Quatrième Année. - Numéro 14 Janvier 1951

Archives Internationales d'HISTOIRE des SCIENCES

Publication trimestrielle
de l'Union Internationale d'Histoire des Sciences

Honorée d'une subvention de l'UNESCO

Nouvelle Série d'ARCHEION

Fondateur : Aldo MIELI

ACADÉMIE INTERNATIONALE
D'HISTOIRE DES SCIENCES
12, Rue Colbert — PARIS - 2e

HERMANN & Cie
ÉDITEURS
6, Rue de la Sorbonne, PARIS-5e

Quantitative Measures of the Development of Science (*)

SUMMARY

The number of scientific papers published each year may be taken as a rough indication of the activity displayed in any general or specialised field of research. Statistics are obtained and analysed for certain cases, from which it appears that during normal times a general field such as Physics increases exponentially to a high degree of accuracy. A specialised field, however, such as the theory of Determinants and Matrices, increases exponentially only to a certain point at which the growth changes to linear variation with time. The growth factor of the exponential portions is such as to double the amount of literature every ten or eleven years in both the general and specialised cases. The effect of the wars is studied in detail and it is shown that the loss in literature is greater than the gain due to increased stimuli. Sudden advances made by individuals do not seem to affect significantly the normal growth of literature in a subject.

Since the usual manner of recording a contribution to scientific knowledge is through the medium of the scientific paper published in some learned journal, one might expect that the number of papers appearing each year would be a useful barometer indicating the amount of activity during that year, and over the range of subject-matter from which a count had been made.

As is customary in such investigations, it is much easier to make a measurement than to ascertain just what has been measured, and considerable caution must therefore be exercised in any

(*) Communication présentée au VI[e] Congrès International d'Histoire des Sciences, Amsterdam, août 1950.

a normal curve of growth being indistinguishable on the diagram. The equation of this period may be written as :

$$\text{Number of papers} = 10{,}000 + 8{,}500 \text{ Exp } (n/15)$$

where n is the number of years elapsed from January 1st, 1900.

The growth constant may alternatively be interpreted as a statement that the total number of papers, measured from a datum level of 10,000, doubles every 10.4 years.

1914-1919 and 1938-1947 : During each of the war periods there is, as one might expect, a fall in the production of scientific papers. In the first period the rate of growth is approximately 85 % of that indicated by an extension of the exponential portion; in the second period the corresponding ratio is of the order of 35 % only. It is perhaps interesting to note that besides the greater magnitude of the loss, recovery was slower in the more recent war, the trend of the curve not returning to normal until 1948.

1948 to present time (1950) : The present absolute magnitude and trend of production of literature is virtually the same as that which would have prevailed about six years ago if the exponential portion had continued without interruption by the war. This is in contradiction with the common assertion that war-time conditions give a stimulus to scientific activity, such stimulus should be indicated by an increased slope of the graph after the period of recovery had been completed. Since this does not appear we must conclude that the net effect of the war has been equivalent to a loss of about six years of activity in this particular field of study. The post-war release of research papers previously kept secret for security reasons does not produce any noticeable increase in slope during the recovery period, but one must not conclude from this that the gain from such source is negligible. Such papers may have an average value greater than those not subject to security regulations, and besides this, it must be remembered that there has been considerable and variable delay in the publication of research papers due to other war-time exigencies.

1900-1914 : The graph rises more rapidly than might be expected from a continuation of the exponential portion to this period. This is probably the consequence of a large change in the organisation of the abstracting journal at the end of the period.

interpretations of the statistics of publication. Perhaps the two greatest difficulties of this method are in deciding whether any particular paper is included in the field under discussion, and if so, whether it is of sufficiently high standard to be counted as a « UNIT » contribution to knowledge.

The boundaries of any particular province of research are necessarily hazy and shifting, but in certain cases special attention has been paid to the delineation of these boundaries for the purpose of producing some comprehensive bibliography or series of abstracts. If this type of compilation extends over a sufficient time interval without much alteration in the basis of selection, it may be conveniently used as a source for the quantitative measures of development in the range covered. In the present work two such collections are studied, the one « Physics Abstracts » covering a wide, general field, and the other « Theory of Determinants and Matrices » dealing with a specialised branch of mathematical knowledge, and based initially on the history of this subject written by MUIR.

