"Memoirs of the Literary and Philosophical Society of Manchester," 3rd Series, Vol. vi., 1879.

"Beiblätter zu den Annalen der Physik und Chemie," Leipzig, 1880, Band iv., Stucke 7, 11.

Considerations respecting the Translation of Series of Observations into Continuous Formulæ. By C. W. MERRIFIELD, F.R.S.

[Read November 11th, 1880.]

It has been the custom of the Presidents of this Society, upon vacating their office, to address to the members a few observations, usually rather general in character, concerning some of the objects for the consideration of which the Society was founded. I am unwilling to break through this good custom, and, although my daily employment leaves me but little leisure for independent investigation, or for reviewing recent progress in mathematics, it may be that I shall not take up your time uselessly if I lay before you a few considerations, which have occurred to me, respecting the Translation of Series of Observations into Continuous Formulæ.

An observation is one of those simple words, each of which represents a group of ideas, the combination of which is very far removed from simplicity. The mere sight of a phenomenon is not an observation. To make it one, it is necessary to connote other phenomena, which must be connected with it by definite numerical relation, in respect of time, or of place, or of some other of the measurable attributes of space or matter. This connotation has always been a sufficiently difficult matter to settle with any degree of exactness; so much so, indeed, that really good observations are difficult to obtain, especially those which are sufficiently exact, and closely enough connected with others, to enable us to establish or to verify formulæ, or to settle constants or units.

Without attempting to go into such deep reasonings as are involved in the ultimate question of whether continuity is an objective or subjective phenomenon, it is not, perhaps, going too far into generalities to say, that our observation of nature tends to discontinuity, and our discussion of it to continuity; and thus there is a tendency, on mere grounds of facility of reasoning, to assume continuity practically, as underlying most connected series of observations. Some of the

particular forms of the abuse of this tendency are what I propose to mention to you this evening.

To fix our ideas, let us consider the variations in the limited population of an island, taking the population as an ordinate, while the abscissa represents time. Every arrival, whether by birth or immigration, gives an abrupt vertical step upwards; every departure, whether by emigration or death, gives an abrupt vertical step downwards, the population line being horizontal from one step to another. However large the population of the island may be, even if it be equal to that of the whole earth, we have actually and ultimately no curve at all, but a zig-zag line with rectangular flexures. When we represent a large population in this fashion by a small scale drawing, we obtain, upon applying any process which has the effect of sinking the detail, what is apparently a continuous curve. But this appearance of continuity postulates defective perception, and has no existence of its own. Any attempt at minute exactness of representation by a continuous law lands us in absolute falsehood; and if we carry the results of continuous calculation to extremes, either by the use of many figures of decimals, or by high orders of differences, we are departing in reality from the precision which we are striving to obtain by an apparent exactness of process.

The entire population of the earth has been roughly estimated at (1,500,000,000) fifteen hundred millions. Taking this as unity, it is clear that, if we go to nine decimal figures, the law of variation must be essentially discontinuous. The best representation by a continuous curve cannot therefore be put as a general question, but must be strictly limited by some controlling condition. It is a definite question, when we ask by what straight line, or by what parabola of the n^{th} order, the variation may be best represented, provided n be not too large; but when n is taken too large, the question is not merely indefinite, but misleading, and wholly inept.

If we take any m points of this broken line and choose a function of m parameters of any selected form, it is a perfectly definite question to determine the values of those m parameters, so that the continuous function shall meet the broken line in those points. But we can go no further, for the actual broken line has but two directions, the horizontal and the vertical, while its continuous representation has ever-varying direction. To make the question definite even to this extent, the form of the function has to be made matter of assumption. In ordinary cases, it is usual and expedient to assume the function to be parabolic, that is to say, of the form

$$y = a + a_1x + a_2x^2 + \dots + a_nx^n,$$

and this leads to the simple application of Finite Differences. It need

hardly be said that this assumption is not by any means generally suitable, or even possible, as is obvious in the case of periodicity, which it fails to represent at all; and in such cases as the ordinary increase or decrease of population, in which the law approaches that of a geometrical series, the interpolation has to be made upon the logarithms of the ordinates instead of upon the ordinates themselves. But the point to which I wish to call attention just now is, not the choice of form, but the consideration that, in a large class of natural phenomena, which it is usual to treat by the application of continuous laws, the attempt to obtain too high a degree of approximation is not a mere difficulty, but an absolute blunder, arising from an oversight of the real character of the problem.

