# The Early History of the Ogive

**Chris Pritchard**

ABSTRACT

The ogive was invoked by Galton in 1874 as the visual embodiment of the method of ranks which he had designed to obviate the need to take actual measurements. Through the mediation of George Darwin, its analytical form was supplied by J W L Glaisher. By 1879, Donald McAlister had extended its domain to a distribution which was not normal, relabelled it the 'curve of distribution' and given its analytical form in Richard Pendlebury's adjustment to Glaisher's notation.

## Introduction

Francis Galton (1822-1911) developed the method of ranks to overcome two problems with different sources and characters but a single solution. The first was the need to avoid taking a direct measurement where to do so would cause offence or social discomfort. In his days as a traveller-explorer in South West Africa he had come across precisely this situation. The second was his interest in discovering the magnitude of a person's intelligence where no metric had ever been established. The key to both problems was to use an ordinal or rank as a surrogate when whatever was being 'measured' had been sorted by size. Ranks provided relative measurement in lieu of absolute measurement, as Galton explained to an audience at the Royal Institution at a Friday evening discourse on 27 February 1874. The record of his talk on the making of the English scientist (more correctly, the British scientist) was given in complementary articles in the next issue of *Nature*. These articles took the form of a letter from Galton to the editor, Joseph Norman Lockyer (1836-1920), 'On a Proposed Statistical Scale' and an anonymous account of the lecture, entitled 'Men of Science, Their Nature and Their Nurture', written in the third person.[1] The crux of the approach was to bind the ranks to the theoretical law of error (what we now term the normal distribution) on the assumption that many phenomena are modelled by such a curve. But the audience was largely ignorant of Quetelet's use of the curve in anthropometry and Galton's use of it in *Hereditary Genius* to describe the distribution of abilities.[2] Galton justified his use of the law of error using a quincunx of his own design but had not yet thought to graph the cumulative function.

## A primary influence: Quetelet's use of cumulatives

In the years leading up to the first use of the ogive, Galton had become thoroughly familiar with the statistical writings of Adolphe Quetelet (1796-1874). In particular, he had studied the Belgian's *Lettres* (1846) in which 'curves' had been fitted to data on the chest girths of Scots militiamen and the heights of French conscripts.[3] Galton had found here the use of both cumulatives and ranks, and so it proved a primary reference for Galton as he developed the concept of the ogive. [4]

Stephen Stigler has given a most detailed account of Queletet's curve-fitting process in his *History of Statistics*.[5] It appears that for pedagogical reasons Quetelet had chosen to fit a discrete binomial (albeit of very large $n$) rather than a continuous normal, though the consonance between the resulting *ligne brisée* and the smooth, idealised *courbe de possibilité* was near perfect. In the construction of the table of values, Quetelet had moved from observed frequencies to relative frequencies, then to cumulative relative frequencies and finally to ranks. After realigning the ranks so that they had a linear scale, he had reversed the whole procedure to arrive at the fitted frequencies.[6] But though cumulatives and ranks had both featured, Quetelet had not drawn a cumulative frequency curve or polygon (to use more modern terms).

## Galton's early use of the ogive

1. The measurement-avoidance ogive (1874)

The reception given to Galton's Royal Institution talk was mixed, with some of the harshest criticism coming from John Morley (1838-1923), then the editor of the *Spectator*, later as Chief Secretary for Ireland, a leading politician in Gladstone's third and fourth administrations. Galton's approach, Morley concluded:

> will never take hold on the world, because it is wholly and intrinsically inapplicable to the purposes for which he recommends it. But if ever it did attain an unfortunate notoriety, we suspect it would be much more likely to be called the scale of Sham Science, than by the name by which its inventor has proposed to christen it.[7]

Galton accepted that he needed to bring his ideas before a wider audience and over the coming months he mixed simple reiteration with exemplification, culminating in the paper on which later historians concentrate almost exclusively, 'Statistics by Intercomparison'. An early opportunity arose through Galton's committee work for the British Association for the Advancement of Science. In 1874, he joined other members of the association's General Committee to produce advice to travellers and explorers.[8] Despite the fact that Karl Pearson (1857-1936) drew attention to this publication in his biography of Galton, later historians do not discuss it.

Galton prefaced the main section of his short article by restating the error theorists' justifications for the 'law of deviation', including an appeal to symmetry, the homogeneity of populations and the plethora of tiny, independent influences.[9] Then he took his first, tentative steps towards the use of ranks, taking the example of there being a thousand of them and drawing attention to the 500$^{th}$ rank, or *c*, which he referred to, simply, as the average. The only significance of the particular label chosen is that C was used to indicate the position of the middle rank, the extreme ranks having already been labelled A and B. The accompanying graph shows a double vertical line at the 500$^{th}$ rank, as if Galton were acknowledging that he was using an even number of data.

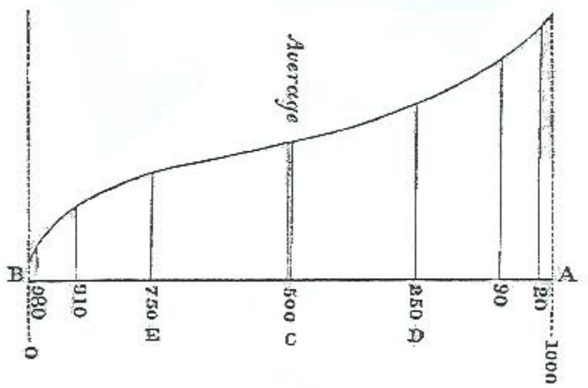



Figure 11.1: The original ogive, depicting the ordered heights of a thousand men

Galton also noted the 250$^{th}$ and 750$^{th}$ ranks, labelled D and E, to which he gave no special name. Neither did he give a name to *r*, the difference in the height at C and D. He did note that if the ranks represent men whose heights are the trait under consideration, then the heights of the 90$^{th}$ tallest and the 910$^{th}$ tallest men are respectively $c+2r$ and $c-2r$; likewise, the heights of the 20$^{th}$ tallest and 980$^{th}$ tallest men are respectively $c+3r$ and $c-3r$. Pearson observed that

‘Galton was proceeding gradually, and the dose was a very small and simple one’.[10] It was the ‘weakness of his brethren’ that demanded they be fed only ‘homeopathic doses’.[11]

It is interesting that Galton did not proceed from rank to measurement but from measurement to rank. This may have been didactic or it may have indicated his reluctance to throw off the mantle of error theory. The former is the more likely, though Galton did refer the novice to Layton’s translation of Quetelet’s *Lettres* for further explanation.

