MARIA NOWAKOWSKA

EPIDEMICAL SPREAD OF SCIENTIFIC OBJECTS: AN ATTEMPT OF EMPIRICAL APPROACH TO SOME PROBLEMS OF META-SCIENCE

ABSTRACT. The paper deals with problems of prediction of spreading out of scientific objects, such as theories, hypotheses, methods etc. Two models for predicting changes in number of publications dealing with the given subjects in consecutive years are suggested; these models are based on the theory of epidemics.

0. INTRODUCTION

In paper [36] a general outline of an approach to meta-science was presented; meta-science was interpreted there as a theory of optimal control of development of science. Several research projects have been suggested; the present paper constitutes an exemplification of one of them, namely of the problem of prediction of spread of certain scientific objects, such as theories, methods, hypotheses, instruments, etc.

The aim of analysis is to investigate the process of changes in numbers of publications which appear in successive years and deal with a given scientific object. This process is interpreted as an epidemic, in which new infection corresponds to the appearance of new publication, resulting from 'infection by an idea'. The intuition behind such an identification is based on analogies between hypotheses concerning the mechanisms of spread of epidemics, and hypotheses concerning the mechanisms underlying the development of population of publications devoted to a given subject (dealing with a given scientific object).

As regards epidemics, it is generally assumed (see for instance [26], [35]) that the number of infections in next period of time depends both on the actual number of infectives, and on the actual number of susceptibles, i.e. those who still might get infected. Somewhat more precisely, it is assumed that the more infectives are present, the quicker is the rate of increase of the epidemic [1], and the less susceptibles present, i.e. more people have already been infected, the slower is the rate of increase of the epidemic.

As regards the population of publications dealing with a given subject, one could conjecture that the number of publications in next period of

Theory and Decision 3 (1973) 262–297. *All Rights Reserved*

time (say, one year) will depend both on the number of papers which appeared recently, and on the 'degree to which the subject has been exhausted'; in other words, when one considers the total number of papers on the subject which appeared up to a given moment, then the following regularity should be observed: the more papers appeared recently, the quicker is the rate of increase of the total number of papers, and the more papers have already appeared altogether, i.e. the more is the subject exhausted, the slower is the rate of increase of number of publications.

To avoid somewhat cumbersome repetitions, we shall from time to time speak merely of 'papers' or 'publications'; it will be tacitly understood that we always refer to papers dealing with a given subject.

In most general terms, the above hypothesis implies that the numbers of publications appearing in successive periods of time should first increase, and after reaching a maximum, start decreasing, as the problem becomes more and more exhausted.

In the sequel, this conjecture will be specified in form of assumptions of two models for change in number of publications. These models will allow us to determine the theoretical behaviour of the curve reflecting changes in numbers of publications, and – after estimating some parameters – to build predictions concerning the future behaviour of the process.

1. Empirical Data

In this section we present histograms for bibliographies concerning

(a) Wechsler tests
(b) theory of games
(c) theory of measurement
(d) multivariate analysis in psychology
(e) psychology of creativity.

The choice of these particular subjects was determined both by the easy access to reasonably complete bibliographies, and by personal interest of the author.

a. *Wechsler Tests*

The histogram (see Figure 1) is based on bibliography [24], which appeared in 1961, and which comprises the period 1939–60.

The data for 1960 were omitted, as they may have been incomplete.

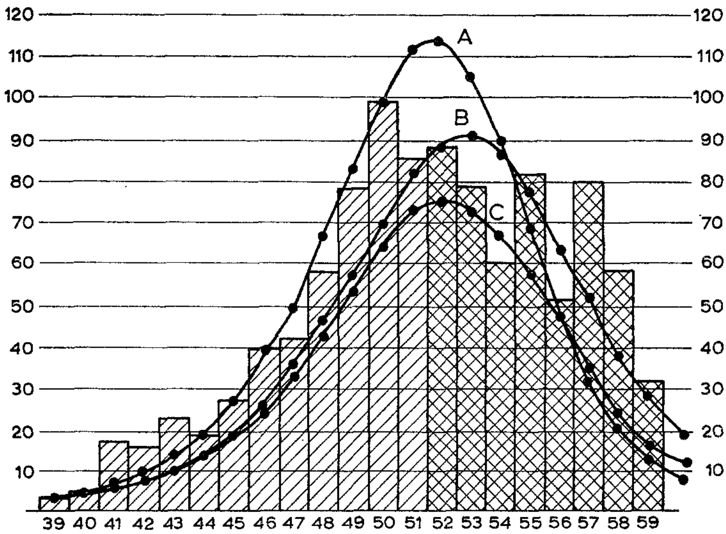

Fig. 1. Prediction for the dynamics of changes of numbers of publications concerning the Wechsler tests. Source: J. Hoskovec, Z. Kanka, *Wechslerovy zkousky W-B, WAIS, WISC Bibliografie*, Prague 1961. Predictions are based on data concerning years 1939–51, and compared with actual data for years 1952–59 (doubly shaded area). Number of papers in the period 1939–59 – 1047; Number of papers in the period 1939–51 – 516; Predictions for values of parameters: A: $N=1800$, $\tau=1.448$; B: $N=1790$, $\tau=1.390$; C: $N=1500$, $\tau=1.390$.

The book [24] gives complete bibliography of tests: Wechsler Adult Intelligence Scale (WAIS, year of appearence 1939), Wechsler Intelligence Scale for Children (WISC, year of appearance 1949), and Wechsler-Bellevue (W-B, year of appearence 1955).

In general, the bibliography comprises primarily the editions of tests in various languages, and publications concerning diagnostic values of these tests, correlations with other tests, and objective criteria, such as EEG, etc.

It should be pointed out that the majority of the papers are more or less of routine character as regards their methodology: they present results obtained by application of one of several simple experimental schemes, for which there exist ready-made statistical techniques.

The large number of publications (1047 in twenty years) is connected both with great interest caused by Wechsler tests (up-to-day they are one of the best available intelligence measures), as well as with the standarization of research procedures.

From Figure 1 it may be seen that the period of maximum interest in Wechsler tests appeared in years 1949–57, while in the following years one may observe gradual 'exhaustion of the topic'.

b. *Theory of Games*

The histogram on Figure 2 is based on bibliography given in book [32]. As stated by the authors of this book, their bibliography is nearly complete as regards theory of games; in histogram, the data for 1957 and 1958

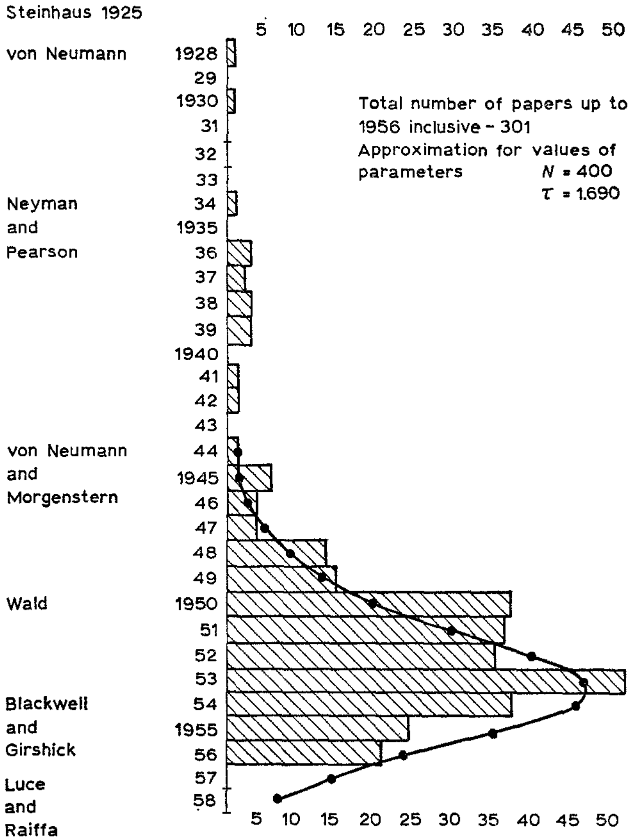

Fig. 2. Publications concerning theory of games Source: R. D. Luce, H. Raiffa, *Games and Decisions*, New York 1958.

were omitted nevertheless, as these were the years immediately preceding the edition of the book.

