion about the sun and then use a different value in treating the motion of the moon about the earth simply because a different value may give a better fit. If Farr's Law that the death-rate is equal to a multiple k times the nth root of the density of population were a full-fledged law of nature both k and n would be universal constants, of which the first would depend in numerical value on the units employed in expressing death-rates and densities, and n, the root, would be a specified number. The search for laws of nature is in part a search for universal constants.

It should go without saying that most of our natural laws are but fledglings. They or their followers will mature with further investigation and in due time. As they stand to-day, particularly in fields of science just coming under quantitative treatment, they will contain fittable parameters. Indeed the further advance of science may demonstrate that some of the constants must remain fittable. This may be illustrated by recalling Kepler's Law that the planets revolve about the sun in ellipses with the sun at one focus. He left the eccentricity, the orientation and the size of the ellipse undetermined. These remain adjustable constants to this day. The law of gravitation does not supply them. They are constants of integration, constants which arise in passing from the differential equations of celestial mechanics, equations involving accelerations, to the integrals which express first velocities and then positions. Although as early as Galileo and Borelli some attempt at deductive physiology was made, and many efforts have been since directed and are even now most intensively being prosecuted toward such an end, it is too much to hope that we shall soon reach a physiology as rationalistic as mechanics. The living organism is not so simple as a chunk of lead. Some very readable reflections of a general nature are put together in D'Arcy Thompson's "Growth and Form." We should not be impatient either of general reflections or of empirical equations; if properly pondered both will mightily help us forward. We are on our way. Whither we go we know not, but the way we must know.

The essential difference between an empirical equation and a law of nature is that the former is a description of our observations, whereas the latter involves induction from the observations. Now induction is the object of the experimental as contrasted with the observational method. A person who performs experiments merely for the sake of describing them seems unnecessarily to be shielding his eyes from the bright visions with which nature on every side surrounds him. We are content to take the limited controlled observation of the laboratory only because we have the greater aim of induction in view. Even though we expect that the induction will fall far short of that degree of universality which we ennoble with designation as a law of nature, we hope to elicit from the immediate occasion something which shall permit us to seize some part of its general significance, some iota of its real meaning. The significant realities of nature are nature's uniformities. At times and to individuals the problem of induction has seemed reasonably simple, a sort of correlative of deduction. Such opinions appear hardly tenable. I can not pursue the discussion here but may refer you to passages in Keynes's "Treatise on Probability" and Whitehead's "Science and the Modern World," and to my DeLamar lecture delivered two years ago at the Johns Hopkins University (Science, 63, 1926, pp. 289–296). My feeling is that unless one is willing to regard induction as a mere matter of chance, a lucky strike in the dark, he must believe it to be the original contribution of a mind lighted by an antecedent rationalism.

You asked me to speak of the statistical methods of treating data. I wish you had not. It is a mean subject. Those of you who have read the biography of the great Lord Rayleigh by his son will recall his statement that he does not believe in statistical methods, that the object of repeating an experiment is to judge of the control acquired, that he even doubts the utility of averaging values to obtain a mean, though he admits that this is carrying disbelief rather far. We find very little statistical analysis in experimental physics or chemistry to-day, a smaller relative amount, I think, than was found a generation ago; and even in astronomy, for which the method of least squares was developed by Gauss and in which it was universally applied in the past, there is a strong tendency to short-cut formal statistical processes. It is now to the biologist or economist that you must go for complicated statistical analysis. Why this state of affairs? May it perhaps lie in a contrast of the experimental and observational methods, in a difference of degree of attainable control? Shall we say that when the control is good, when we are working in a field in which control is easy or when we are sufficiently astute or fortunate to design experiments so that those consequences in which we are interested are independent of the other variations, then we have no need of statistics and can go along with Lord Rayleigh? Shall we admit that statistics belongs rather in the field of observation and serves to replace control when that is not attainable or is repugnant to the nature of the investigator?