The most convenient practice in the assessment of papers is to regard them equal weight, in spite of the fact that some are obviously much more important than others. This will not affect the use of such a count as an index of activity provided that the distribution of merit amongst the papers remains effectively constant during the period studied. So far as could be ascertained this was certainly the case, although the extension of the range to include earlier dates would raise grave doubts about the validity of this assumption.

A GENERAL FIELD (*Physics Abstracts*) FIGURE I

The total number of abstracts appearing in this journal since January 1st, 1900 is illustrated graphically, the slope of the curve at any date, indicating the rate of production of new papers. The use of a cumulative total is beneficial since smoothing is automatically effected without loss of accuracy. The graph falls into a number of distinct regions.

1920-1937 : The growth is exponential to a high degree of accuracy during this period between the wars, the actual variation from a normal curve of growth being indistinguishable on the diagram. The equation of this period may be written as :

$$\text{Number of papers} = 10{,}000 + 8{,}500 \text{ Exp } (n/15)$$

where n is the number of years elapsed from January 1st, 1900.

The growth constant may alternatively be interpreted as a statement that the total number of papers, measured from a datum level of 10,000, doubles every 10.4 years.

1914-1919 and 1938-1947 : During each of the war periods there is, as one might expect, a fall in the production of scientific papers. In the first period the rate of growth is approximately 85 % of that indicated by an extension of the exponential portion; in the second period the corresponding ratio is of the order of 35 % only. It is perhaps interesting to note that besides the greater magnitude of the loss, recovery was slower in the more recent war, the trend of the curve not returning to normal until 1948.

1948 to present time (1950) : The present absolute magnitude and trend of production of literature is virtually the same as that which would have prevailed about six years ago if the exponential portion had continued without interruption by the war. This is in contradiction with the common assertion that war-time conditions give a stimulus to scientific activity, such stimulus should be indicated by an increased slope of the graph after the period of recovery had been completed. Since this does not appear we must conclude that the net effect of the war has been equivalent to a loss of about six years of activity in this particular field of study. The post-war release of research papers previously kept secret for security reasons does not produce any noticeable increase in slope during the recovery period, but one must not conclude from this that the gain from such source is negligible. Such papers may have an average value greater than those not subject to security regulations, and besides this, it must be remembered that there has been considerable and variable delay in the publication of research papers due to other war-time exigencies.

1900-1914 : The graph rises more rapidly than might be expected from a continuation of the exponential portion to this period. This is probably the consequence of a large change in the organisation of the abstracting journal at the end of the period.