Compared with the preceding bibliography the number of positions is rather small (301); all, or nearly all of them are of theoretical character, and were written by mathematicians. One can easily point out the names of founders and principal contributors to the theory of games; they are marked next to time axis on histogram 2. For technical reasons, on this and some of the next histograms, the time axis was drawn going vertically down.

The 'biography' of theory of games is, roughly, the following: in 1928 J. von Neumann [51][2] (and three years earlier Steinhaus [43] in Poland) published first papers which contain abstract formulation of problem of game, and suggest the principle of minimax as a criterion for choice of strategy.

For a considerable time (until 1944) these papers received very little attention; thus, the period of 'social incubation' of theory of games was rather long.

In the 30's there appear a series of papers by Neyman and Pearson (see [34]) concerning the theory of testing statistical hypotheses. Today these papers are included in the domain of theory of games, or – more general – decision theory (statistical problems are interpreted as 'games against Nature'). In one of these papers Neyman and Pearson [34] suggest using the minimax principle as a criterion for choice of statistical test (see references to Chapter I in [27]).

The research concerning statistical decisions was carried out in 1939–51 by Wald, and crowned in 1950 by the appearence of his book *Statistical Decision Functions* [53].

In 1944 there appears the first, and in 1947 the second edition of the book by von Neumann and Morgenstern *Theory of Games and Economic Behaviour* [52], in which the authors present general principles of theory of games, and apply the concepts introduced to economy.

Outlining at once a rich theoretical domain, with appealing possibilities of application to social sciences, caused the 'epidemical' increase of interest in theory of games (from 4 papers in 1947 to 37 papers in 1950). The most intense development occurs in years 1950–54. During this period there appears the greatest number of new names; the most notable contributors are Blackwell (connections with statistics), Bellman (foundations of the theory of dynamic programming), and Nash and Shapley (N-person games).

This period ends in appearance of the book by Blackwell and Girshick '*Theory of Games and Statistical Decision Functions*' [2] (1954), which contains the first synthesis of mathematical aspects of theory of games.

Next, starting from 1954 Luce publishes a series of papers concerning the foundations of theory of games, in particular utility theory; in 1958 there appears the book by Luce and Raiffa *Games and Decisions: Introduction and Critical Survey* [32], giving a critical appraisal of foundations of theory of games and its applications to social sciences.

It seems that the growth of number of papers dealing with theory of games, and the subsequent decline of this number starting from somewhere around 1955 is connected with two phenomena. One of them is, according to the main hypothesis of this paper, the exhaustion of the subject: for zero-sum two person games the majority of the problems have already been solved; for other games, considerable difficulties were encountered in building a theory satisfactory from either normative or descriptive point of view.

The second factor which influenced the development and the loss of interest in theory of games is connected with the change of interest in theory of games by non-mathematicians. Since the starting point of the theory of games is the analysis of situation of conflict, it was generally expected that this theory would be applicable to study of such situations in social sciences. This caused originally a great demand for mathematicians dealing with theory of games: the demand may be judged by relatively large number of papers in bibliography in [32] which were probably financed by army, and other institutions interested in potential applications of theory of games to their problems. Thus, the interest in theory of games, at least in the United States was reflected in financing mathematicians. As it gradually turned out that theory of games cannot serve as satisfactory universal model of situations of conflict, the interest of non-mathematicians (and consequently, also that of mathematicians) faded.

At present only those branches of theory of games which found fruitful applications, and became developed into self-contained theories are intensely expanding; notably the theory of linear programming and theory of dynamic programming.

c. Theory of Measurement

The histograms (Figures 3 and 4) were based on combined bibliographies of two positions: the monograph (1966) *Basic Concepts of Measurement* by Ellis [13] and the paper *Basic Measurement Theory* by Suppes and

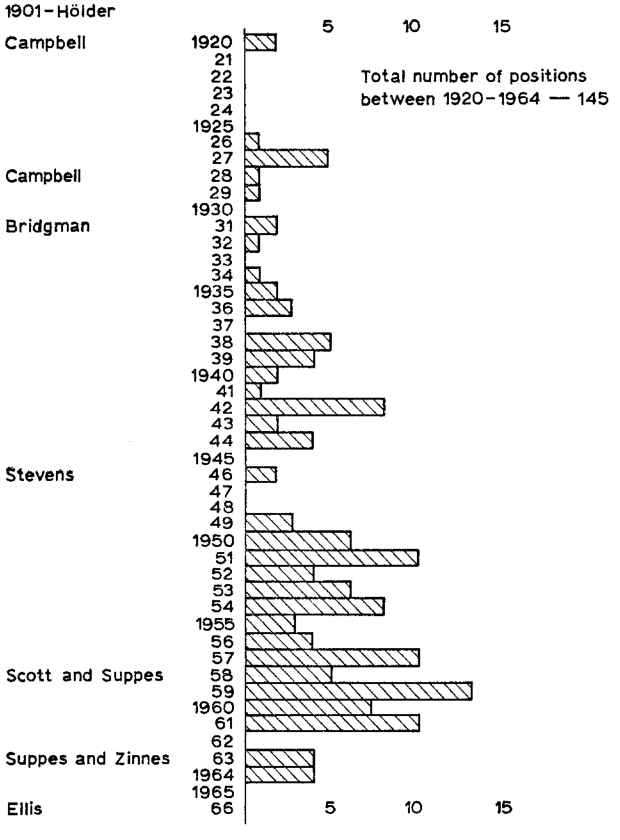

Fig. 3. Bibliography of the theory of measurement Sources: B. Ellis, *Basic Concepts of Measurement*, Cambridge 1966; P. Suppes and J. L. Zinnes, 'Basic Measurement Theory', in *Handbook of Mathematical Psychology*, Vol I (ed. by R. D. Luce, R. R. Bush and E. Galanter), New York 1963.

Zinnes [46]. Both of these positions give a synthesis of problems of theory of measurement, the latter understood as that branch of knowledge, which deals with the question: 'what is measurement' (as distinct from the branches which deal with questions 'how to measure', and 'what can be inferred about the value measured from the value observed'. The latter question is studied in the theory of statistical estimation; the theory dealing with the former question bears no specific name).

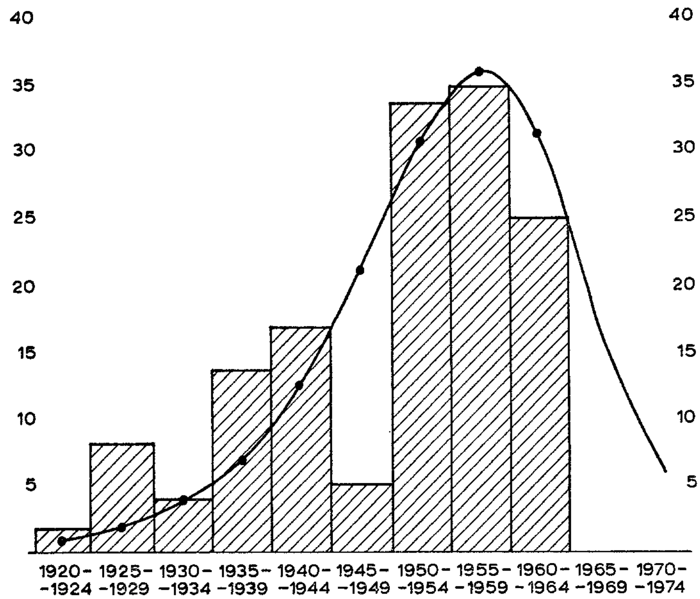

Fig. 4. Bibliography of the theory of measurement Data from Figure 3 grouped in five-year periods. Approximation for values of parameters $N = 200$, $\tau = 1.93$.

As distinct from the previous two bibliographies, the one presented in this section cannot be considered complete. However, it is not likely that the complete bibliography would be larger by any considerable amount, so that certain conclusions can be drawn from histograms 3 and 4, incomplete as they may be.