It has been suggested by Eliot Slater that the ogive held a special appeal for Galton because his concern was with the extremes of the normal distribution. Graphically, the extremes of the normal curve (*i.e.*, the curve of the density function) have little effect on the eye of the beholder, whilst the area of mediocrity bulges impressively. Slater observed that it is exactly the opposite in the ogive, which ‘runs almost level over the middle part of its course, but dips or rises with increasing steepness towards either extreme’.[12] Here mediocrity is aligned with a small gradient, the visual drama being in the tails. This was not without consequence, for as Slater argued:

> The man at the extreme right would exceed his neighbour to the left by a larger margin that the man would exceed his other neighbour. This led Galton to the misleading idea that there was more variation at the extremes of the distribution than about its middle. The correct view is that variability is a quality of the group as a whole, but that its effects will be more openly manifested than near the mode.[13]

## 2. The ogive as empirical and theoretical model (1875)

As already mentioned, Galton brought together all his ideas on indirect measurement in the autumn of 1874 in the paper ‘Statistics by Intercomparison’.[14] Published in the *Philosophical Magazine* the following January, it presented ‘a more complete explanation and a considerable development of previous views’.[15] Galton explained once more how it is possible to ‘dispense with standards of reference … being able to create and afterwards indirectly to define them’.[16] It is possible ‘to replace the ordinary process of obtaining statistics by another, much simpler in conception, more convenient in other cases, and of much wider applicability’.[17] The key to the new method was the ability of individuals to ascertain which of two objects possesses to the larger extent the quality under investigation. By making a series of such binary judgements the objects can be put in series. Galton did not allude to Gustav Fechner (1801-1887) but the nature of the exercise is not unlike the discrimination tasks of the psychophysicists.

Two populations can be compared by examining their means and ‘probable errors’, as determined from the ‘first and third quarter points’.[18] Galton expressed a preference

> to reckon the divergencies in excess separately from those in deficiency. They cannot be the same unless the series is symmetrical, which experience shows me to be very rarely the case.

This remark stands in stark contrast to the view often expressed of the universality of the law of frequency of error, and is contradicted forcefully later in the same paper. It certainly recognizes the potential usefulness of such a representation even where no symmetry exists. The magnitudes of the quality under consideration can be arranged on an evenly spaced base and a ‘curve of double curvature’ drawn through their tops.[19] Galton noted that ‘such a curve is called, in the phraseology of architects, an “ogive”’.[20] He labelled the ogive in his diagram with the onomatopoeic OG, marked the ordinates of the quarter points *p, m* and *q* and noted that fine discrimination is required only when sorting objects near to these three stations.[21]

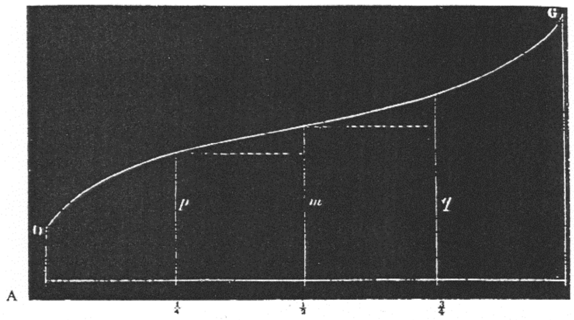



Figure 11.2: Galton's ogive with the quarter points marked

Galton used the ogive as both an empirical model and a theoretical model. When using it as an empirical model he made no assumption that the data are distributed according to the law of frequency of error. The data determine the ogive's parameters and the term 'ogive' is adopted regardless of symmetry. So the objects at the quarter points on the horizontal axis of relative frequency can be identified and the values of $m$, $p$ and $q$ can be estimated from the scale of magnitude on the vertical axis of the empirical ogive. These stations constitute 'permanent standards of reference' or benchmarks against which a comparable set of data may be judged, either now or in the future.[22]

As to the theoretical model, it is underpinned by an assumption that the data do, indeed, conform to the law of frequency of error. The ogive is fixed by the 'mean' and one other quarter point. With a regular arithmetical graduation on the horizontal axis, the scale on the vertical axis is automatically graduated in units of $q - m$, the probable error.[23] This allows a reversal of the usual argument, from a simple proportion or relative frequency to a measurement expressed in terms of the probable error, rather than from a measurement to a relative frequency.

Galton believed the assumptions on which the law of frequency of error had traditionally rested to be unduly stringent. In particular, the number of causes or influences affecting a phenomenon is often not very large and yet the theoretical model is postulated on its being infinite. Further, the law of frequency of error is an idealization of an underlying binomial law in which the number of influences that determine a character is simply the degree of the binomial expansion. Galton believed that number to be relatively small.

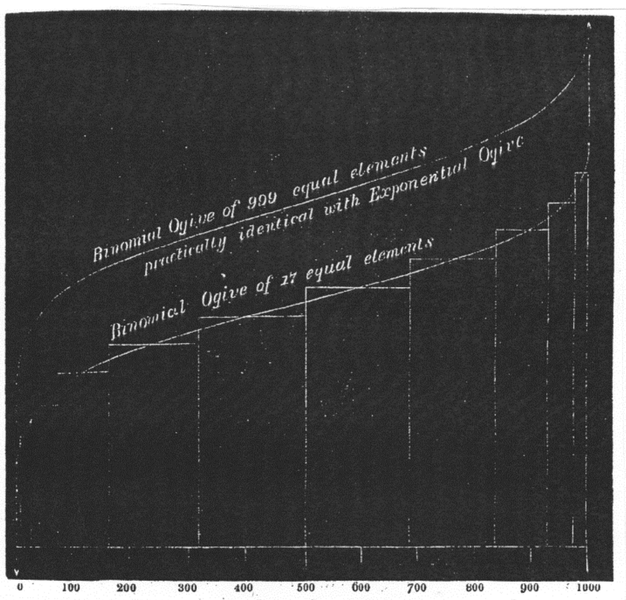



Figure 11.3: Galton's comparison of binomial ogives of 17 and 999 elements

He set about demonstrating it by comparing the 'binomial ogive of 17 equal elements' to the ogive of Quetelet's binomial of degree 999 (known to be almost identical with the law of frequency of error). He superimposed the latter onto the former by bringing together the means and stretching the 'Quetelet ogive' until the quarter points of the two curves also coincided.[24] Galton commented that the 'closeness of the resemblance is striking', and noted that it is difficult to separate the two visually when the power is as high as 30.[25]

The preliminaries complete, Galton offered a plausible argument that the law of frequency of error is self-referencing. In other words, a combination of binomial laws produces a law which differs little from another binomial or, equivalently, for characters with a moderate to large number of influences, a combination of laws of frequency of error differs little from another law of frequency of error.