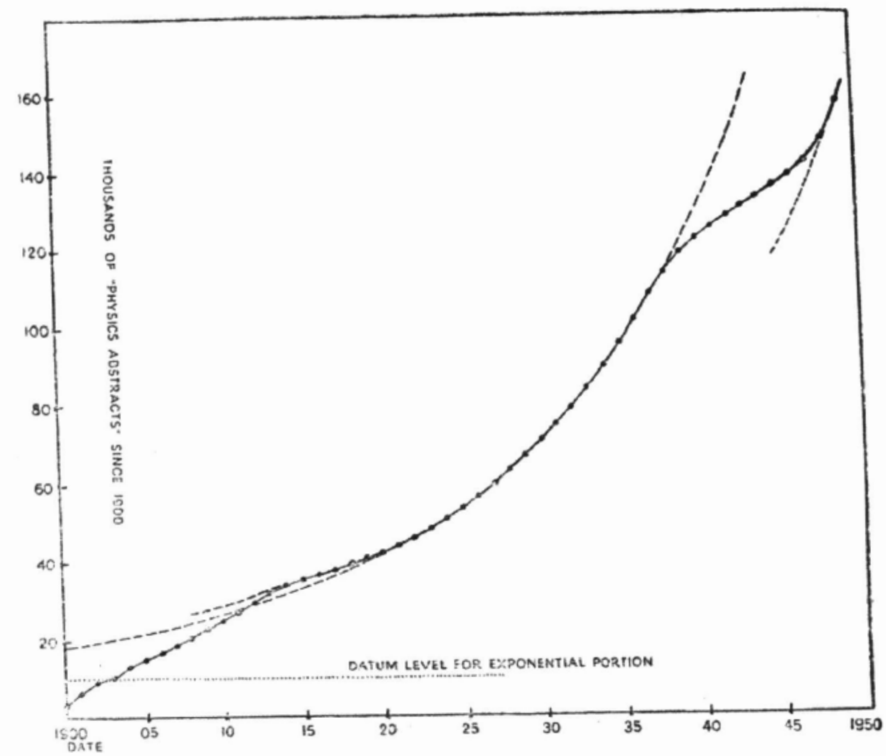

Figure 1. — Total number of "Physics Abstracts" published since Jan. 1st, 1900. **The full curve gives the total, and the broken curve represents the exponential approximation. Parallel curves are drawn to enable the effect of the wars to be illustrated.**

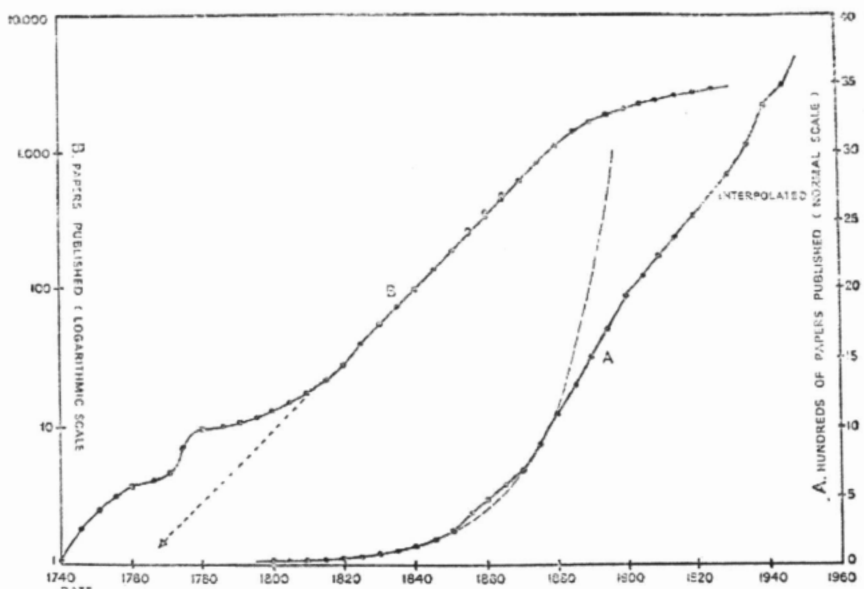

Figure 2. — Total number of papers published on "Theory of Determinants **and** Matrices" since commencement of field.

A) Normal plotting, the broken curve indicating the exponential approximation.

B, Logarithmic plotting to illustrate development before 1840. The broken line **shows** the course if growth had been exponential starting from one paper in 1760.

Prior to the 1914 War papers of engineering research were included in the field of physics, but after this time the boundaries of physical science became more definite and moved to their present position, where many of the papers included previously amongst « Physics Abstracts » would not today find place. One may therefore regard this portion of the curve as representing the genesis of what is now called « physics » from a previous corpus of wider scope. It would be interesting to know whether the datum-level of 10,000 papers obtained from the analysis of the exponential portion of the graph is capable of interpretation as the equivalent number of scientific papers in the field of physics existing prior to 1900. One is tempted to make comparison with the present rate of production of approximately 8,000 papers per annum!

A Specialised Field
(Theory of Determinants and Matrices) Figure II

The source of data for this study is *History of the Theory of Determinants* by Muir, for the period 1693-1919, followed by information collected from *Zentralblatt für Mathematik* and *Mathematical Reviews* for the intervals 1930-1939 and 1940-1949 respectively. The period 1920-1930 is not covered, and no accuracy is claimed for the arbitrary interpolation introduced to give continuity to the diagram.