First thing to be noted is the small number of papers (145); this is probably due to the fact that theory of measurement deals with epistemological problems mainly, and (at least until recently) was of interest mostly for philosophers.

The first papers on theory of measurement (Hölder, 1901; Campbell, 1920, 1928; Bridgman, 1931) were connected specifically with the needs of physics, where (parallel to the problems of increase of precision of measurement) there appeared problems of epistemological, logical and methodological status of measurement.

The basic development of theory of measurement is, however, connected not with physics, but with psychology.

The measurement in psychology was introduced rather early, and in a sense, by two ways, which remain more or less independent until today, constituting two different schools of thought in psychology.

On the one hand, towards the end of XIX Century Galton, and Binet in 1904 introduced the test measurement in psychology: this measurement allowed to assign to each subject tested a value on a certain numerical scale. Oddly enough, for quite a long time the testing methods developed intensively in spite of the fact that there was no consistent theory of such measurement. Such a theory has been suggested as late as in 1950 by Gulliksen (*Theory of Mental Tests* [23]), and its new version appeared in 1968 (Lord and Novick, *Statistical Theories of Mental Test Scores* [28]).

On the other hand, towards the end of XIX Century, Wundt and Fechner gave the origin to another approach, called psychophysics. Its main object was the so-called scaling, i.e. assigning numerical values to stimuli in such a way, that these values should reflect perception of certain properties of these stimuli. The reflexion on the types of scales obtained in this manner led Stevens (1946) to formulate the general theory of measurement ('On the Theory of Scales of Measurement' [44]).

The basic contribution of Stevens consists of defining the measurement, and introducing the classification of scales of measurement, this classification being based on admissible transformations of these scales (such as, for instance, change of unit, etc.).

The theory of measurement suggested by Stevens was subsequently formalized by Scott and Suppes in 1958 ('Foundational Aspects of the Theories of Measurement' [38]). These results were later presented in an expanded form in paper [46] by Suppes and Zinnes (1963) *Basic Measurement Theory*. According to this approach, the scale of measurement is a model (in the sense used in foundations of mathematics, as introduced by Tarski) for the set of empirical relations defined on a given set of objects.

The philosophical aspects of theory of measurement are discussed in detail by Ellis in his monograph *Basic Concepts of Measurement* (1966) [13].

It is worth remarking that in spite of development of theory of measurement, the epistemological status of the test measurement has not been satisfactorily settled as yet.

When one analyses the histograms on Figures 3 and 4, one can see that the problems of theory of measurement have not been studied too extensively. This is due partially to the fact that these problems are difficult, and also to the fact that the results obtained were of no direct consequence: the methodology of measurement could very well develop independently of philosophical reflexions concerning the epistemological status of measurement; this is true both for physics and psychology.

On the other hand, it appears that the problems of theory of measurement become slowly exhausted: the period of maximum of interest falls sometime in between 1950–60, and in subsequent years one notes a decline in numbers of publications.

The decrease in numbers of publications in the period 1945–49 is probably of a local character, and is connected specifically with the war, which stopped the inflow of new scientists in years directly following the war.

d. *Multivariate Analysis in Psychology*

The histogram (see Figure 5) was prepared from bibliography of book [10]; it comprises 1505 positions, mostly of methodological character (the data for 1964, 1965 and 1966 were omitted, as referring to the years immediately preceding the edition of the book). In spite of a large number of positions, this bibliography is not complete; moreover, it may be expected to be rather biased (some authors, representing the same direction of research and the same methodological point of view are rather poorly represented: thus, for instance, the editor of the book, R. B. Cattell, is represented by 142 positions, while Guilford and Eysenck only by 22 positions each, and Gulliksen only by 3 positions).

It appears, however, that in spite of the above objections, the bibliography reflects certain general trends in methodology of multivariate analysis.

According to Cattell, the founders of this methodology are Galton and Spearman; it developed, in a sense, as an opposition to experimental methodology of Wundt-Pavlov type. As distinct from the bivariate controlled experiment recommended by the latter, the methodology of multivariate analysis is based on statistical techniques such as analysis of variance, regression analysis, and factor analysis.

The most prominent representatives of multivariate analysis are Spearman, Thurstone, Thorndike, Cattell, Guilford and Eysenck (the periods of their activity are marked on the right hand side of diagram 5, and the dates of appearance of most important books – on the left hand side, next to the time axis. For references about these books, see the list of bibliography at the end of this paper).

It follows from the histogram, that in spite of the efforts of many prominent specialists, the multivariate analysis was spreading out somewhat slowly; this is due to the conceptual difficulties, as well as to the

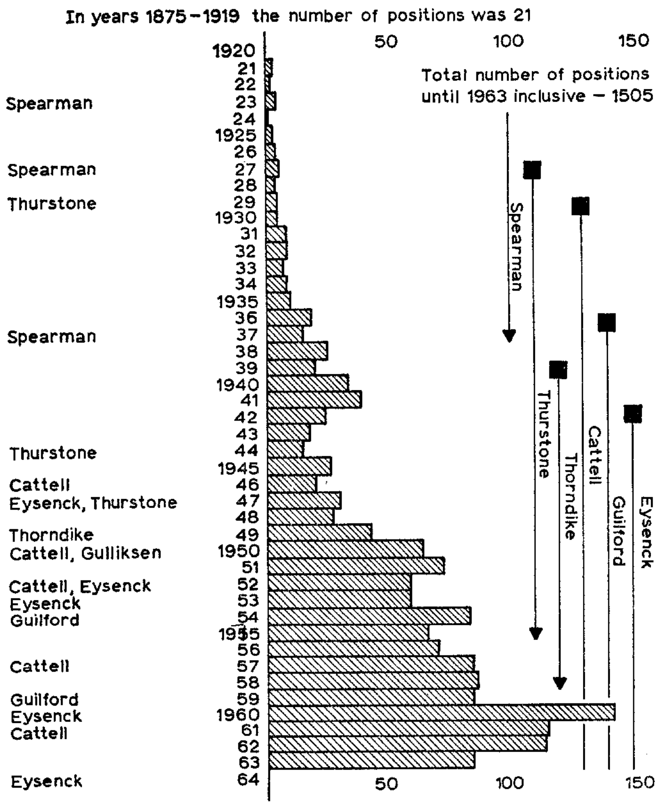

Fig. 5. Publications on multivariate analysis in psychology Source: *Handbook of Multivariate Experimental Psychology* (ed. by R. B. Cattell), Chicago 1966.

criticism (lasting till today) directed against factor analysis. An additional cause of slow progress was the technical (computational) difficulty connected with factor analysis; it seems that it is only in the 50's, with more common use of computers, combined with intense editorial activity of mature already pioneers of factor analysis, that one notes a rapid increase in numbers of publications, with maximum in years 1960–62.

The histogram seems to indicate, however, a certain tendency towards decline in recent years. This may suggest that the problems of methodology of multivariate analysis are close to being exhausted.

e. *Psychology of Creativity*

The histogram (Figure 6) was prepared from bibliography in the book [45]

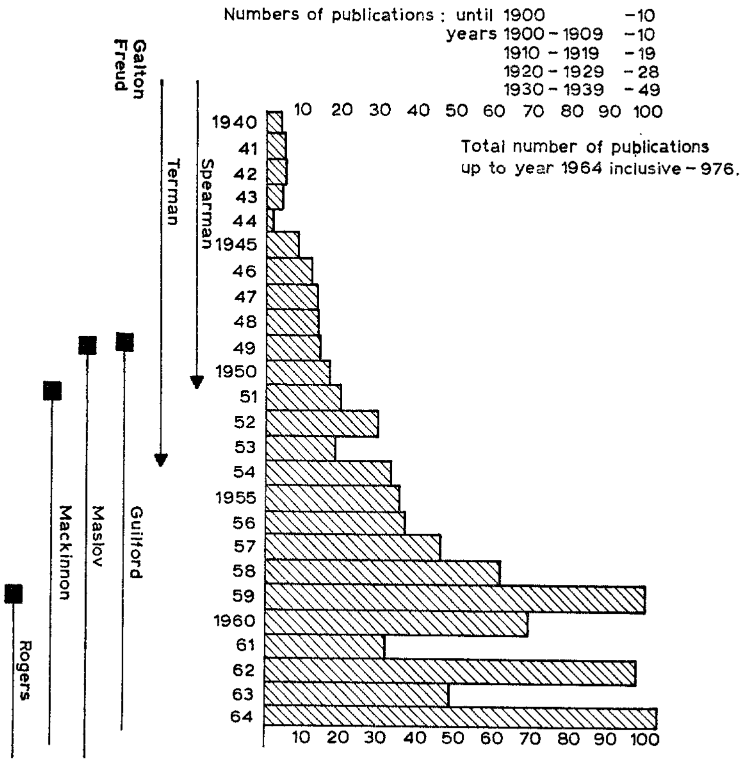

Fig. 6. Publications on psychology of creativity Source: A. Strzałecki, *Wybrane zagadnienia psychologii twórczości* (Selected Topics in Psychology of Creativity), Warszawa 1969.