## 3. The comparison ogive (1875-76)

In much the same period, Galton had also begun research into the growth of the human form, now referred to as 'auxology'.[26] He had approached the Anthropological Institute in 1874 with a proposal to gather anthropometric data on boys from a range of 'Public Schools, middle class schools and others, down to those of pauper children'.[27] The first two schools to respond were Marlborough School and Liverpool College, but 'the statistics came out very differently, so that it would have been impossible to combine them' and only the Marlborough data and their analysis were published initially.[28] Galton must have suspected that the disparity between the two sets of data reflected the fact that one was a rural school, the other located in a city. Certainly, by the time he had received returns from nine schools (all public schools) he had decided to undertake a comparative study of physical development at schools in rural and urban locations. Galton made public this larger set of data at a meeting of the Anthropological Institute in the spring of 1875 and his analysis was carried by its journal the following year.[29]

The statistics of only eight of the schools were used in the main part of the study, four town schools, four country schools.[30] In total a little fewer than 4000 boys were measured.[31] One criterion Galton insisted upon was that the number of boys in the town and country categories be roughly the same. With many boys leaving the town schools before their sixteenth birthdays the best balance existed for 14-year olds. This reduced the number of boys' measurements to a rather 'scanty' 296 country boys and 509 town boys.[32]

It was only fourteen months since Galton had unveiled his new method of ranking data and he took the opportunity to illustrate it once more.[33] For four data sets, heights and weights of country boys and heights and weights of town boys, the observed frequencies were reduced to 'per centages' and then summed and tabulated.[34] It was plain to Galton that the 'curves of height (of course, not those of weight) conform fairly to the Law of Error'.[35] The weight statistics were put aside.

The summed percentages for height were graphed to produce two *ogives* or 'curves of contrary flexure', one each for the country and town schools.[36] The distribution of heights was summarised by the lengths of five ordinates or vertical 'rods': the 'middle ordinate', together with ordinates one and two probable errors either side of that average, *i.e.*, at the 8$^{th}$, 25$^{th}$, 75$^{th}$ and 92$^{nd}$ divisions. Galton commented that to measure any other ordinates would be a 'misdirection of labour', such is the form of the law of errors. He directed readers to the methods of Quetelet's *Anthropométrie*, published in 1870, for comparison with his own and quoted the average height of Belgian 14-year olds as given by Quetelet .[37]

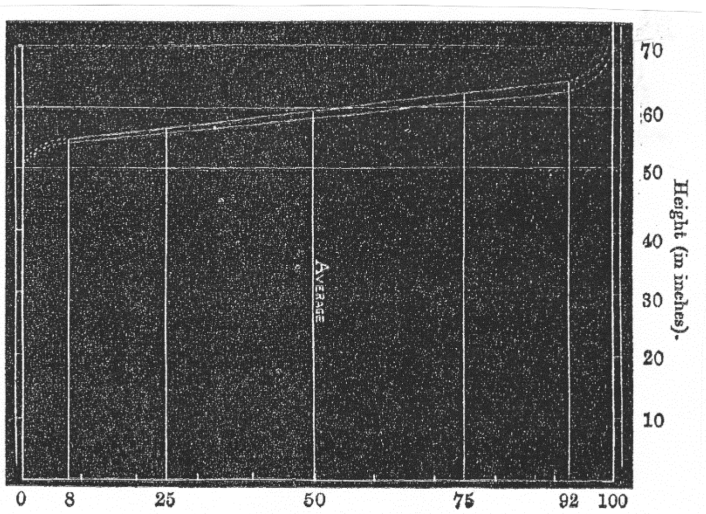



Figure 11.4: Comparison of the distribution of heights of 14-year old boys
[upper ogive (urban schools), lower ogive (rural schools)]

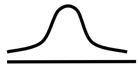
From his study Galton noted that the boys at the country schools were on average 1¼ inches taller and 7 lbs heavier. He concluded that the ‘difference in height is due, in about equal degrees, to retardation and to total suppression of growth’, disregarding the possibility that the social classes of English society were represented in different proportions in the two groups. When Karl Pearson reviewed Galton’s papers, he noted that the boys of Clifton, Eton, Haileybury, Marlborough and Wellington were of the professional and administrative classes. The boys of Christ’s Hospital, City of London School, King Edward’s Birmingham and Liverpool College belonged to a different social grouping, being the sons of shopkeepers and clerks. Pearson therefore argued that Galton, without fully recognising it, confounded demographic factors (town versus country) and social factors.[38]

### Glaisher on the analytical form of the error function

Over the winter of 1874-75, George Darwin (1845-1912) cast a mathematician’s eye over Galton’s ogive, but in this he was joined by his colleague and friend at Trinity College Cambridge, James Whitbread Lee Glaisher (1848-1928). It was Glaisher who explained how the ogive could be expressed in the form of a mathematical function. In fact, he provided two such forms.

Galton and Glaisher were first introduced in the late autumn of 1874. The encounter was brief, though long enough to leave Galton with a most favourable impression. If Galton and Glaisher had time for more than the usual pleasantries, then their common interest in the error function may well have been discussed, however dissimilar the sources of that interest. Galton and Glaisher’s father, James Glaisher (1809-1903) shared an interest in the error function as used in instrument calibration, but the son was interested in it purely as an analytical function.

Meanwhile, Galton wrote to Darwin on Christmas Day, 1874:

> As regards that “ogive” of which we were talking, I was stupid & explained myself ill,
>
> & boggled. In the ordinary  $x$ is the magnitude & $y$ the frequency.
>
> In my plan, $y$ is the magnitude & $x$ the sum of the frequencies, the frequencies being taken
>
> from the $e^{-\frac{x^2}{c^2}}$ tables & the sum of the frequencies from the tables of the integration

of it. viz Tables I & II respectively of the usual publications (? II & III in the Encycl[opædia] Metropolitana).[39]

It appears that Darwin had not initially appreciated that Galton's ogive was an inverse function.[40] The independent variable is *y* and is shown graphically on the horizontal axis. This variable is under control, its values taken at fixed intervals. The cumulative frequencies constitute the dependent variable, and on the assumption that the character under consideration is distributed according to the error function, its values can be looked up in tables. If Galton had asked him to find a formula for the ogive, as seems likely from Darwin's reply, such an appreciation would have been essential.