We must think here not so much of statistics as a method of description, but more of it as a basis for induction. If we believe that induction requires rationalism in experimental procedures which are relatively exact, we must feel that it requires an even higher degree of insight and thoughtfulness, of care and of diffidence when maneuvering in crude and complicated realms. It pays to begin the analysis with simple methods, to make diagrams, to rearrange the data and make new pictures, to dwell with the material until one knows its excellencies and its defects in respect to those particular items that may be important for the conclusions, to take into account not alone the data themselves but any general scientific considerations that may be germane to the induction. This attitude any one may cultivate; it is consonant with the admonitions of Bowley and of Westergaard to eschew elaborate processes which carry one out of touch with the original figures and which by their very elaboration give a false sense of security.

ACETIC ACID

EXPERIMENTS OF NEUMANN

		Temperature								
		78°	100°	110°	120°	130°	140°	150°	160°	185°
A	Pressure	...	393.5	411	432	455	477	498.5	...	565
	D. calc.	...	3.39	3.23	3.06	2.90	2.75	2.61	...	2.28
	D. obs.	...	3.44	3.31	3.14	2.97	2.82	2.68	...	2.36
	Exc. of D. obs.	...	+.05	+.08	+.08	+.07	+.07	+.07	...	+.08
B	Pressure	...	342.3	359.3	377.5	398.5	417.5	436.5	...	495
	D. calc.	...	3.35	3.18	3.02	2.85	2.70	2.57	...	2.26
	D. obs.	...	3.37	3.22	3.06	2.89	2.75	2.63	...	2.31
	Exc. of D. obs.	...	+.02	+.04	+.04	+.04	+.05	+.06	...	+.05
C	Pressure	...	258	...	...	...	...	...	...	382
	D. calc.	...	3.26	...	...	...	...	...	...	2.22
	D. obs.	...	3.17	...	...	...	...	...	...	2.25
	Exc. of D. obs.	...	−.09	...	...	...	...	...	...	+.03
D	Pressure	...	232	...	252	274	287.5	300	...	335
	D. calc.	...	3.23	...	2.87	2.72	2.58	2.46	...	2.21
	D. obs.	...	3.12	...	2.94	2.68	2.54	2.44	...	2.23
	Exc. of D. obs.	...	−.11	...	+.07	−.04	−.04	−.02	...	+.02
E	Pressure	164	186	197	209	221	232	243	253	269
	D. calc.	3.53	3.15	2.97	2.81	2.65	2.52	2.41	2.32	2.18
	D. obs.	3.41	3.06	2.91	2.75	2.61	2.50	2.40	2.31	2.22
	Exc. of D. obs.	−.12	−.09	−.06	−.06	−.04	−.02	−.01	−.01	+.04
F	Pressure	149	168	...	...	201	...	...	...	...
	D. calc.	3.50	3.12	...	...	2.62	...	...	...	...
	D. obs.	3.34	3.01	...	...	2.56	...	...	...	...
	Exc. of D. obs.	−.16	−.11	...	...	−.06	...	...	...	...
G	Pressure	137	156	166.5	180	188	199	208.2	...	230
	D. calc.	3.48	3.09	2.92	2.75	2.60	2.47	2.37	...	2.17
	D. obs.	3.26	2.98	2.81	2.61	2.50	2.40	2.29	...	2.14
	Exc. of D. obs.	−.22	−.11	−.11	−.14	−.10	−.07	−.08	...	−.03
H	Pressure	113	130	138.5	149	157.5	168.2	175	...	191.5
	D. calc.	3.42	3.03	2.85	2.69	2.55	2.43	2.33	...	2.15
	D. obs.	3.25	2.94	2.78	2.60	2.47	2.32	2.26	...	2.13
	Exc. of D. obs.	−.17	−.09	−.07	−.09	−.08	−.11	−.07	...	−.02
J	Pressure	80	92	98.5	106	112.5	117.3	...	129.2	...
	D. calc.	3.32	2.91	2.73	2.58	2.45	2.35	...	2.21	...
	D. obs.	3.06	2.76	2.61	2.46	2.34	2.27	...	2.11	...
	Exc. of D. obs.	−.26	−.15	−.12	−.12	−.11	−.08	...	−.10	...
K	Pressure	66	77.7	84	89.5	93	98	103	...	110.5
	D. calc.	3.26	2.85	2.68	2.53	2.40	2.31	2.24	...	2.12
	D. obs.	3.04	2.66	2.49	2.37	2.32	2.24	2.16	...	2.11
	Exc. of D. obs.	−.22	−.19	−.19	−.16	−.08	−.07	−.08	...	−.01