'*Wybrane zagadnienia psychologii twórczości*' ('Selected Topics in Psychology of Creativity') by Strzałecki. As in the previous bibliographies, the data for 1965 and 1966 preceding the edition of the book were omitted. Thus, the histogram comprises 976 items. This bibliography is far from complete: as stated by Strzałecki, in 1965 there appeared a bibliography of papers concerning creativity, with some 4000 positions.

According to intentions of Strzałecki, the book [45] is devoted mostly to American studies on creativity, which "... concern mainly the description of traits of creative people, hence for instance the role of talents and their structures in creative thinking" ([45], p. 4).

Thus, the bibliography comprises the papers concerning not only creativity, but also talents, personality, motivations, etc.

The periods of activity of most prominent, according to Strzałecki, psychologists dealing with creativity, are marked on Figure 6 next to time axis.

As distinct from the preceding histograms, no tendency towards decline is discernible; violent fluctuations in numbers of positions in years 1961–64 are connected, most likely, partially with periodical conferences on problems of creativity organized in the United States, and partially with fluctuations of financial support given to research. It seems, therefore, that the problems of psychology of creativity are far from being exhausted.

2. THEORETICAL ANALYSIS

In this section we shall outline the methods of building predictions, based on analogous methods used in description and prediction of epidemics.

It is worth noting that mathematical theories of epidemics are particular cases of a more general model, the so-called birth-and-death process (see [19]). Simplifying the matter somewhat, birth-and-death processes describe the evolution of size of population of abstract 'particles' (sometimes one also speaks of a certain abstract 'system', which can change its 'state'). The particles in birth-and-death process may 'give birth' to other particles, or 'die' (as a result, the 'system', interpreted as population of particles, may change its 'state', the latter identified with the number of particles in population).

Depending on specific assumptions on probabilities of particular events which may occur to particles (or, equivalently, changes of state of the system), one obtains different variations of birth-and-death processes.

The above described model is very general and flexible, and has been successfully applied to many phenomena, such as nuclear chain reactions, penetration of cosmic rays through various media, development of biological populations, operation of telephone exchanges, or other servicing systems, epidemics, etc. In each case, the application was obtained by a suitable interpretation of the concept of particle or system (see [19]).

In models of epidemics, the particles are interpreted as human beings; the state of the system is (in models of epidemics on finite populations, see for instance [26]) identified with a pair of numbers (x, y) where x is the number of infectives, and y is the number of susceptibles. In a time interval short enough to be able to neglect the probability of two or more infections of recoveries within this time interval, the system may pass from state (x, y) to one of the states $(x-1, y)$ or $(x+1, y-1)$. In other words, one assumes that no infections, deaths or recoveries occur simultaneously.

The passage from state (x, y) to the state $(x-1, y)$ corresponds to death or recovery of one infective (usually the term 'removal' is employed), hence the number of infectives decreases from x to $x-1$, and the number y of susceptibles remains unchanged[3].

Passing from the state (x, y) to the state $(x+1, y-1)$ corresponds to a new infection: the number of infectives increases from x to $x+1$, and the number of susceptibles decreases from y to $y-1$.

It is assumed that the probability of passing, during a short time interval, from (x, y) to $(x-1, y)$ is proportional to x, and that of passing from (x, y) to $(x+1, y-1)$ is proportional to both x and y.

These assumptions allow us to determine the probability that at a given moment (counting from the beginning of the epidemic) there will be exactly k infectives; one may therefore build predictions in probabilistic terms.

In spite of conceptual simplicity, this model leads to exceedingly complicated formulae (see, for instance, [20]). One can, however, achieve a relative computational simplicity by considering, so to say, a deterministic version of this model; the prediction is then obtained in terms of expected values (hence it is less exact). To start with, one should note that under the above assumptions expected changes of number of infectives in short time intervals are proportional to the lengths of these intervals. In other words, the expected net change (positive or negative) of number of infectives during the period of, say, two days is twice as large as the expected change during one day. More precisely, if at a certain moment t the state of epidemic is (x_t, y_t), then for a sufficiently short time interval Δt one may expect that the number of infectives $x_{t+\Delta t}$ at the moment $t+\Delta t$ will be equal to $x_t - ax_t\Delta t + bx_t y_t \Delta t$, while the number of susceptibles $y_{t+\Delta t}$ will be equal to $y_t - bx_t y_t \Delta t$; here a and b are appropriate coefficients of proportionality. The justification of the above formulae is the following: $ax_t\Delta t$ represents the expected number of individuals who either die or recover during the interval $(t, t+\Delta t)$, while $bx_t y_t \Delta t$ is the expected number of new infections. Thus, the number of infectives x_t will decrease by the amount $ax_t\Delta t$ and will increase by the amount $bx_t y_t \Delta t$; the number of susceptibles y_t will decrease by the amount $bx_t y_t \Delta t$.

Thus, the equations are $x_{t+\Delta t} = x_t - ax_t\Delta t + bx_t y_t \Delta t$ and $y_{t+\Delta t} = y_t - bx_t y_t \Delta t$. It ought to be remarked, that the quantities on the left hand side are expectations of the corresponding variables at the moment $t+\Delta t$, while x_t and y_t on the right hand side are actual values of the process.

A particular case of the above model of epidemic is obtained when one assumes that neither death nor recovery is possible, and only new infections may occur; in terms of parameters it corresponds to the case $a=0$. Such a model is considered by Daley in paper 'Concerning the Spread of News in a Population of Individuals who Never Forget' [12]. Daley investigates spread of news, such as gossips; individuals who had not heard the news are 'susceptible' and those who heard the news are 'infectives'. Recovery is not possible, as it is assumed that the individuals have perfect memory, and never forget.

In this paper we shall suggest two models for analysis of the dynamics of changes in numbers of publications on a given subject. The first of these models, in which only the population of publications is considered, will be based directly on Daley's model. In the second model, somewhat more complex, simultaneous development of two populations will be considered: that of publications on a given subject, and that of scientists dealing with the subject in question. This model will be a combination of Daley's model applied to population of papers, and the model of epidemic with possible 'recoveries' or 'deaths' applied to population of scientists. 'Death' as well as 'recovery' will be interpreted simply as abandoning the subject by a scientist for whatever reasons.

MODEL I. When one considers the model of Daley, or equivalently, model of epidemic with $a=0$, one obtains the equation (taking $\Delta t=1$):

$$(1) \qquad x_{t+1} = bx_t\left(N - \sum_{i=1}^{t} x_i\right).$$

We shall interpret $x_1, x_2, \ldots$ as numbers of papers concerning the subject which appear in successive periods of time; here b and N are certain parameters.

It seems that the length of period of time for observation of x_t cannot be smaller than one year, as it is not possible to pinpoint the date of appearence of a paper with precision better than up to one year.