Assuming the representative continuous function to have been suitably selected, it becomes a question of some interest, to know what order of differences should be used,—or, more generally, how many parameters should be left to be determined from the results of observation,—so that the error of interpolation should be fairly comparable with the departure from continuity. I cannot find that this problem has received much attention, even to the extent of being definitely stated, except in the case of quadratures. The course has rather been, among the older calculators generally, and also among later calculators who have not bought their experience, either to go to an enormous elaboration of terms and figures, or else to seek some form of function which shall represent the data almost absolutely.

As an example of the former, I may quote Abraham Sharp, who calculated common logarithms to 61 figures, and the areas of circular segments to 17 figures.* In the introduction to the segment table, he gives coefficients for bisection, as far as the twentieth mean difference, upon which he goes on to remark:—"The Operation, I presume, will rarely, if ever, proceed so far, having observ'd that ordinarily if the twelfth Difference exceed 5 Figures, the remoter Differences will encrease; consequently the Subdivision cannot be perform'd truly by any other method save by Calculation." This abuse of figures is quite common with young pupils applying arithmetic to mechanics or physics, as most of my hearers have probably not failed to observe in the course of their own teaching. It is, however, by no means confined to students. I have seen the results of an examination, in a group of schools attended by from 20 to 150 pupils each, brought to a comparison by percentages to four figures of decimals, and the distinguished mathematician who conducted the examination found serious fault with me for cutting down his decimals, when they came to be printed, to two figures; which, being percentages, were considered by

* See "*Geometry improved*," by A. S., Philomath (London, 1717, 4to), p. 7.

me to be exactly two figures too many. Again, there exists a table for the comparison of English and metric measures, in rather wide circulation, in which the metre is given in inches, to two integers and six decimals,—this being assuredly a superior limit to any possible accuracy,—while the cubic metre is given to five integers and eighteen decimals. It probably involved less mental effort to perform the multiplication simply, than to reflect that the third decimal place marked the superior limit of accuracy in the comparison of the cubic measures. The practice which I have found best, is to keep as much as possible to the same number of significant figures throughout a calculation. The cases are rare in which this needs to be departed from. It is indeed sometimes convenient to use the figures as if they were exact, and not to curtail them in the arithmetical work, in order to retain such useful verifications as the proof of nines; but that convenience affords no justification for recording a result, carried for mere convenience of checking to more figures than are wanted, so as to affect an impossible exactness. That is quite a different thing from using a figure or two more than the observations justify, so as to make sure of not increasing the probable error of the result by arithmetical inaccuracy.

Some of the recent investigations concerning the ellipsoidal figure of the earth have not unfrequently been interpreted with greater generality than the actualities of the problem really permit. The difference between the greater and lesser equatorial axes is given as nearly 10,000 feet, on absolute lengths of a little under 42 million feet. The departure from a spheroidal figure is therefore only 2,500 feet at its maximum. Now, when we have regard to the actual conformation of the earth's surface—with mountains rising above the sea level to a height of fully ten times this departure from spheroidal figure—it becomes necessary to enquire what particular mean this represents. It is evident that the mean figure obtained with reference to mere external form would by no means correspond with the statical mean obtained by considering the earth at rest, but having regard to local density, and making such assumptions as to the mean distribution of density as would render the problem determinate. Again, a different result would be obtained if we used the ellipsoid which would represent the earth dynamically for any specified purpose. Taken in its proper sense, as a mean sea-level,—that is to say, as the figure which would be given by water in a pipe open to the sea at one end, and brought to every possible latitude and longitude at the other end,—it is of course a definite thing. But when we consider how very small is the difference from a much coarser mean, and how very large compared with it are the local irregularities of the earth's surface, regarded either geometrically or statically, it does not appear that much reliable use can be made of

it for general purposes; but it is unfortunately finding its way into the elementary books, with as much precision of statement, and as little restriction of generality, as the eccentricity.

The principal divisions of the general problem of interpolation, setting aside periodicity, are—

1. When the ordinates are numerous, and fairly close to one another.
2. When, for some reason or other, but few ordinates are available, and these usually separated by wide intervals.

In the former case the approximation is good, subject to the condition, in the general case, of not affecting a very extreme accuracy; and in the case where the law of variation is known to be continuous, the approximation may be pushed on without limit; but this is rather a question of computation from formulæ, than of representation of observations. The true test of accurate correspondence is, that all the forms of interpolation should give concordant results, in this sense, that it should be indifferent whether we interpolate directly to the required result, or whether we interpolate to some function of the observations, and then take the inverse functions. If these results differ materially, either result is only reliable within the limits of that difference. Care must, of course, be taken that the function selected shall not itself introduce an approach to discontinuity, as in the case of a reciprocal or logarithm in the neighbourhood of a zero of the variable function.