Except by accident, mathematical conclusions are no sounder than the premises that went into the formulation of the problem, and they may be much weaker because of those accessory assumptions which are so apt tacitly to creep in during the progress of the work. If resort must be had to the technical processes of correlation, it is well first to master the treatment of association in Yule's "Theory of Statistics," and it is important to bear in mind the words of R. A. Fisher in his excellent "Statistical Methods for Research Workers," namely:

> If we choose a group of social phenomena with no antecedent knowledge of the causation or absence of causation among them, then the calculation of correlation coefficients, total or partial, will not advance us a step towards evaluating the importance of the causes at work. . . . In no case can we judge whether or not it is profitable to eliminate a certain variate unless we know, or are willing to assume, a qualitative scheme of causation.

This means antecedent rationalism.

It may be instructive to illustrate the method of correlation on material from physics or physical chemistry, even though the physicist would not apply it. Willard Gibbs, in the course of his memoir on the equilibrium of heterogeneous substances, developed from theoretical considerations a formula for connecting the density *D*, pressure *p*, and absolute temperature *t* of a mixture of gases with convertible components, and tested it in a rough way on available data. Shortly thereafter in a paper on the vapor-densities of certain substances he returned to the matter in detail. His result was

$$\log \frac{D_1(D-D_1)}{(2D_1-D)^2} = -A - B \log t + \frac{C}{t} + \log p,$$

where D_1 is the density of the rarer component and may be computed from the molecular formula. This equation contains three constants A, B, C; but of these B is connected with the specific heats and is therefore not fittable in the empirical sense. He showed, however, that, for the cases he was treating, the term B log t could be neglected, provided its omission was compensated in the determination of A and C. The result for acetic acid was

$$\log \frac{2.073\ (D-2.073)}{(4.146-D)^2} = \frac{3520}{t_c+273} + \log p - 11.349$$

when the two constants $C = 3520$ and $A = 11.349$ were derived from the experiments of Cahours and Bineau, which he appears to have judged to be the best then attainable.

Calculations from this numerical equation were then compared by Gibbs with sixty-five experiments of Neumann over a wide range of temperatures and pressures. The result of the comparison is given in the table. There are divergences, systematic as well as accidental. The discussion which Gibbs gives to this matter is interesting. He can discuss it because he has a rational formula based on excellent determinations. If he had had merely an empirical formula fitted directly to these observations of Neumann, his discussion would have had to be changed, if indeed it could have been given at all. If we apply correlation, which is a method of least squares, to the figures of Neumann we shall inevitably weight most heavily those observations which tend to depart most from the linear regressions automatically set up by the method. This is proper if the departures are accidental; if they are due to systematic errors we shall be led off the track both of a correct rational explanation of the phenomenon in question and of a sound criticism of the observational material.