Thus, the expected number x_{t+1} of papers in year $t+1$ is proportional both to the number x_t of papers which appeared in year t, and to the number

$$N - x_1 - x_2 - \cdots - x_t = N - \sum_{i=1}^{t} x_i$$

which expresses, in a sense, 'the degree to which the problem has already been exhausted'. The more papers appeared up to a certain moment of time, the smaller is the above number.

In general, the behaviour of process $x_1, x_2, \ldots$ is such as presented on Figures 1, 2 and 4 (in the latter, the length of the unit of time is five years). The values of x_t initially increase, and then decrease. Of course, the actual shape of curve depends on the values of parameters b and N.

a. *Estimation of Parameters b and N*

For a given empirical histogram, the evaluation of parameters b and N may be obtained by method of least squares; for such 'fitting' of curve (1) to empirical data it is necessary to use electronic computers. To get a rough estimates of b and N, or at least, to get an orientation as to their order of magnitude, one can apply the following reasoning.

For initial moments of time, that is, for small t, the sums $x_1+\cdots+x_t$ are small in comparison with N. We can therefore write

$$x_2 = bx_1(N - x_1) \approx bNx_1$$
$$x_3 = bx_2(N - x_1 - x_2) \approx bNx_2 \approx (bN)^2\, x_1$$

and so on. In general, for small t we have approximately

$$x_{t+1} \approx (bN)^t x_1,$$

that is, for initial periods of time, the growth of number of publications is approximately exponential.

Thus, we can write for small t

$$\tau = bN \approx \sqrt[t]{x_{t+1}/x_1}. \tag{2}$$

Formula (2) provides us with an approximation of product bN, which was denoted here by τ.

Next, to estimate N one can use the fact, that in the neighbourhood of 'peak' of curve we have

$$\frac{x_{t+1}}{x_t} \approx 1.$$

Using the basic Formula (1) we get

$$\frac{bx_t\left(N - \sum_{i=1}^{t} x_i\right)}{x_t} = b\left(N - \sum_{i=1}^{t} x_i\right) \approx 1,$$

that is,

$$N \approx \frac{1}{b} + \sum_{i=1}^{t} x_i.$$

It follows that

$$N \approx \frac{N}{bN} + \sum_{i=1}^{t} x_i = \frac{N}{\tau} + \sum_{i=1}^{t} x_i$$

hence

$$N \approx \frac{\sum_{i=1}^{t} x_i}{1 - \frac{1}{\tau}}. \tag{3}$$

For example, in histogram on Figure 1 (Wechsler tests), to estimate $\tau = bN$ the value $x_{11} = 78$ was taken for $t = 11$. Since $x_1 = 3$, by Formula (2) we obtain

$$\tau \approx \sqrt[10]{78/3} = 1.385.$$

To estimate the value of N, Formula (3) was applied: it was assumed first that the peak occurs in years 1950–51. As the number of papers up to 1950 inclusive was 430, the estimate given by (3) yields value N of the order of 1400.

Next, if one assumes that the peak occurs in years 1951–52, the total number of papers up to year 1951 inclusive was 516, and Formula (3) gives N of the order of 1800..

The values of τ and N estimated by this method for the three histograms on Figures 1, 2 and 4 served as a starting point for searching for curves which would provide 'the best fit' to empirical data. The histograms on Figures 2 and 4 were approximated by curves (1) by trial and error method (for Figure 1, the procedure was somewhat different and will be discussed below).

Because of technical difficulties connected with calculation by hand, and because the aim of this paper was to outline methods rather than analyse particular sets of empirical data, no attempt has been made to find 'best fitting curves' (i.e. curves which give minimum of sum of squared deviations between actual data, and value on the curve).

The curves were not fit to the data on histograms on Figures 5 and 6. In both cases the bibliographies were not complete, and in addition to that, the data on Figure 6 do not show any decline, so that it is impossible to localize the peak, while the suggested methods of estimation are based on the knowledge (at least approximate) of joint number of papers which appear before the moment of reaching the maximum.

Incidentally, one could try to estimate the value of N from observation of the deviation from exponential growth for initial values of t. Such an estimate, however, would be much less precise than the one based on knowledge of the position of maximum.

b. *Interpretation of Parameters*

It follows from equation (1) that $\tau = bN$ is a parameter characterizing the dynamics of initial growth of number of publications (the interpretation of parameter b separately will be given in model II). The parameter N may be interpreted as the number of papers which have to appear in order to exhaust the problem.

One can also imagine another interpretation of parameter N, which will be used in model II. It may be assumed, namely, that the problem under consideration may be partitioned into N 'sub-problems', such that solving any of them is worth a separate publication, and these subproblems are solved successively by scientists working on the subject.

c. *Interpretation of Results*

In general, the behaviour of curves on figures 1, 2 and 4 suggest that the model describes reasonably well the phenomenon of development and decline of the three problems considered: Wechsler tests, theory of games, and theory of measurement. If one is willing to assume that the model presented provides an adequate description of the dynamics of changes of numbers of publications on other subjects also, then one can apply the model for prediction of future behaviour of process of changes of numbers of publications on a given subject; the prediction, of course, has to be based on such a fragment of history of the process to be predicted, which allows us to evaluate parameters τ and N, and consequently, to fit the curve (1).

To get a sketch what type of prediction is available by means of model I, the following 'experiment' was performed: having complete data for Wechsler tests (up to year 1959), the data for 1952–59 were omitted, and three prediction curves were constructed on the basis of data for years 1939–51. The predictions thus obtained were then compared with the actual data for 1952–59.

On the basis of data for 1939–51 approximate values of τ and N were calculated; roughly, it turned out that τ is about 1.4 and N is somewhere between 1400 and 2000. There curves were drawn for three sets of values of parameters τ and N; the choice of particular values of parameters was made on the ground of whatever experience was gained in fitting these and other curves (on Figures 2 and 4). However, the aim was not to find the best fitting curve for data for years 1939–59, but rather find out to what extent the modification of values of parameters (within the limits obtained from rough estimations (2) and (3)) influences the form of curves, and how far they would fall from empirical data.

As can be seen from Figure 1, at the beginning of 1952, when the total number of papers published on Wechsler tests was equal to 516, it could have been predicted that the total number of papers which will ever appear on Wechsler tests is of the order of 2000, and that starting from sometime near 1960, the yearly numbers of papers on Wechsler tests will not exceed 20.

When one takes into account that a rather specific fact is predicted for 8 yr ahead, and that this prediction is valid, one can accept such a prediction as quite satisfactory.

MODEL II. We shall now suggest a somewhat more complex model, whose aim will be

(1) prediction of numbers of publications in successive years, as in model I;

(2) inference concerning certain variables, whose direct observation may be quite difficult, and at any rate, would likely be subject to serious errors, such as the number of scientists working on a given subject in the given year (counting also those, who have not succeeded in obtaining publishable results).

This model will have, in addition, the property that it can be easily modified so as to investigate the effects of control of number of scientists. This effect will be expressed in terms of time needed up to the relative exhaustion of the problem.

d. *Parameters of the Model and Their Interpretation*

The model will depend on four parameters, denoted by N, a, b and c. The interpretation of these parameters will be the following:

N, as in model I, will denote the number of 'sub-problems' of the given problem. Parameter N may also be interpreted as the number of publications necessary to exhaust the subject.

a will denote the probability of 'death' of a given scientists working on the subject in a given year (here, of course, 'death' need not be interpreted literally; it simply signifies abandoning the research on the subject for whatever reasons).

b will denote the probability of obtaining a solution to a given subproblem by one scientist during one year of research (thus, it is implicitely assumed that all problems are equally difficult, and all scientists have equal ability to solve every subproblem). It will be also assumed, that each scientist working on a given subject tries, in each year, to solve all subproblems which are still not solved in this year. This is obviously a considerable simplification; the last assumption might presumably be replaced by a more realistic one, but it would require introducing one or more new parameters, and consequently, would require collecting other type of empirical data than those used in this paper.

c will denote the coefficient of attractiveness of the subject.

e. *Basic Variables of the Model are the Following:*

u_t – number of scientists working on the subject in year t;
x_t – number of publications on the subject which appear in year t.

f. *Basic Equations*

We shall now present equations which will serve as postulates of the model. Following each equation, the justification of it will be given. Such a justification will consist of deriving the expression for expected value of corresponding variable in year $t+1$ given the values of the variables in years preceding t. As in model of epidemic outlined above, the equation serving as a postulate will be obtained by replacing expected values by actual values.