As we have seen, Galton's paper, 'Statistics by Intercomparison', was published at the turn of 1875, some ten months after his lecture at the Royal Institution on the subject of ranking methods.[41] Darwin wrote to Galton on 4 January 1875 immediately upon reading it and about a week after receiving Galton's letter.[42]

> You talk of filling in y[ou]r 'ogives' with free hand; I suppose you know those 'French curves' which enable you to draw in curves thro' any number of points very neatly.
>
> One can obtain the Equation to y[ou]r binomial curves in terms of what we call Γ functions; i.e. $\Gamma(n+1) = 1.2\ldots..n$ when *n* is integral, but it can't be expressed algebraically when *n* is fractional: The equation to this Expon[entia]l Ogive is
>
> $$x - \frac{n}{2} = a\int_m^y e^{-\left(\frac{y-m}{c}\right)^2} dy$$
>
> where *m*, *n*, *x*, *y* have y[ou]r meaning, *c* is the modulus & (*a*) some const[ant]. At least I think this is so.
>
> I do not see any def[initio]n of 'grade' in so many words & it took me a few minutes before I saw what you meant.
>
> It w[oul]d be well not to speak of the  as a curve of 'double' curv[atur]e, as that term is already occupied to mean a curve with torsion and flexure e.g. a helix but a curve of contrary curv[atur]e or with a p[oin]t of contrary reflexure. You w[oul]d of course be aware of the distinction.

In response, Galton wrote, 'thank you much for the Equation to the ogive' and acknowledged that '*curve of double curvature* was a sad slip for *curve of contrary flexure*'.[43]

Darwin was an accomplished young mathematician, but his familiarity with the mathematics of the error function and related functions could not rival that of Glaisher, one of the leading authorities on the subject. In contrast to the support Darwin provided to his own father and his cousin in his early career, which may be interpreted as largely familial, Glaisher's long career was characterised by selfless and magnanimous support for his scientific colleagues. If Galton required help, and whether approached directly or by Darwin on Galton's behalf, that help would have been forthcoming.

James Whitbread Lee Glaisher was known in the family as Lee and in the scientific community by his initial letters.[44] As a youth, he acted as a guinea-pig on two of his father's balloon flights, the first to a height of over 14,000 feet. He was monitored for changes in colour, pulse and breathing rate, with increasing altitude.[45] Glaisher was educated at Trinity College, Cambridge and graduated as Second Wrangler in 1871. Such success in the Mathematical Tripos assured him of a college fellowship and thus began his lifelong affiliation with Trinity College. He had been lecturing at Cambridge for over two years when Darwin returned there in October 1873.

Half a century later, when it fell to the second Sadleirian Professor at Oxford, Andrew Forsyth (1858-1942), to pay tribute to Glaisher, he portrayed him as no great pioneer but, rather a spur to others.[46] Forsyth focused on Glaisher's lengthy service on the council of the London Mathematical Society and as the editor of two journals for mathematicians over a fifty year period, the *Quarterly Journal of Pure and Applied Mathematics* and the *Messenger of Mathematics*. He believed that Glaisher's greatest legacy as a mathematician was related to the calculation of mathematical tables. Glaisher was certainly the leading authority of his day on tables of the error function and their history, and an accomplished error theorist in his own right.[47] It was his special knowledge of error theory in his early years at Trinity College, Cambridge which helped Galton to better understand the mathematics of the ogive. Darwin may have acted as prompt and intermediary but it was from Glaisher that Galton learnt of the ogive's functional form.

Glaisher was still a student when his first paper was communicated to the Royal Society by Arthur Cayley.[48] In it he argued that there are numerous functions that cannot be integrated analytically and neither is it convenient to resort to numerical methods. However, if such functions can be expressed in terms of 'fundamental' functions already evaluated numerically, then they too can be evaluated. He noted that 'one function, … the integral $\int_x^\infty e^{-x^2}dx$, well known for its use in physics, is so obviously suitable for the purpose, that, with the exception of receiving a name and a fixed notation, it may be said to have already become primary'.[49]

In the first part of the paper, Glaisher called the function the 'Error-function', labelled it Erf $x$, and then expressed some two dozen results in terms of it. In the second part, he noted that, expressed in the notation of Legendre, Erf $x$ is the gamma function $\frac{1}{2}\Gamma\left(\frac{1}{2}, e^{-x^2}\right)$.[50] It was in a similar gamma function form that Darwin expressed the ogive in his letter of 4 January 1875 to Galton, adding the codicil that he only thought he had the correct form. It is unlikely that he was able to consult Glaisher directly in those days around the New Year, when Trinity College was not in session. But it is clear that he was aware of Glaisher's paper and capable of adjusting the gamma function so that the abscissae are expressed in terms of the ordinates, *i.e.* $x$ in terms of $y$.

Also in the second part of the paper, Glaisher defined the 'Error-function-complement' Erfc $x$, to be Erfc $x = \int_0^x e^{-u^2}du$, where Erf $x$ + Erfc $x = \pm\frac{1}{2}\sqrt{\pi}$, depending on the sign of $x$.[51] A further thirty identities were then deduced. And the pure mathematics of the two related functions having been explored, Glaisher turned to the applications of Erf $x$ to physics and to the various mathematical tables of the function which had been produced. He explored Kramp's application of the error function to refraction, dissecting his methods of calculation as he went.[52] He also expressed, in terms of Erf $x$, Fourier's function for the conduction of heat in a metal bar and noted William Thomson's use of it to estimate the age of the earth.

Kramp had previously provided tables of Erf $x$ for 0.00 (0.01) 1.24, to between 7 and 11 decimal places, together with tables of $\log_{10}\text{Erf } x$ and $\log_{10}\left(e^{x^2}\text{Erf } x\right)$, each for 0.00 (0.01) 3.00, to 7 decimal places. Subsequently, they had been corrected and supplemented by Bessel, Legendre, Encke and De Morgan.[53] Glaisher extended the range of the tables of Erf $x$, providing values of the function for 3.00 (0.01) 3.50, to 11 decimal places; for 3.50 (0.01) 4.00, to 13 decimal places; and for 4.00 (0.01) 4.50, to 14 decimal places.[54] So that accurate values of Erfc $x$ could easily be calculated from them, he also gave the value of $\frac{1}{2}\sqrt{\pi}$ to 14 decimal places.