The tables necessary for the treatment of Neumann's data by the method of correlation are:

TABLE A

PRESSURE AND DENSITY

Pressure / Density	50–99	100–149	150–199	200–249	250–299	300–349	350–399	400–449	450–499	500–549	550–599	Totals
210–224	1	3	1	1	1	1						8
225–239	2	2	2	1	1		1		1		1	11
240–254	1	1	3	2	1	1						9
255–269	2	1	1	2	1			1	1			9
270–284	1	1	1	1				1	1			6
285–299		1	2		1		1		1			6
300–314	2		2	1			1	1				7
315–329		2			1		1					4
330–344		1	1			1	1	1				5
Totals	9	12	13	8	6	3	5	4	4	0	1	65

TABLE B

TEMPERATURE AND DENSITY

Temperature / Density	78	100	110	120	130	140	150	160	185	Totals
210–224						1	1	1	5	8
225–239				1	2	2	2	1	3	11
240–254			1	1	2	3	2			9
255–269		1	1	2	3		2			9
270–284		1	2	1		2				6
285–299		2	1	1	2					6
300–314	2	3		2						7
315–329	2	1	1							4
330–344	2	2	1							5
Totals	6	10	7	8	9	8	7	2	8	65

TABLE C

TEMPERATURE AND PRESSURE

Temperature / Pressure	78	100	110	120	130	140	150	160	185	Totals
50–99	2	2	2	1	1	1				9
100–149	3	1	1	2	1	1	1	1	1	12
150–199	1	3	2	1	2	2	1		1	13
200–249		1		1	2	1	2		1	8
250–299		1		1	1	1	1	1	1	6
300–349		1					1		1	3
350–399		1	1	1	1				1	5
400–449			1	1		1	1			4
450–499					1	1	1		1	4
500–549										0
550–599									1	1
Totals	6	10	7	8	9	8	7	2	8	65

The biometric constants computed from these tables are:

Mean density	= 2.682	Mean temperature	= 128.3	Mean pressure	= 229.6
Stand. Deviation: Density	= .375	S. D. temperature	= 30.0	S. D.: Pressure	= 126.8
S. D. of mean	= .046	S. D. of mean	= 3.7	S. D. of mean	= 15.6
S. D. of S. D.	= .33	S. D. of S. D.	= 2.6	S. D. of S. D.	= 10.8

The correlation coefficients between pairs of variables are

Pressure-Temperature	$\gamma_{23} = .38$	Density-pressure	$\gamma_{13} = .14$	Density-Temperature	$\gamma_{12} = .-.82$
Standard Deviation is	.10		.12		.04

This indicates no significant cogrediency of density and pressure, only a moderate cogrediency of pressure and temperature, but a very strong contragrediency of density and temperature. The partial correlation coefficients, however, tell a different story.

Correlation of pressure and temperature with density constant	$\gamma_{23 \cdot 1} =$	.88 = .03
Correlation of density and pressure with temperature constant	$\gamma_{13 \cdot 2} =$	.87 ± .03
Correlation of density and temperature with pressure constant	$\gamma_{12 \cdot 3} = -$	.95 ± .01
Standard deviation of density with pressure and temperature constant	$\sigma_{1 \cdot 23} =$	.105
Standard deviation of temperature with density and pressure constant	$\sigma_{2 \cdot 13} =$	9.0
Standard deviation of pressure with temperature and density constant	$\sigma_{3 \cdot 12} =$	58.

Regression equation of density on pressure and temperature

$$\frac{D-2.68}{.105} = .87\ \frac{p-127}{58} - .95\ \frac{t_c-128}{9.0}$$

Regression equation of pressure on density and temperature

$$\frac{p-127}{58} = .87\ \frac{D-2.68}{.105} + .88\ \frac{t_c-128}{9.0}$$

Regression equation of temperature on density and pressure

$$\frac{t_c-128}{9.0} = .88\ \frac{p-127}{58} - .95\ \frac{D-2.68}{.105}$$

(As the calculations are merely for illustrative purpose they have been run through with only slide rule precision and an accumulation of errors may appear in the final results. It is noticeable that $\sigma = .105$ for the root mean square residuals of the departures of the density as calculated from the regression equation fitted to the data happened to be the same value, estimated from the average departure .085, found by Gibbs when comparing his equation, fitted to other data, with this particular series.)