The equation expressing the *number u_{t+1} of scientists working on the subject in year $t+1$* is the following:

$$(4) \qquad u_{t+1} = (1-a)\,u_t + cx_t\,.$$

The justification of Equation (4) is the following: in year $t+1$ the expected number of scientists working on the subject will be

u_t – the number of scientists working on the subject in year t,
minus
au_t – the expected number of scientists who stopped working on the subject,
plus
cx_t – the expected number of scientists who became attracted to the problem by reading papers which appeared[4] in year t. Equation (4) expresses the above balance.

The equation expressing the *number of publications in year $t+1$* is

$$(5) \qquad x_{t+1} = [1-(1-b)^{u_t}]\left(N - \sum_{i=1}^{t} x_i\right).$$

The justification of Equation (5) is the following: x_{t+1} equals to the number of papers which appear in year $t+1$, hence it equals to the number of subproblems which were solved in the preceding year, i.e. in year t (it is assumed, for simplicity, that the period of waiting for publishing of the paper is one year).

First, the probability that a given subproblem will be solved in year t by a given scientists equals b, hence the probability of the opposite event, i.e. of event that a given scientists will not solve a particular problem equals $1-b$. As there are u_t scientists working on the subject in year t, probability that a given subproblem will not be solved by any of them is

$$(1-b)^{u_t}.$$

Consequently, the probability that a given subproblem will be solved in year t (by any of the u_t scientists working on the subject) equals

$$1-(1-b)^{u_t}. \tag{6}$$

Next, in year t there remained

$$N - x_1 - x_2 - \cdots - x_t = N - \sum_{i=1}^{t} x_i \tag{7}$$

subproblems to be solved. Knowing the probability (6) of solving a given subproblem by at least one of the scientists, and the joint number of subproblems (7) which remain to be solved, the expected number of subproblems solved in year t is equal to the product of (6) and (7), which gives the right-hand side of equation (5). But, as already pointed out, x_{t+1} equals to the number of subproblems solved in the preceding year, i.e. in year t.

Equation (5) deserves some comments. This equation takes into account all subproblems which were solved up to year $t-1$ inclusive, hence whose publications appear in years $1, 2, \ldots, t$. Thus, it is implicitely assumed that scientists working in year t know which subproblems have already been solved. Part of this knowledge comes out of reading publications: results obtained in years $1, 2, \ldots, t-2$ are published in years before t; the knowledge about most recent results, i.e. those obtained in year $t-1$, which are published only sometime during the year t, is assumed to come from preprints, personal contacts, etc.

Incidentally, it is possible that a more realistic picture would be obtained, if one assumed that the unit of time is not one year, but two years (this being a closer approximation of waiting time for publication). Clearly, the interpretation of results obtained by means of the model would be unchanged, provided that the term 'tth year' will be interpreted as 'tth successive two-year period'.

g. *Auxiliary Equations*

The basic Equations (4) and (5) contain too many parameters (N, a, b and c) to be determined from the data on x_1, x_2,.... Thus, we shall use additional data, concerning

– number of papers written by new authors in year t;

– number of new names of authors in year t.

Clearly, number of new authors need not be equal to the number of

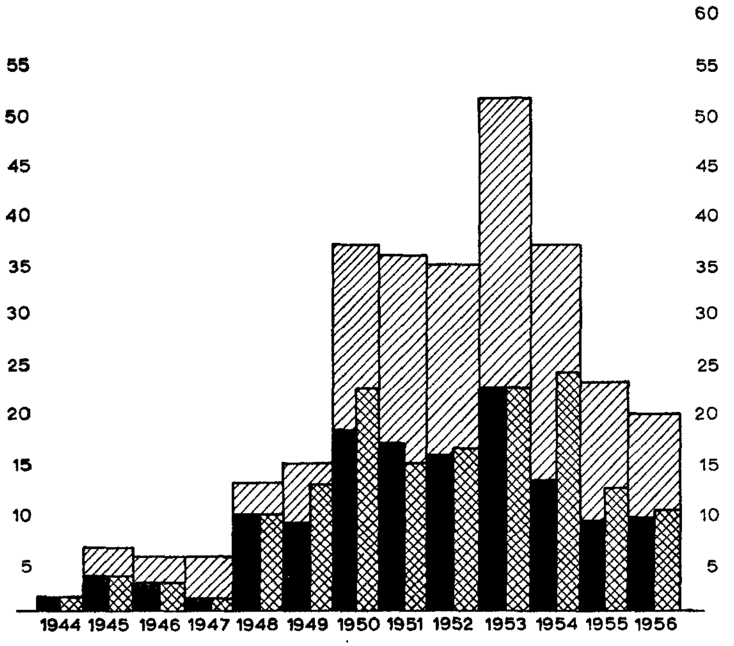

Fig. 7. Papers on theory of games Source: see Figure 2. *Shaded area*: numbers of papers in successive years (as in Figure 2); *Doubly shaded areas* (right hand side interior histogram): numbers of papers written by new authors in successive years; *Black area* (left hand side interior histogram): numbers of new names of authors in successive years (see explanation in text).

papers written by new authors (in year t), since some of them may publish more than one paper in the same year.

Such data for theory of games are presented on Figure 7. One can see there three histograms: the outer one is simply a fragment of histogram of Figure 2, comprising the data for 1944–57. The right of the interior histograms represents successive numbers of papers by new authors, and the left interior histogram represents the numbers of new names in publications in successive years.

The following principle of counting joint papers have been accepted in preparing this histogram: a paper which appears in year t written by two authors, say A and B, such that A has already published (in preceding years) papers on theory of games, and B has not published yet, is counted as $\frac{1}{2}$ of paper by 'new author' (B), while that author B is counted as new name. This explains why in some years the number of new names exceeds the number of papers written by new authors.

An analogous principle was accepted for papers written by three or more authors.

The following notations will be used in the model:

u_t^0 – the number of scientists working on the subject in year t who have not solved any of the subproblems (state at the beginning of year t);

z_t – number of papers by new authors which appear in year t;

n_t – number of new names in publications in year t.

For simplicity, it is assumed in the model, that there are no joint papers.

The equation for *number* u_{t+1}^0 *of scientists working on the problem in year* $t+1$ *who have not solved any of the subproblems up to the beginning of year* $t+1$ is

$$u_{t+1}^0 = (1-a)\left(u_t^0 - n_{t+1}\right) + cx_t. \tag{8}$$

The justification of Equation (8) is the following: at the beginning of year t there were u_t^0 scientists who have not solved any of the subproblems yet. Some among them succeeded in solving at least one of the subproblems – the number of such scientists equals to the number of new names which appear in publications in year $t+1$, i.e. it equals to n_{t+1} (as the publications in year $t+1$ concern the results obtained in year t).

Thus, at the end of year t, the number of scientists who still have not solved any of the subproblems is equal to

$$u_t^0 - n_{t+1}.$$

Out of this number, fraction a of it abandons the problem, while cx_t scientists become attracted by the problem. We have therefore the balance

$$u_{t+1}^0 = u_t^0 - n_{t+1} - a\left(u_t^0 - n_{t+1}\right) + cx_t$$

which gives Equation (8). The appearance of term cx_t on the right hand side is connected with the fact, that the 'new' scientists who become attracted by the problem have not solved any of the subproblems, by definition.

The equation expressing the *number* z_{t+1} *of papers by new authors in year* $t+1$ is

$$(9) \qquad z_{t+1} = x_{t+1}(u_t^0/u_t).$$

The justification of Equation (9) is the following: the joint number of papers which appear in year $t+1$ is x_{t+1}, and, as already pointed out, this number equals to the number of subproblems solved in year t. Now, there were u_t scientists working on the problem in year t, out of which u_t^0 have not solved any subproblem until the beginning of year t. Thus, the number z_{t+1}, equal to the number of subproblems solved in year t by 'new' scientists working in year t, will be equal to the product of x_{t+1} (=joint number of subproblems solved in year t), and the ratio u_t^0/u_t which expresses the proportion of 'new' scientists to the number of all scientists (working on the problem in year t). This gives Equation (9).