Where but few ordinates are available, any attempt at general inferences from them is futile, and not to be mended by elaboration of formula, unless we have some extraneous source of information as to the law which the observations must, with more or less approximation, follow. That is the case with new asteroids and comets. We know, from physical considerations, that their paths in space can differ but little from focal ellipses about the sun. Hence a number of observations, sufficient to determine the parameters of this elliptic motion, render the problem of interpolation determinate. That is not so, when there is no indication of law. In that case we must remember that all that our observations really tell us is the state of the function at the point of observation. This is obvious enough when generally stated, but, nevertheless, the caution is not needless, for when curves of stability for ships first came into ordinary use for ships of war—I am sorry to say they are not yet in common use for merchant ships—a suggestion was made that it might be useful to combine the ordinates by Stirling's higher rules, instead of the simpler ones. In the calculations made for the navy, however, the officials of the Constructor's department preferred to use the lower rules, and to select the points of discontinuity for the final ordinates of separate summations; and this was evidently better and sounder work.*

* See the "Transactions of the Institution of Naval Architects," Vol. XII. (for 1871), pp. 96—125, for the methods actually employed in the Admiralty.

In the direct application of Finite Differences, James Stirling rightly lays great stress on the selection of a proper form for interpolation, by means of transformation, if necessary. His words (translated from the Latin) are:—"For interpolation is not to be rashly undertaken; but before beginning the work we must inquire what is the simplest series, upon the interpolation of which that of the proposed series depends. And this preparation is usually absolutely necessary, in order to arrive at neat and definite conclusions."* His own illustration consists in forming a series whose n^{th} term is the product of the n^{th} terms of three other series, and then remarking that it is simpler to interpolate to these three series separately, and to take the product of the results of these separate interpolations, than to interpolate directly in the more complex series. This is evidently meant as an illustration of the wider truth,—that such transformation, as is necessary to make the interpolation as small a matter as possible, should always be adopted. A familiar case of this is the interpolation of the logarithms of numbers differing but little from unity, or of the logarithmic sines or tangents of small angles, in which a transformation is always necessary to exactness.

It is but an extension of this principle in another direction, that when any definitely known law governs the phenomenon, to which the observations relate, either as a whole, or in part, that law should enter into the formulæ of interpolation. Thus, for accurate work, the motion of a projectile subject to gravity is often better treated by interpolating its departure from a parabola, than by direct treatment of its ordinates. This advantage, however, disappears when the departure is excessive, as when the resistance of the medium is very great, or when the trajectory is followed to an extreme length. In these cases a formula must be used which takes account of resistance, and the subject of interpolation or comparison must be the departure from the path thus indicated. This caution is even more necessary to be observed where there is periodicity. When that exists, no interpolation which fails to take notice of it is worth anything if it covers more than a very small fraction of the period.

This consideration emphasizes the importance both of inferring a continuous law from such observations as we can command of any phenomenon suspected to be continuous, and also of testing that law when we have inferred its form. The actual inference of the form is a matter subject to no law. Sometimes it is an obvious deduction from physical considerations; at others it is simply that application of wide experience which is called wisdom; occasionally it seems to involve that deep and instinctive insight into the nature of things, which is called genius. The inference once made, its source is immaterial, except

* See Stirling, "*Methodus Differentialis*" (London, 1730, 4to), p. 88.

as a matter of history; but there are some points relating to its comparison with the observations which are worthy of our attention.

First in order, I may mention the case in which the phenomena are essentially discontinuous. I have already mentioned one instance, in the law of population of an island. Another is to be found in the formula which represents the number of primes below a limit x , namely,

$$x : (\log, x - a).$$

There are two things to be observed about this. In the first place, the irregularity in the occurrence of the prime numbers renders any exact formula necessarily untrue, when we go into minute detail. Secondly, the quantity a appears to be a function of x , and not a constant.*

There is a remarkable instance in which an experimental hypothesis was found to represent with very great exactness extensive series of statistics, which, to casual observation, entirely failed to suggest any obvious law. The late Benjamin Gompertz suggested that the law of human mortality might be very nearly expressed upon the hypothesis that the force of vitality diminishes by equal quantities in equal times, reckoned from birth. This leads to the following expression for the table of decrements of life:—

$$dy = -ab^x y dx,$$

or

$$y = p^{q^x},$$

where a and b , or p and q , are parameters to be determined from the observations, and y represents the number living at the end of x years. It is found that this form, with different values for the parameters, represents all the ordinary life tables over a very large portion of their extent. Nevertheless, it does not represent the best tables from end to end. If adjusted to the middle portion of the table, it fails to represent either infantile or senile life with exactness. It has been attempted to meet this difficulty by treating these parts of the table discontinuously. This is a very reasonable plan, so far as regards arithmetical accuracy, in any results we may deduce from it. Otherwise it is not very satisfactory, for it leaves in doubt whether this mode of representation is intended to cover want of continuity in the phenomenon, or the inexactness of the formula applied to it. In justice to Mr. Gompertz, it should be stated that he is not responsible for this addition to his theory.