In getting forward with quantitative rationalistic science one of the chief purposes of expressing the experimental or observational results as an equation or formula is to have the relationship in such a form that it may be subjected to mathematical manipulation in combination with other formulas. It should be remarked that a regression equation is unhappily not subject to such manipulation. Even the simple process of solving a linear equation can not be carried out. If we should solve the second of the given regressions for D we should find:

$$\frac{D-2.68}{.105} = 1.15\ \frac{p-127}{58} - 1.01\ \frac{t_c-128}{9.0},$$

which is by no means identical with the first, the proper, linear expression for D, and will give results decidedly wide of the mark. This is awkward, but it is inevitable by virtue of the very nature of such an equation. As a matter of fact a regression is designed to give us the best average value of one variable, say D, when the precise values of the other variables are known. What the rationalist wants for his purposes is the best functional relationship between the variables all treated alike with respect to their different degrees of precision. For many purposes he prefers to know an appropriate type of function, not necessarily linear, than to emphasize any special numerical values, and for the discovery of functional types the method of partial correlation is awkward, to say the least. Such desires and preferences are coordinate with his aim of analyzing relations in an ideal system free from variations other than those of the variables on which he is concentrating his attention. I feel sure that many a physiologist of to-day feels this, even if he does not formulate it, and pursues his course in the manner of Galileo, of Newton, and of Helmholtz rather than after the style of Galton and Pearson.

Let it be clearly understood that I do not condemn the statistical method; in many cases it is indispensable. I come not to condemn, but to analyze. One has to admit that any method, by the very fact that it is a method, may tempt persons unintelligently to confide their fortunes to it to their own destruction (an observation hardly necessary here scarce fifty blocks from Wall Street). The rationalist who becomes too much idealist is peculiarly liable to defeat his own ends, to permit himself to be led too far afield by the imagined beauties of his own speculations, and never find his way back to nature. The illustrious investigator for whom this society is named said of Bacon that he wrote of science like a Lord Chancellor. That you may have less excuse to say of me that I have treated of your problems like a mathematical physicist, I wish at the close of this discourse to say a little about obtaining data; that is decidedly germane to our subject. I am much impressed with the great elaboration of experimental apparatus, especially in the hands of the young investigator. Such a condition is perhaps inevitable here in America, where we manufacture so many doctors of philosophy and where so many of these newer doctors enter on positions really as research assistants to maturer scientists. It is easy for the teacher or investigator to incorporate the candidate or the graduate into his own investigative system, technique and all, to start him with a material equipment comparable to that to which he himself has attained only after years of work and of thought. It is perhaps the selfish thing to do; in some cases it may be a necessary defensive action of the mature student if he is to have any further hope of progress with his own work. But the neophyte has not had those years of work and of reflection; he is not yet in intimate contact with the facts, the irreducible and stubborn facts, of nature. May not too elaborate equipment shield him from such contact? Might it not be better for him at first to perform more qualitative, fewer quantitative, experiments, to range around his field a bit and really become acquainted with it?

The inspiration to a fruitful scientific life comes from seeing nature not through some elaborate intermediary darkly as in a glass, but face to face. Galton's life is interesting not so much for its record of his accomplishments as for the revelations of his freshness of mind. He was a great amateur, and such are the salt of the earth. It is their followers who systematize and reduce to method. Years ago, as I was wandering through the Jefferson Physical Laboratory in search of some former fellow student with whom to exchange ideas, I came upon my old teacher, B. O. Peirce. Said he: "Colonel Wilson, we are all poor physicists here in the Jefferson." "How so?" I asked. "Why, we have to have $14,000 a year to get our uninteresting results when a real physicist would get new stuff with a ball of twine and a jackknife!" Exaggeration, of course; nobody begrudges the Jefferson its budget, it is well spent; it should be larger. The adult must have his means of livelihood. My question regards the young. Is it well too urgently to transform the natural freely imaginative organization of the child's hide-and-seek into the supervised play of the school and the massed phalanx of the stadium under the direction of the professional coach? It is metaphor; but does it not somewhat apply to our conduct of graduate instruction and of initiation into a life of scientific research? Should we not distrust all over-elaboration of method, whether of obtaining or of treating data, whether of apparatus or mathematics or statistics? Howsoever inevitable such development is, should we not be actively on guard lest it lead us and particularly our youth inward to mooning over artificialities instead of out to live with the stubborn facts of a real world. If I knew a young fellow who sought advice about love I should not send him to his room to study Balzac's "Physiologie du Mariage" or Bourget's more ponderous "Physiologie de l'Amour moderne," nor yet to a clinic to be "psyched" à la Freud; I should tell him to go see some girls.