Finally, the equation expressing the *number* n_{t+1} *of names of new authors in publications in year* $t+1$ is

$$(10) \qquad n_{t+1} = [1 - (1 - 1/u_t^0)^{z_{t+1}}]\, u_t^0, \quad \text{for} \quad u_t^0 > 0.$$

The justification of this equation is the following: n_{t+1} equals to the number of scientists who satisfy the following two conditions

(a) in year t they solved at least one of the subproblems (so that their names appear in publications in year $t+1$),

(b) before year t they have not solved any of the subproblems, so that up to the year t inclusive they had no publication on the subject.

Thus, u_t^0 scientists have written z_{t+1} papers altogether (as z_{t+1} is the number of subproblems solved by 'new' scientists), and the problem reduces to computing the expected number of those individuals, among u_t^0 scientists, who have written at least one of the z_{t+1} papers.

Probability that a given paper is written by a given scientist (out of u_t^0 scientists) equals $1/u_t^0$, hence the probability that a given scientist will not publish a given paper, i.e. will not be the one with priority of solution to a given subproblem is

$$1-(1/u_t^0).$$

Consequently, probability that a given scientist will not publish any of the z_{t+1} papers (papers of new authors which appear in year $t+1$) equals

$$(1-1/u_t^0)^{z_{t+1}}.$$

Therefore the probability that a given scientist has published at least one of z_{t+1} papers (hence his name is represented in publications in year $t+1$) equals

$$(11)\qquad 1-(1-1/u_t^0)^{z_{t+1}}.$$

Consequently, the expected number of new names in year $t+1$ is the product of (11) by the number u_t^0 of new scientists working on the subject in year t, which justifies the Equation (10).

h. *Numerical Simulation*

To get an idea of the behaviour of curves described by model II, specific values of parameters were accepted, and the curves were drawn for some initial values. The data are presented in Table I, and on Figure 8. The following values of parameters were assumed:

$N=400$; the problem consisted therefore of 400 subproblems.

$a=0.3$; in each year 30% of scientists working on the subject abandon it;

$b=0.005$; the degree of difficulty of each of the subproblems is such that out of 1000 scientists working in the given year on the given subproblem, on the average only 5 will get the solution;

$c=1.3$; on the average, each new publication attracts 1.3 new scientists.

TABLE I

t	1	2	3	4	5	6	7	8	9	10
u_t	5	11.3	20.6	41.2	76.1	125.4	198.9	249.4	267.6	219.4
x_t	6	9.8	20.6	36.3	60.9	84.4	84.7	61.6	24.7	7.3
z_t	–	5.9	14.6	23.4	41.8	57.3	61.4	44.4	18.2	5.3
n_t	–	2.7	7.1	11.1	22.1	34.7	44.6	38.1	17.0	4.9
u_t^0	3	8.0	13.3	28.4	51.7	90.9	143.5	183.9	195.8	165.6
$N-\Sigma x_t$	394	384.2	363.6	327.3	266.4	182.0	97.3	35.7	11.0	3.7

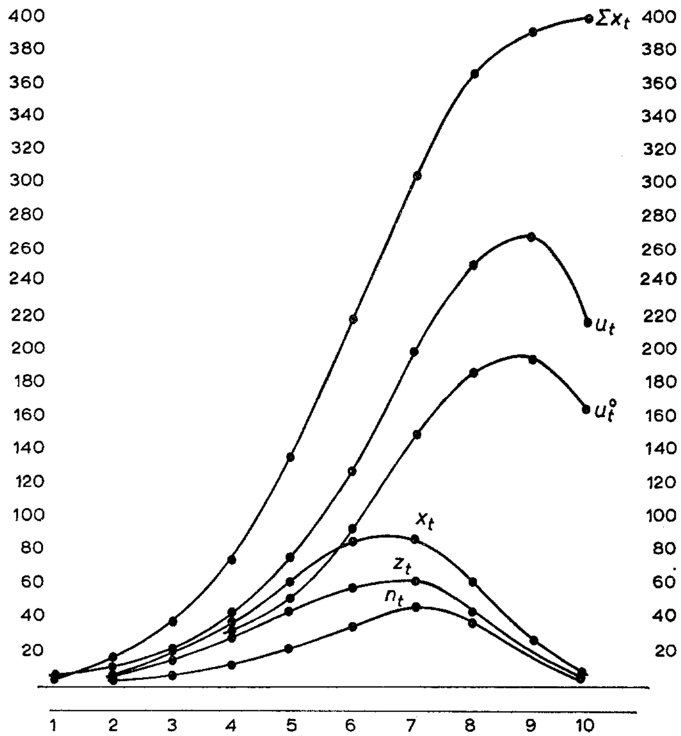

Fig. 8. Numerical illustration for model II Values of parameters: $N=400$, $a=0.3$, $b=0.005$, $c=1.3$. Initial values: $u_1=5$, $u_1{}^0=3$, $x_1=6$.

u_t – number of scientists working on the subject in year t

$u_t{}^0$ – number of scientists who work on the problem in year t, but at the beginning of this year have not solved yet any of the subproblems (obtained no publishable results)

x_t – number of publications on the subject in year t

z_t – number of publications of new authors in year t

n_t – number of new names of authors in publications in year t

Σx_i – the degree to which the problem has been exhausted in year t.

The initial values were assumed to be $u_1=5$, $u_1^0=3$, and $x_1=6$, hence it was assumed that in year 1 there appeared 6 papers, and five scientists were working on the problem, out of which three did not solve any of the subproblems as yet.

It may be seen from Figure 8, that after 10 years the problem has been almost completely exhausted (Σx_i reached almost the value 400). The maximal number of scientists (267) worked on the problem in year 9;

at the same time there was also the peak in number of scientists who have not succeeded in solving any of the subproblems (195).

Three curves, simulating the numbers of publications (x_t), numbers of publications by new authors (z_t) and numbers of new names of authors in publications in year t (n_t) reach their maxima in 7-th year, equal respectively to 84, 61 and 44. It is interesting that the differences between curves z_t and n_t are large at the beginning (up to 6-th year), and then decrease rather fast. This is explained by the fact that at the beginning of the process the number of papers per one scientist is larger than in the later periods, when the subject is becoming exhausted, and the number of scientists still continues to increase. Thus, at the beginning of the process there is a larger number of scientists who write more than one paper in one year – this gives the difference between z_t and n_t.

The fact that the curve z_t lies always above the curve n_t is connected with the assumption of the model which excludes the joint papers.

Because of computational difficulties, no attempts have been made to find values of parameters for which the curves x_t, z_t and n_t would give best fit to empirical histograms (in particular those of theory of games, presented in Figure 7). This will be done with the help of electronic computers.

i. *Problems of Control of the Process*

Model II can be easily modified so as to investigate the influence of controlled changes on numbers of scientists working on the problem. Such a control is described in the model simply in form of adding to values u_{t_0} and $u_{t_0}^0$ (for a fixed t_0) a certain number d of scientists. Then in determining the values of variables for the next period of time, i.e. for value t_0+1 one should use $u_{t_0}+d$ and $u_{t_0}^0+d$ in place of u_{t_0} and $u_{t_0}^0$ in corresponding formulae.

TABLE II

t	1	2	3	4	5	6	7	8	9
u_t	5	11.3	20.6	41.2	97.1	216.5	312.9	358.9	
				+30	+30	+30	+30		
x_t	6	9.8	20.6	36.3	98.1	108.0	86.0	28.8	5.2
z_t	–	5.9	14.6	23.4	80.3	74.4	64.1	21.9	4.1
n_t	–	2.7	7.1	11.1	43.7	50.1	53.8	20.1	4.0
u_t^0	3	8.0	13.3	28.3	57.6	153.8	231.4	280.1	
				+30	+30	+30	+30		
$N-\Sigma x_i$	394	384.2	363.6	327.3	229.2	121.2	35.2	6.4	1.2

Clearly, such a control by adding (or subtracting, if $d<0$) of scientists need not be restricted to one time moment.