There are two distinct regions to be considered :

Before 1880 : If Muir's data may be accepted as reasonably complete, and not redundant during this interval, it may be said that the growth has been accurately exponential from an effective epoch of one paper in 1760. This is illustrated on the logarithmic graph which approaches the straight line indicating the normal growth curve for a total number of published papers of only ten. At this level, however, the sample is too small for definite conclusions to be obtained from the remarkable regularity exhibited. In spite of a few papers appearing in the interval 1693-1760 it may be said that the accumulation of literature in this field commences at about 1760 and proceeds exponentially until 1880. Within this range the equation of the curve may be written as :

$$\text{Number of papers} = \text{Exp}\ (n/17)$$

where n is the number of years elapsed from 1760.

Alternatively one may say that the number of papers, starting from unity in 1760, doubled every 11.8 years.

The accuracy of curve-fitting, although striking, is not so good as with « Physics Abstracts », partly because of the smaller sample used, partly because of greater difficulty in setting the boundaries of the field, and partly because of the possible existence of other small but significant sources of variation.

After 1880 : At or about 1880 the character of the curve changes from exponential to linear. Small variations occur from time to time in the linear portion, but on the whole it is evident that the growth of literature no longer exhibits the character of normal growth but is only increasing at a fairly constant rate approximating to that obtaining in 1880. This could be due to a progressively deteriorating completeness in the bibliography used, it is more likely that some other explanation is needed to account for the continuation of this phenomenon over a long period.

The effect of the wars is much less marked than in the previous example; no disturbance is noticeable in 1914-1918, but between 1940 and 1945 there is an equivalent gap of about three years.

Exponential Growth and Linear Growth

It would be unfortunate to generalise from merely two examples, but other investigations indicate that many of the features described above are to be found in other quantitative investigations of the growth of science and scientific literature. For example similar types of curves are obtained for the number of British Patents since the Law Amendment of 1863-1864 and for the membership roll of various learned societies.

In particular a growth constant leading to doubling every ten or eleven years seems to be characteristic of periods of exponential growth, and in the case of specialised fields or activities, these periods are followed by the transition to linear growth. In some cases the field becomes « dead » and growth decreases steadily towards zero.

Exponential growth is a property of systems in which the rate of increase at any time is proportional to the amount already achieved; in the two cases studied this constant of proportionality

Alternatively one may say that the number of papers, starting from unity in 1760, doubled every 11.8 years.

The accuracy of curve-fitting, although striking, is not so good as with « Physics Abstracts », partly because of the smaller sample used, partly because of greater difficulty in setting the boundaries of the field, and partly because of the possible existence of other small but significant sources of variation.

After 1880 : At or about 1880 the character of the curve changes from exponential to linear. Small variations occur from time to time in the linear portion, but on the whole it is evident that the growth of literature no longer exhibits the character of normal growth but is only increasing at a fairly constant rate approximating to that obtaining in 1880. This could be due to a progressively deteriorating completeness in the bibliography used, it is more likely that some other explanation is needed to account for the continuation of this phenomenon over a long period.

The effect of the wars is much less marked than in the previous example; no disturbance is noticeable in 1914-1918, but between 1940 and 1945 there is an equivalent gap of about three years.

Exponential Growth and Linear Growth

It would be unfortunate to generalise from merely two examples, but other investigations indicate that many of the features described above are to be found in other quantitative investigations of the growth of science and scientific literature. For example similar types of curves are obtained for the number of British Patents since the Law Amendment of 1863-1864 and for the membership roll of various learned societies.

In particular a growth constant leading to doubling every ten or eleven years seems to be characteristic of periods of exponential growth, and in the case of specialised fields or activities, these periods are followed by the transition to linear growth. In some cases the field becomes « dead » and growth decreases steadily towards zero.