Gompertz's formula, however, may be made to represent, by a proper choice of values for its parameter, the main portion of all the principal tables of the decrements of life, in some cases including infantile or senile life, and even both, with a degree of exactness running very close

* On this subject see Glaisher's "Introduction to the Factor Table for the Fourth Million."

upon the limit of continuity of the tables; for, with the exception of the Registrar-General's tables, which go to millions, most of the ordinary life tables only deal with from 10,000 to 100,000 individuals. In these cases, therefore, minute accuracy, beyond three or four figures, lands us in the discontinuity already mentioned, as arising from our falling back upon the individual unit. There is, therefore, no object in pushing the discussion of the formula too far, especially as all these inductions are open to another source of error. As I have already remarked, it is an interesting, and almost untouched problem, to adjust the degree of interpolation to the limit of continuity.

The inaccuracy of a formula may sometimes be masked, almost completely, by giving an average constant value to some really variable parameter. Thus the observations of the trajectory of a rifle bullet may be represented with very great accuracy on the supposition that the resistance throughout the flight varies simply as the cube of the velocity. As the observations extend from initial velocities considerably exceeding that of a sound wave in air, to a final velocity considerably below that of sound, it is obvious that the same physical law cannot hold, with an unchanged parameter, throughout the whole course; and the inference is, that, instead of having obtained the actual law, we have merely obtained an approximate law, sufficiently close to the real one not to be arithmetically distinguishable from it in the results which fall under our observation. This source of error is too often overlooked in the practical verification of assumed laws which are supposed to govern phenomena.

When we are free to choose form, instead of assigning values to the parameters of an arbitrarily chosen form, like that of parabolic interpolation, it is obvious that a suitably selected form may enable us to represent a series of observations with far greater exactness, and yet with a small number of parameters. In fact, the solution of certain differential equations by the variation of parameters, is little more than a particular application of this principle, with some generalisation of it. The advantage of such selected forms is, that they afford some degree of probability that they are approximations to the law governing the phenomena, as well as arithmetically approximate representations of the observations. But mere arithmetical fitting, obtained by the assumption of complex forms with numerous parameters, is not only no evidence that the true law has been approximately found, but affords no justification for the use of the formula beyond the extremes of the observations. The process is, in fact, of no more value than the application of parabolic interpolation to a series not ascertained to be free from periodicity, or from infinite values of differential coefficients, within the limits; and it is usually more troublesome.

The foregoing remarks are irrespective of any uncertainty as to the

individual observations. When a number of observations of a phenomenon, which can yield but a single numerical value, have to be compared, the ordinary theory of the errors of observation furnishes the most probable numerical amount of that value, or of any given function of that value; and this whether the observations be all equally good, or have definite numerical weights attached to each. A further refinement has been introduced by attaching weights, themselves derived from the departure of the individual observations from the first mean. This is a perfectly definite process, and the only remark which needs to be made upon it here is, that the most probable value of a given function of the result is not the same thing as the given function of the most probable value of the result.

When an unknown curve is only known by a number of points, each determined subject to some unknown but appreciable error, the problem of finding the curve is absolutely indeterminate, unless some assumption be made as to the nature of the curve. This will be best seen by taking an easy problem, in which the indeterminateness is removed by a simple supposition. Let us assume that a right line has been observed, and is to be plotted by means of a set of equidistant ordinates, but that, upon setting them off, the heads are not in a right line. It is then a perfectly definite problem, to find a right line such that the squares of the distances of the points from it shall be a minimum; and, in accordance with the fundamental principle of the ordinary theory, we shall find the same right line in whatever uniform direction we measure the distances. But the assumption that the line through the observed points is a right line, is exactly what we want to avoid in the general problem. On the other hand, when points of a curve are definitely given, we may make the curve determinate by assuming that its continuity is of the highest order possible. In its simplest form this is effected by assuming the curve to have a parabolic equation; but this is not essential, and we may settle it by circular curvature instead of by parabolic order. But, whatever law of facility we take for the nature of the curve connecting the points, that is evidently independent of the law by which the points are assumed, and there is no law connecting the two systems of probability. Any attempt to attain determinateness is therefore of necessity futile.