EDWIN B. WILSON
SCHOOL OF PUBLIC HEALTH OF HARVARD UNIVERSITY

SIDNEY IRVING SMITH

PROFESSOR SIDNEY IRVING SMITH, Ph.B., Yale, 1867, was born February 18, 1843, in Norway, Maine, and died May 6, 1926, in New Haven, Conn. He had been in feeble health several years, due to various complications combined with his age. Immediate cause of his death was cancer of the throat. He had been partially blind since 1906, due to hereditary glaucoma. Although surgical operations were made on both eyes, he became totally blind several years ago. Professor Smith married in New Haven, June 29, 1882, Eugenia Pocahontas, daughter of Edward Brady Barber, a music-teacher from Canada. Mrs. Smith died March 14, 1916. There were no children. He is survived by a sister-in-law, Mrs. Clarence M. Smith, of Norway, Maine; his brother-in-law, Professor emeritus Addison E. Verrill (B.S. Harvard, 1862), M.A., Yale, 1867; Major George E. Verrill, '85 S. Yale, U. S. Engineer; A. Hyatt Verrill, *ex*-'91, Yale, Art, artist and author (nephews); Edith B., m. V. Akers; Lucy Lavinia, *ex.* Art, m. S. H. Howe, Jr. (nieces).

His preparatory training was received in the Norway Liberal Institute, and Bethel, Maine, Preparatory School. Professor Smith before coming to Yale had, under the instruction and encouragement of A. E. Verrill, collected and studied about all the flowering plants and ferns of Norway, Maine, and vicinity, discovering many rare species. He always retained his interest in botany and gardening. At the same time he made large collections of the local insects and found many undescribed species, some of them of great interest. His insects, obtained prior to 1864, were purchased by Professor Louis Agassiz for the Museum of Comparative Zoology. After that he collected insects for the Yale Museum. He joined Professor Verrill in various dredging expeditions in Long Island Sound and to the Bay of Fundy in 1864 to 1870, making collections for the Yale Museum, in which he had charge of the Crustacea for many years.

He was assistant in zoology at Yale, 1867–1874; instructor in comparative anatomy, 1874–75; professor of comparative anatomy, 1875–1906; and since then professor emeritus. He had charge of deep water dredging in Lake Superior for the United States Lake Survey in 1871, and for the United States Coast Survey about St. George's Banks in 1872; and was associated with the biological and deep-sea work of the United States Fish Commission, 1871–1887. Later, he gave his share of the deep-sea Crustacea to the Yale Museum.

He was state entomologist of Maine and Connecticut for a number of years and contributed to the annual reports of the Maine and Connecticut Boards of Agriculture. In 1890 he revised the definitions in comparative anatomy in Webster's International Dictionary. He organized and conducted one of the first, if not the first, biological course in this country of studies intended as a preparation for a medical school. He was an excellent teacher, using laboratory anatomical work extensively.

When the first Peabody Museum was planned, in 1875, he and Professor Verrill made all the plans and detailed drawings and specifications for the furniture and exhibition cases on the third floor, and part of those on the second floor of the museum. He was also one of the early promoters of the Biological Station at Woods Hole, Mass., and for several years one of its trustees. He had been a member of the National Academy of Sciences since 1884; and was also a member of various other scientific societies.

He was the author of numerous zoological papers