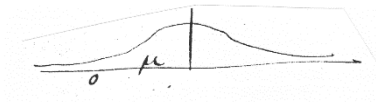
Immediately following the publication of the first part of Glaisher's paper, Richard Pendlebury (1847-1902) of St John's College, Cambridge, also perceived the need to have a notation for $\int_0^x e^{-x^2}dx$ and proposed erf $x$.[55] In a paper of 1875, Glaisher chose instead to interchange his original definitions of the error function and its complement, using Erf $x$ = $\int_0^x e^{-x^2}dx$ and Erfc $x$ = $\int_x^\infty e^{-x^2}dx$.[56]

Following a request for information either made directly by Galton or, more likely, indirectly via Darwin, Glaisher wrote to Galton on the subject of ogives.[57] As Galton explained to Darwin

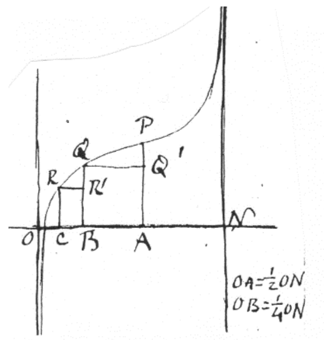
> I got a letter from Glaisher a short time back about my "exponential ogive" whereof he much approves, name and all, and he gives me a compact expression for it, in terms of his "Error function". I enclose a copy of part of what he says.[58]

Galton had transcribed Glaisher's explanation and attached it to his letter.[59] It reads:

> In the ordinary theory, $\mu$ being the distance of the greatest ordinate from the origin, the chance of an error lying between $x$ and $x + dx$ is $\frac{h}{\sqrt{\pi}}e^{-h^2(x-\mu)^2}dx$ which is of course the curve usually drawn




> while the curve that you draw in place of this is $x = \frac{h}{\sqrt{\pi}}\int_{-\infty}^{y} e^{-h^2(y-\mu)^2}dy$ which may also be
>
> written $x = \frac{1}{\sqrt{\pi}}\left\{\frac{\sqrt{\pi}}{2} - \text{Erf } h(\mu - y)\right\}$
>
> or $x = \frac{1}{\sqrt{\pi}}\text{Erfc } h(\mu - y)$ and is like this




> [Erf $x$ is written for $\int_0^x e^{-x^2}dx$, Erfc $x$ = $\int_x^\infty e^{-x^2}dx$,
>
> so that Erf $x$ + Erfc $x$ = $\frac{\sqrt{\pi}}{2}$.]
>
> If we put $x$ = ½ we have Erf $h(\mu\text{-}y) = 0$ so that $y = \mu$ and AP is as you say the 'average'. Put $x$ = ¼, we have Erf $h(\mu\text{-}y) = \frac{\sqrt{\pi}}{4}$ and $h(\mu\text{-}y) = \zeta$ so that $y = \mu - \frac{\zeta}{h}$ ($\zeta$ = .4769) and, as you say, PQ' = the probable error.

Galton was clearly unaware of Glaisher's papers in which the two complementary functions are used extensively. They include no graphs but Glaisher had made the connection. The definitions of Erf $x$ and Erfc $x$ are here juxtaposed, compared with the original definitions. Glaisher later explained the reasons for the switch:[60]

> This notation … was not well chosen, for it is the latter integral that occurs in the theory of errors, viz. The chance of an error not exceeding $x$ is $\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. Soon afterwards, therefore, I interchanged the functional symbols, putting Erf $x$ = $\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ and Erfc $x$ = $\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$ and this notation I used in subsequent papers.

The letter also revealed to George Darwin that Galton and Glaisher had attempted to tabulate the ogive. Galton wrote that:

> it occurred to me that it would be uncommonly convenient to calculate an 'exponential ogive' table, which I did, and since receiving Glaisher's letter I sent it to him to see if he c[oul]d get it properly recalculated for me directly from his formula. You see, by knowing *any* two ordinates, you know the whole curve & can at once get the value of any other ordinate in it. I need not bother with particulars about the table, further than that it gives ordinates from 1 to 50 in an ogive of 100 places, from 1 to 50 in an ogive of 1000 places — ditto 10,000, 100,000 & a million. So that all goes into a page.[61]

In fact, Galton intimated four months later (in May 1876), the Galton-Glaisher tables would be made available in a joint publication:

> Glaisher & I are going to send in a joint paper about that ogive "Exponic Ogive" we are going to call it. He has calculated tables de novo & we have got them into a most handy form for reference. I really think many future statisticians will be grateful for them. For by them, given *any two* magnitudes in a "statistical series" of *any* (given) number of things, whose places in that series are known, you find by the simplest rule-of-three arithmetic, both the mean & the modulus.[62]

Yet, in a paper of 1908, Glaisher gave a comprehensive listing of tables of the error function and related integrals but made no mention of any tables produced by him in the late 1870s in connection with Galton.[63] The proposed tables were never published, possibly because of their limited scope.

### McAlister's 'curve of distribution'

In 1879, Galton approached a young mathematician, Donald McAlister (1854-1934), to find the function having the same relationship to the geometric mean as the law of frequency of errors has to the arithmetic mean.[64] In his autobiography, he wrote:

> I have received much help at various times from Mathematical friends. On one occasion, being impressed with the probability (owing to Weber's and Fechner's Laws) that the true mean value of the qualities with which I dealt would be the Geometric and not the Arithmetic Mean, I asked Mr. Donald Macalister … to work out the results. [65]

Galton had attacked the problem himself in the summer of 1877 and had reached a result which 'seems to come out very prettily & simply'. His intention had been to send it to *Nature*, but wishing 'to be assured that it is correct' had asked George Darwin's opinion.[66] The advice he received was such that Galton told Darwin three weeks later that 'I rewrote the thing I sent you & have simply docketted it & laid it by for some future use'.[67] This suggests that Darwin had cast aspersions on the original analysis and that Galton had been unable to proceed along the lines suggested. It is likely that Darwin had been unwilling to spend time on it.[68]

McAlister was a 'vigorous mathematician', according to Galton, and this was borne out by his record as an undergraduate.[69] He read mathematics at St. John's College, Cambridge from 1873, arriving to take up his studies just as George Darwin returned to Trinity College. Glaisher was already a tutor at Trinity College, the immediate geographic neighbour of St John's. McAlister graduated as Senior Wrangler and First Smith's Prizeman in January 1877. He later pursued a brilliant career as a doctor and administrator, presiding over the General Medical Council for 27 years (1904-1931).[70]

Prior to approaching McAlister with his mathematical challenge, Galton had met him socially, almost certainly through his wife's brother, the Headmaster of Harrow School, Henry Montagu Butler (1833-1918). Once McAlister had completed his analysis, Galton sent two papers to George Gabriel Stokes (1819-1903), as Secretary of the Royal Society, with the intention that they should be published together. The first was a short introductory paper 'The Geometric

Mean, in Vital and Social Statistics', in which Galton outlined the remit he had given McAlister; the second was McAlister's paper, 'The Law of the Geometric Mean'.[71] A short correspondence ensued between Galton and Stokes following the papers' arrival. From one such letter from Galton it can be seen that Stokes was keen to learn of suitable domains of application for this mathematics:

> About MacAlister's paper; it might be well to look at the marked passages in the enclosed letters from him, sent to me a few days back …
>
> The principal people who have used the law of error for vital statistics, since Quételet, are the compilers of the War department Statistics of the N. American Forces after the war between the N. and S. States. And again, curiously enough, Fechner himself in his *Psychophysik* (I, 108) introduces a long mathematical investigation by his mathematical colleague (I have lent the book and forget his name) wherein a series of law of error tables, 'Methoden der richtigen und falshen Fälle,' are formed to help him in his own investigations. In short, he ignores his own law! He uses tables on the Arithmetic Mean principle to discuss results of observations on phenomena that have the Geometric Mean condition. So the question treated in the paper is really one of importance to statisticians.[72]

The two papers were subsequently read by Galton at the Royal Society on 20 November 1879. A second version of McAlister's paper was published in 1881 in the *Quarterly Journal of Pure and Applied Mathematics* in the style of its editor, Glaisher.[73] This full, mathematical version — in which there is a reference to the former version as a mere 'abstract' — has been largely ignored by historians.[74] The first version relies on the geometric or visual, and is essentially in the style of Galton, the second on the analytical and very reminiscent of Glaisher's papers.

The context for McAlister's papers is to be found in Galton's introductory paper. Here he rejected the notion that the errors deemed equally likely either side of the true value by error theorists can be adopted in vital and social statistics.[75] From Fechner's experimental studies of the lower mental faculties it was plain that equal intervals were defined by taking ratios rather than differences, and hence the true central value was the geometric mean rather than the arithmetic mean of the observations.[76] He not only alluded to the psychophysical studies of the senses but also sought to extend the range of applications to the growth of money, business and population.[77]

Both versions of McAlister's paper contain two fundamental results. Firstly, the analogue of the law of frequency of error is the 'law of facility', given by the curve $y = \frac{h}{x\sqrt{\pi}} e^{-h^2(\log x)^2}$, where $h$ is a measure of dispersion termed the 'modulus of precision' or 'weight'.[78] Secondly, if $x$ and $y$ obey the law of facility, with moduli of precision $h_1$ and $h_2$ respectively, then $z = xy$ obeys the law of facility, with modulus of precision $h$, where $\frac{1}{h^2} = \frac{1}{h_1^{\,2}} + \frac{1}{h_2^{\,2}}$.

Throughout the two papers, the discussion is related closely to Galton's 'Statistics by Intercomparison', whilst extending not just the mathematics but the nomenclature. And amongst that terminology is a new name for the ogive, the 'curve of distribution':

> Another method of exhibiting the law, suggested by Mr. Galton's Method of Intercomparison, is the following. Let the series of measures be represented by a series of ordinates: arrange these side by side at equal small distances and in order of magnitude. Their extremities will then lie on a curve of contrary flexure, which Mr. Galton calls an Ogive; we may speak of it as the '*curve of distribution*'.[79]

Galton did not use the term 'curve of distribution', either in a paper on mental imagery of 1880 or in his *Inquiries into Human Faculty* of 1883, both of which contained discussion of such curves.[80] It is likely, therefore, that McAlister and not Galton, coined the term 'curve of distribution' and Galton retained a preference for his own 'ogive'.

It is in the first version of the paper that graphs of the law of facility are shown, including the asymmetrical (lognormal) distribution curve:

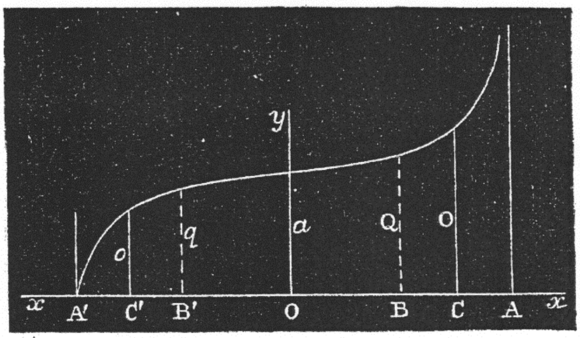



Figure 11.5: McAlister's 'curve of distribution'

McAlister gave its analytical form as $h\kappa x = \int_0^{h\log y} e^{-t^2} dt = \text{erf}.(h\log y)$.[81] This is not Glaisher's notation — he always used 'Erf', rather than 'erf' — but Pendlebury's adjustment to it.[82] Pendlebury had been a Fellow at St. John's College since graduating there as Senior Wrangler in 1870, and was one of McAlister's teachers.[83] For Glaisher's own journal, *The Quarterly Journal of Pure and Applied Mathematics*, McAlister diplomatically switched to the editor's notation and invoked his name in the phrase, 'Mr. Glaisher writes it $h\kappa x = Erf\,(h\log y)$'.[84]

## Conclusion

The originality of the ogive was Galton's alone. There were some small ways in which Darwin's influence was felt directly on Galton — offering advice on terminology, for example — and he acted as a sounding board throughout, but it was through his contacts that Darwin's influence was brought to bear. From Glaisher, who was steeped in the mathematics of the ogive, George Darwin learnt of the curve's analytical form and of the 'erf' designation and this he intimated to Galton. The designation was promoted through Glaisher's journals and by way of Pendlebury, a contributor to one of those journals, it also became known to McAlister. He adopted it for the little-discussed, mathematical version of his paper on the law of facility, an analogue of the law of frequency in which the geometric mean is the measure of central tendency.

---

1 Francis Galton, 'On a Proposed Statistical Scale' *Nature* 9 (5 March 1874), 342-343; idem, 'Men of Science, Their Nature and Their Nurture', *Nature* 9 (5 March 1874), 334-345.

2 Francis Galton, *Hereditary Genius* (London: Macmillan, 1869). It might be argued that in a section on, 'The comparative worth of different races', (pp. 325-337), Galton took a step towards the notion of ranks when discussing the assigning of 'grades' to different races for the purpose of comparison, just as today we might assign grades for examinations. But these grades were classes or ranges rather than individual values, whether cardinal or ordinal.

3 Adolphe Quetelet, *Lettres à S. A. R. le Duc Régnant de Saxe-Coburg et Gotha, sur la théorie des probabilités, appliquée aux sciences morales et politiques* (Brussels: Hayez, 1846). See letters 19-21.

4 Stephen M. Stigler, private communication, received 18 July 2008.

5 Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900* (Cambridge, MA: Belknap Press of Harvard University Press, 1986), pp. 206-215.

6 Ibid., 212.

7 [John Morley], 'Mr. Galton on Nuts and Men', *The Spectator*, 16 May 1874, 623-624, 624.