Exponential growth is a property of systems in which the rate of increase at any time is proportional to the amount already achieved; in the two cases studied this constant of proportionality is such that the whole of the previous work could have been done in about sixteen years if publication had proceeded constantly at the current rate. Linear growth, on the other hand, implies a rate of increase which remains constant and independent of the amount already done.

What is it in the form or content of a field that determines whether its growth shall be exponential or linear? In a certain sense, it may be that the problem is created artificially by the imposition of field boundaries upon the general body of knowledge. To use the metaphor of geographical exploration, sometimes it is *terra incognita* that is being investigated, and sometimes it is familiar territory that is being developed and consolidated. In the former case the law of growth should be exponential because each new discovery brings increased stimulus, in the latter it is more probable that progress would be at steady level, and growth therefore linear. There is then the possible explanation that as the research front advances certain fields become cut off and experience a change from normal to linear growth. If this linear increase ultimately slows down it may be because the field is so far behind the research front as to receive but little stimulus for activity.

The same type of explanation may also be stated in terms of research workers instead of publications. It is reasonable to suppose that the amount published in any year is directly proportional to the number of people engaged in research on the field examined. Approximate calculations confirm this conjecture for recent years. We may therefore interpret the rate of growth as being an index of the scientific man-power mobilised around the field studied. Exponential growth then implies that the expanding subject is attracting new workers at a rate proportional to the activity in that subject, linear growth implies that the number of workers is remaining constant. We have then the picture of workers being attracted to the research subjects in ever increasing numbers, and as the research front moves forward recruiting to some fields slows and stops, so leaving a constant body of workers in those particular fields. The above interpretations are crude, and in any complete picture due allowance would have to be made the replacement of old workers by new, and for the effect of « research schools » on the growth of individual fields of study.

Conclusions

The absence of large deviations from the exponential law indicates that sudden outstanding advances made by individual research workers do not significantly affect the normal pattern of growth in the quantitative measure of development in science. The regularity of the growth should make it possible for reasonable estimates to be obtained of the « size » of general or specialised scientific knowledge at any future date. For example one may extrapolate from present information to an approximate knowledge of the number of research physicists required in twenty years time, or of the size of a scientific library at that same time. In both cases, with the usual limitations of assumed constant growth and the absence of disturbing factors (such as wars) the estimated magnitude should be just less than four times the present value.

More information is needed to establish or disprove the constancy of the coefficient of increase in exponential growth. If this is a universal constant of the order of eleven years for doubling to take place this same magnitude must surely be involved in other measures associated with the development of science, and some connection may be sought with other types of numerical estimation.

The change from exponential to linear growth is an indication of the removal from the research front of the field being examined. This should enable one to diagnose the nature of changes currently taking place, and to plan accordingly in the disposition of research facilities at university and other laboratories. For example, it would be interesting to ascertain whether the present dominance of nuclear physics has produced a transition to linear growth in those portions of physical science not directly connected with nuclear and atomic studies.

Certain sources of error in any quantitative measures as those proposed must be emphasised. As already pointed out there is some difficulty in deciding the boundaries of any given field and in weighting the value of published papers in that field. There is also an unknown and variable delay between the research work and the publication of the paper or abstract. Another factor to be considered is the consolidation of previous knowledge in textbooks and reviews, for it is by this means that the research paper passes eventually into the body of knowledge that must be assimilated before the student can reach the accelerating research front. The process of consolidation must therefore be continuous and increasing for exponential growth to be maintained. If there is a limit to the extent to which previous work may be assimilated through reviews and other means the implication is a rapidly increasing bulk of scientific literature to be read by the student before research in any field, however narrow, may be commenced. One may wonder whether the present difficulties of overspecialisation in scientific education are not due to an overall approach to a limit beyond which exponential growth is no longer possible without drastic change in the structure of science.

References

Science Abstracts, Section A (Physics). Institute of Electrical Engineers, London.

Muir : *History of Determinants*. Macmillan. Vol. I to IV, 1906-1923.

Muir : *Contributions to the History of Determinants*. Blackie, 1930.

Singapore.

Derek J. Price.