This indeterminateness is experienced in practice as well as indicated by theory. One of the commonest modes of "fairing" a curve through given points is by using a flexible batten, or spline, which is pinned down by lead weights to the points through which the curve is to be drawn, and the pen is then drawn along the batten. Now, in practice, it is found impossible to use similar battens for all curves. The batten has to be weakest where the curvature is the greatest, and it is a matter of taste and discrimination to select a batten with the proper

taper, and to use it discreetly, so as to get a reasonable and presentable result. The use of moulds or curved patterns is still more a matter of eye.

In the case of a curved surface, such as that of a ship, the problem is rendered somewhat more determinate by the consideration that all the sections, and all their projections, must be fair curves. The two sets of vertical sections, and the water sections, thus correct one another, and it is not an uncommon thing to complete the "fairing" by means of diagonal lines. Another mode, nearly equivalent, is to make a model, and to work it until it is not only quite smooth, but until, when it is held up in every possible light, the shadows fall evenly and fairly upon it. This is quite as severe a test as the drawing. Nevertheless, in either case, the adjustment is not a matter of rule, but of taste and judgment. Apart from the mechanical skill necessary to produce such work, there are many people whose perceptions are not sufficiently delicate to appreciate or test it.

While the problem is thus really and intrinsically indeterminate, all the solutions being strictly *secundum quid*, instead of being general, the difficulty is by no means beyond the reach of practical skill in the most useful cases. A comparatively small number of sections in two dimensions will enable two experienced draughtsmen to produce a couple of ships which shall differ very little in size or shape when they come to be built.

It may be worth while to repeat that the indeterminateness really turns upon the want of any arithmetical comparison between two independent systems of variation of error, or of any analytical means of combining them so as to give a single determinate result.

When the observation, regarded as a point in space, is not only uncertain in linear position, but is uncertain in two or three dimensions, so as to be limited by a circle or sphere of indefinite radius, there is further indeterminateness. This appears, at first sight, to be indeterminate in an independent manner, as being in another dimension. Whether any method can be devised which shall render the errors in different dimensions comparable, I do not know; but I have seen none as yet, and there will still remain the insuperable difficulty about the arithmetical comparison of the probable error of the line or surface through definite points, with the probable error in the position of the points.

Many of the foregoing remarks would seem to apply to the discussion of mean values, so often met with in meteorological works, and especially to means of means, both in respect of what the mean result really represents physically, and also as to whether the processes do not "plane down," or "smooth away," features essential to the theory, as well as local irregularities.

I fear that I have brought before you little that is new, and nothing that would fail to occur to any careful thinker, who gave real attention to the subject. Nevertheless, I hope that you will not think your time wasted, by my thus bringing before you, with some degree of system, a discussion of some of the limits of precision in the application of continuous formulæ to series of observations.

December 9th, 1880.

S. ROBERTS, Esq., F.R.S., President, in the Chair.

Mr. William Ralph Westropp Roberts, M.A., Ex-scholar, and Mr. Ralph Augustus Roberts, B.A., Scholar of Trinity College, Dublin, were elected Members of the Society.

The Auditor (Mr. J. Stirling, M.A.) stated that he had been to the Bank of England and found that the Treasurer's statement as to the monies therein deposited was true, and that he had examined the accounts and found them correct. A vote of thanks was passed to the Auditor on the proposal of the Chairman.

The following communications were made :

"Note sur la Dérivation des Déterminants:" Prof. Teixeira, Coimbra, Portugal.

"Solution of the Binomial Equation $x^p - 1 = 0$, Quinquisection:" Prof. Cayley, F.R.S.

"A General Theorem in Kinematics:" Prof. Minchin.

"On the Solution of the Inverse Logical Problem:" Mr. W. B. Grove, B.A.

"Motion of a Viscous Fluid:" Mr. T. Craig.

"On the Electrical Capacity of a Conductor bounded by two Spherical Surfaces cutting at any angle:" Mr. W. D. Niven, M.A.

The following presents were made to the Library :—

"Atti della Reale Accademia dei Lincei,—Transunti," Vol. v., fasc. 1°, 2°; Roma, 1881.

"Proceedings of Royal Society," Vol. xxxi., No. 207.

"Educational Times," December, 1880.

"Beiblätter zu den Annalen der Physik und Chemie," Band iv., Stück 12.

"Mémoires de la Société des Sciences Physiques et Naturelles de Bordeaux," 2° série, tome iv., 1^{er} cahier; Paris, 1880.

"Jahrbuch über die Fortschritte der Mathematik," zehnter Band, Jahrgang, 1878, Heft 3; Berlin, 1881.