[8] Francis Galton, 'Notes and Queries on Anthropology for the Use of Travellers and Residents in Uncivilized Lands. No.C – Statistics', *Handbook of the British Association for the Advancement of Science* (London: Edward Stanford, 1874), 143-146. Galton also wrote sections on 'Population', 'Count wherever you can', 'Communications', 'Causes that limit Population' and 'Astronomy' and other contributors included Charles Darwin and John Lubbock.

[9] Ibid., 144-145; Pearson, *Life* II, 334-335.

[10] Pearson, *Life* II, 334.

[11] Ibid., 335.

[12] Eliot Slater, 'Galton's Heritage', *Eugenics Review* 52 (1960), 91-103, 100.

[13] Ibid.

[14] Francis Galton, 'Statistics by Intercomparison', *Philosophical Magazine* [4] 49 (1875), 33-46.

[15] Ibid., 33.

[16] Ibid., 34.

[17] Ibid.

[18] Ibid., 34-35.

[19] Ibid., 35. George Darwin immediately pointed out the problem in using this terminology. See Francis Galton, Letter to George Darwin, 8 January 1875, UCL Galton Archive, 190A.

[20] Francis Galton, 'Statistics by Intercomparison', 35.

[21] Ibid., 35-37.

[22] Ibid.

[23] Ibid.

[24] Ibid., 41.

[25] Ibid., 42.

[26] The best treatment of Galton's auxological studies is to be found in James Mourilyan Tanner, 'Galton on Human Growth and Form', in *Sir Francis Galton, FRS: The Legacy of His Ideas*, edited by Milo Keynes (Macmillan, 1993), 108-118.

[27] Francis Galton, 'Proposal to Apply for Anthropological Statistics from Schools', *Journal of the Anthropological Institute of Great Britain and Ireland* 3 (1874), 308-311, 311.

[28] Francis Galton, 'Notes on the Marlborough School Statistics', *Journal of the Anthropological Institute of Great Britain and Ireland* 4 (1874), 130-135, 135. The data were given in the preceding paper: Walter Fergus and G. F. Rodwell, 'On a Series of Measurements for Statistical Purposes, Recently Made at Marlborough College', *Journal of the Anthropological Institute of Great Britain and Ireland* 4 (1874), 126-129.

[29] Francis Galton, 'On the Height and Weight of Boys Aged 14, in Town and Country Public Schools', *Journal of the Anthropological Institute of Great Britain and Ireland* 5 (1876), 174-181. This paper was read on 27 April 1875.

[30] The country schools were Marlborough, Clifton College, Haileybury, Eton College and Wellington. The town schools were the City of London, Christ's Hospital, King Edward's in Birmingham and Liverpool College.

[31] At the country and town schools the numbers of boys measured were 1410 and 2394. The statistics gathered at Wellington School were stated but not used in the analysis.

[32] Francis Galton, 'On the Height and Weight of Boys Aged 14', 174.

[33] Tanner was mistaken in claiming that 'this is the paper that introduced the system of percentiles for characterising a distribution, which is one of the two lasting contributions of Galton to the field of auxological analysis' J. M. Tanner, 'Galton on Human Growth and Form' in *Sir Francis Galton, FRS: The Legacy of His Ideas,* edited by Milo Keynes (Macmillan, 1993) 108-118, 111.

[34] Francis Galton, 'On the Height and Weight of Boys Aged 14', 174.

[35] Ibid.

[36] Ibid., 176.

[37] Adolphe Quetelet, *Anthropométrie, ou Mésure des Differentes Facultés de l'Homme* (Brussels: Muquardt, 1870), 266. Quetelet gave his anthropmetric data relative to the height of Belgians, overall and for three age groups. They appear as indices, therefore, relative to a height of 1. For example the index for the distance between the outer extremities of the two eyes is given as 0.056.

[38] Pearson, *Life* II, 125-126.

[39] Francis Galton, Letter to George Darwin, 25 December 1874, Galton Archive, 245/6; also in Pearson, *Life* IIIB; 463.

[40] Galton's ogive is what we would now refer to as an inverse cumulative frequency curve.

[41] Francis Galton, 'Statistics by Intercomparison, with Remarks on the Law of Frequency of Error', *Philosophical Magazine* [4] 49 (1875), 33-46. The lecture was given on the evening of Friday, 27 February 1874.

[42] George Darwin, Letter to Francis Galton, 4 January 1875, UCL Galton Archive, 190A. The letter is actually dated 4 January 1874, but it is clear that Darwin had momentarily forgotten that the year had changed, for he included a correct reference to details on p. 43 of the published paper.

[43] Francis Galton, Letter to George Darwin, 8 January 1875, UCL Galton Archive, 190A.

[44] Indeed when Airy congratulated James Glaisher on the success of his son in the Tripos, he referred to the son as Lee. See J. L. Hunt, 'James Glaisher FRS (1809-1903). Astronomer, Meteorologist and Pioneer of Weather Forecasting: "A Venturesome Victorian"', *Quarterly Journal of the Astronomical Society* 37 (1996), 315-347, 327.

[45] J. L. Hunt, 'James Glaisher FRS (1809-1903). Astronomer, Meteorologist and Pioneer of Weather Forecasting: A Venturesome Victorian', *Quarterly Journal of the Astronomical Society* 37 (1996), 315-347, 328.

[46] Ibid.; idem, 'Glaisher, James Whitbread Lee', *Dictionary of National Biography, 1922-30* (London, 1937), 339-340. Also, John L. Hunt, 'J. W. L. Glaisher, FRS, ScD (1848-1928)', *Quarterly Journal of the Royal Astronomical Society* 37 (1996), 743-757. The sketch of Glaisher's life and career given by Phillip S. Jones in the *Dictionary of Scientific Biography* adds little to that given by Forsyth, Hardy and Hunt.

[47] For a summary of Glaisher's early papers on the method of least squares, see Pritchard, PhD Thesis, 292-293.

[48] J. W. L. Glaisher, 'On a Class of Definite Integrals', *Philosophical Magazine* [4] 42 (1871); Part I, 294-302; Part II, 421-436.

[49] Ibid., 296.

[50] Ibid., 433. Legendre wrote $\Gamma(a,x)$ for $\int_0^x (\log(1/x))^{n-1} dx$, whence $\Gamma(a,e^{-x}) = \int_x^\infty e^{-v} v^{a-1} dv$. See Adrian Marie Legendre, *Traité des Fonctions Elliptiques* (2 vols., Paris: Hachette, 1825-1826), vol. 2, ch.17.

[51] Glaisher's choice of Erfc *x* (the error-function-complement) was made by analogy with the cosine of an angle being the sine of its complement and *vice versa*. During 1871-72, Glaisher was using the gamma function $\Gamma(1+\frac{1}{n}) = \int_0^\infty e^{-x^n} dx$ to investigate $\int_0^\infty \sin x^n dx$, $\int_0^\infty \cos x^n dx$, $\int_0^\infty \sin x^2 dx$ and $\int_0^\infty \cos x^2 dx$, though without invoking Erf *x*. See J. W. L. Glaisher, 'On the Integrals $\int_0^\infty \sin x^n dx$ and $\int_0^\infty \cos x^n dx$', *Messenger of Mathematics* 1 (1871), 106-111, dated 1 September 1871; idem, 'On the Integrals $\int_0^\infty \sin x^2 dx$ and $\int_0^\infty \cos x^2 dx$', *Quarterly Journal of Pure and Applied Mathematics* 13 (1875), 343-348, dated 15 December 1872.

[52] Christian Kramp, *Analyse des Réfractions Astronomiques et Terrestres par le citoyen Kramp*, Strasbourg and Leipzig: Dannbach, 1798-99.

[53] It seems likely that Galton would have been aware of such tables for a few years prior to be learning of the erf notation, for example in Augustus De Morgan, 'Theory of Probabilities', *Encyclopaedia Metropolitana* (1836-37), reprinted as *Treatise on the Theory of Probabilities*, London: Clowes, 1837.

[54] This degree of accuracy is equivalent to 7 significant figures across the range [3.00, 4.50].

[55] Richard Pendlebury, 'Note on Some Definite Integrals', *Philosophical Magazine* [4] 42 (1871), 437-440.

[56] J. W. L. Glaisher, 'Numerical Values of Certain Continued Fractions', *Quarterly Journal of Pure and Applied Mathematics* 13 (1875), 255-259.

[57] The fact that the Galton Archive contains no letter requesting such information suggests a face-to-face contact in the New Year, though between whom is not known.

[58] Francis Galton, Letter to George Darwin, 30 January 1876, UCL Galton Archive, 245/6. Pearson makes no mention of Glaisher at any point in his biography of Galton, except to quote his name in this letter.

[59] Ibid.

[60] J. W. L. Glaisher, 'List of Tables of $\int_0^x e^{-t^2}\,dt$ and Connected Integrals and Functions', *Messenger of Mathematics* 38 (1908), 117-128. Glaisher also conceded that even this is unsatisfactory because each integral should be preceded by the factor $2/\sqrt{\pi}$. The former would then be the error-function and the latter the error-integral or eri *x*. There was even a case, he argued, for replacing erf *x* and erfc *x* by eiq *x* and eiq′ *x*, the q indicating the squaring.

[61] Francis Galton, Letter to George Darwin, 30 January 1876, UCL Galton Archive, 245/6.

[62] Francis Galton, Letter to George Darwin, 17 May 1876, UCL Galton Archive 245/6.

[63] J. W. L. Glaisher, 'List of Integrals of $\int_0^\infty e^{-t^2}\,dt$ and Connected Integrals and Functions', *Messenger of Mathematics* 38 (1908), 117-128.

[64] The surname appears in three freely interchangeable forms in the secondary literature — McAlister, Macalister, MacAlister — though he used the first of them in signing the two extant letters written to Galton during this period. See Donald McAlister, Letter to Francis Galton, 20 & 21 October 1879, Stokes Papers, Cambridge University Library, ADD.7656: M6 & M7. The latter contains reference to a note sent by McAlister to Galton from the British Association for the Advancement of Science meeting in Sheffield. The approach to McAlister had been made before this meeting which was held in the summer of 1879 because the note contained a consideration of the 'practical outcome' of the extended theory.

[65] Francis Galton, *Memories*, 304.

[66] Francis Galton, Letter to George Darwin, 19 June 1877, UCL Galton Archive, 245/6.

[67] Francis Galton, Letter to George Darwin, 8 July 1877, UCL Galton Archive, 245/6.

[68] George Darwin commented just a few days later that 'I have got too much on hand just at present'. See George Darwin, Letter to Francis Galton, 11 July 1877, UCL Galton Archives, 39H.

[69] Francis Galton, Letter to Karl Pearson, 18 August 1910, in Pearson, *Life*, IIIA, 430.

[70] H. J. W. Hetherington, 'MacAlister, Sir Donald', *Dictionary of National Biography 1931-1940*, (Oxford University Press, 1949). E. B. MacAlister, *Sir Donald MacAlister of Tarbet* (London: Macmillan, 1935).

[71] Francis Galton, 'The Geometric Mean, in Vital and Social Statistics', *Proceedings of the Royal Society* 29 (1879), 365-367. Donald McAlister, 'The Law of the Geometric Mean', *Proceedings of the Royal Society* 29 (1879), 367-376.

[72] Francis Galton, Letter to George Gabriel Stokes, 23 October 1879, in Pearson, *Life*, IIIB, 468. Galton could not lend Fechner's book to Stokes because he had lent it to Darwin at the turn of the year. See Francis Galton, Letter to George Darwin, 30 December 1878, UCL Galton Archive, 245/6.

[73] Donald McAlister, 'On the Law of the Geometric Mean in the Theory of Errors', *Quarterly Journal of Pure and Applied Mathematics* 17 (1881), 175-194. Glaisher had started the journal in 1878 and remained its editor for half a century.

[74] As he wrote to Galton in October 1879, one reason that McAlister produced two versions of the paper was that 'my time is so fully occupied that I have not been able to give an example of the method'. See Donald McAlister, Letter to Francis Galton, 20 October 1879, Stokes Papers, Cambridge University Library, ADD.7656: M6.

[75] Francis Galton, 'The Geometric Mean, in Vital and Social Statistics', 366.

[76] Ibid.

[77] Ibid., 366-367.

[78] The 'law of facility' is now referred to as the lognormal distribution.

[79] Donald McAlister, 'The Law of the Geometric Mean', *Proceedings of the Royal Society* 29 (1879), 367-376, 373.

[80] Francis Galton, 'Statistics of Mental Imagery', *Mind* 5 (1880), 301-318. Idem, *Inquiries into Human Faculty and its Development* (London: Macmillan, 1883).

[81] Donald McAlister, 'The Law of the Geometric Mean', *Proceedings of the Royal Society* 29 (1879), 367-376, 373.

[82] Richard Pendlebury, 'Note on Some Definite Integrals', *Philosophical Magazine* [4] 42 (1871), 437-440.

[83] Pendlebury was a Fellow of St John's, 1870-1902, and a College Lecturer in Mathematics when MacAlister was an undergraduate.

[84] Donald McAlister, 'On the Law of the Geometric Mean in the Theory of Errors', *Quarterly Journal of Pure and Applied Mathematics* 17 (1881), 175-194, 186.