Assuming a given process of control, one can analyse the behaviour of curves, and obtain answers to questions concerning the changes in the dynamics of the process, and changes in the time up to the exhaustion of the problem.

Table II below gives the results for such a control in the case when to

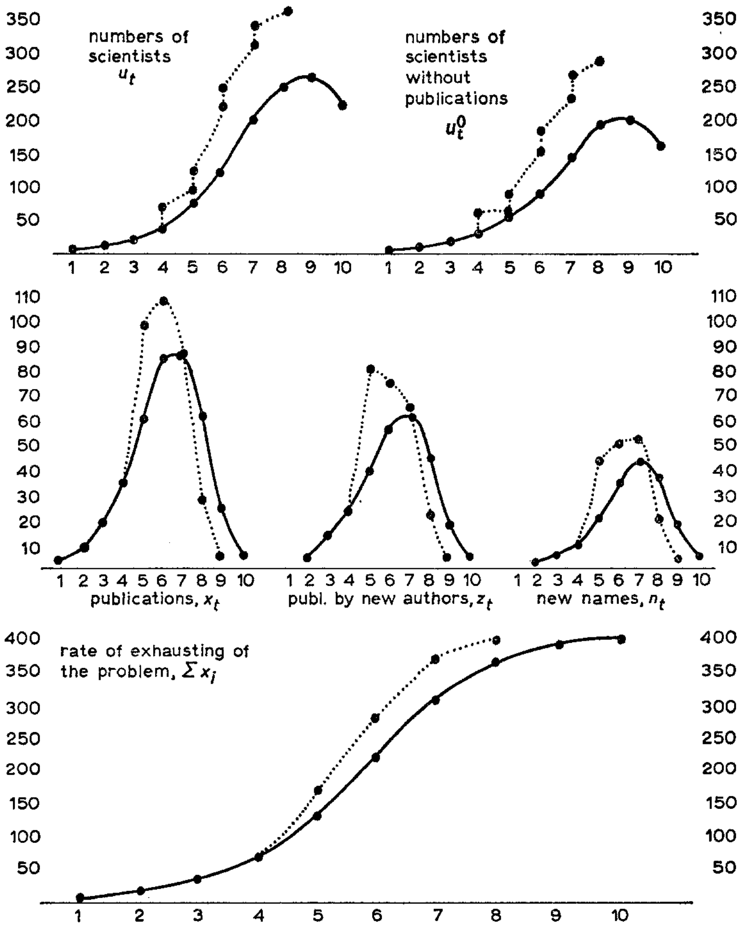

Fig. 9. Numerical illustration of control of process described by model II Continuous curves: process from Figure 8. Dotted curves: process with parameters as in Figure 8, to which 30 new scientists were added in each of the moments 4, 5, 6 and 7.

the process with parameters from the preceding example 30 new scientists are added in each of the years 4, 5, 6 and 7. Continuous curves on Figure 9 present the results for uncontrolled process (that from Figure 8), and dotted curves represent the results for the controlled process.

As can be seen, maximal effects of control occur in years 6 and 7, when the differences in the degree of exhausting of the problem reach 60 papers, which is about 15% of all subproblems. Besides, such a control caused the exhaustion of the problem in about two years earlier than it would have occurred for the uncontrolled process.

The control has also caused changes in the dynamics of the number of scientists working on the problem: the appearance of a larger than normal number of publications in year 5 (after adding first 30 scientists) caused 'attracting' to the problem of a larger number of scientists than it would have occurred otherwise (for the uncontrolled process). These changes, in turn, caused the changes in curves z_t and n_t.

3. Conclusions

The approach suggested allows us to build predictions for spreading out of scientific objects, such as theories, hypotheses, methods, etc.

The basis for each prediction lies always in the correct definition of mechanisms underlying the predicted phenomenon (expressed by an appropriate model), and in the exact estimation of parameters of the model and those variables which serve as a foundation of prediction.

It is clear, therefore, that in case of correct definition of these mechanisms, one obtains not only the possibility of prediction, but also the possibility of control of the considered process.

In case of models suggested in this paper, the premises which suggest the possibility of using them in the problem of control of science are the following:

If model II would give valid predictions (which can be reasonably expected on the ground that even a simplified model I gives a good specific prediction), we would have a premise for conclusion suggesting that its assumptions (expressed by equations) are adequate.

In turn, if these equations constitute an adequate[5] description of the mechanisms of the underlying phenomenon, then the knowledge of these equations, coupled with the knowledge of parameters appearing in them will give a basis for control. More precisely, one should be able to answer the following questions:

(1) to what extent the time needed for exhausting the subject will shorten if at particular moments of time given numbers of scientists will be added to those already working on the subject;

(2) how many scientists, and at which moment of time, ought to be added, so as to exhaust the problem not later than at the given year?

Clearly, the answers to questions (1) and (2) is more specific than the obvious recommendation (for which no reference to any model is needed) that 'the more scientists, the faster the problems will be exhausted'.

The models introduced in this paper require (when one wants to use them for actual predictions, or control) the information about the values of parameters for various problems. The material thus collected, besides immediate application for prediction, might be of the following use:

The characterization of each scientific problem in terms of a set of parameters (say, N, a, b and c, if model II is to be used), makes it possible to introduce a taxonomy consisting of grouping these problems which have similar values of parameters.

The empirical material concerning values of parameters for various problems might, therefore, be classified according to the introduced taxonomical categories. It is possible that such a classification would lead to generalizations, in form of relations between qualitative characteristics of problems, and their taxonomical categories. In other words, it is possible that the qualitative characteristics of the problem would determine its taxonomical category, hence values of its parameters.

If this turned out to be the case, one could build predictions of growth of a given problem even before a sufficient number of papers appeared; in an obvious way, this would make predictions easier and more available.

Clearly, one can consider the stochastic versions of the models considered, obtaining predictions not in terms of expectations, but in terms of confidence intervals for predicted values. This would allow us to evaluate probable magnitudes of fluctuations around the point predictions used in this paper.

Besides that, using the intuitions concerning epidemical spread of scientific objects, one could try to build other models, based on more extensive empirical data. Thus, it would be interesting to study the growth of a problem not in isolation, but against the background of related problems. In connection with this, one can study the 'diffusion' of scientists (or methods) from one branch to another. As regards scientists, one could attempt to predict the magnitude of flow of them in and out of dominating problems.

Another possibility would be to consider models taking into account geographical and-or economical aspects of the process.

Quite independently from solutions based on models, some interesting meta-scientific information may be obtained directly from histograms for particular problems.

Thus, for instance, the histogram 7 for theory of games shows that when the problem is approaching the exhaustion (and becomes more difficult), the number of joint papers increases (apparently it becomes more efficient to work in teams). This may be judged from the fact that numbers of new names exceed numbers of papers written by new authors.

Besides, histograms for theory of games (Figure 2), theory of measurement (Figure 3) and multivariate analysis in psychology (Figure 5) seem to suggest the following hypothesis: the papers or books which give a synthesis, or critical appraisal, of a given problem appear several years after reaching the maximum by number of publications.

One can imagine that the ideas contained in those papers or books which give critical synthesis of specific problems become the elements of the objective universe of science, as understood by Popper [37].

4. Summary

In this paper meta-science is understood as the theory of optimal control of development of science. The problem of prediction of spreading out of scientific objects, such as theories, hypotheses, methods, etc. is considered. The analysis in this paper is concerned with the process of changes of numbers of publications dealing with a given subject in successive years; this process is treated as an 'epidemic' in which new infections correspond to the appearance of new publications, resulting from 'infection by a scientific idea'.

The data presented concern, among others, the Wechsler tests, theory of games and theory of measurement. Two models are suggested (based on models of epidemics by Kendall and Daley), for analysis, prediction and control of the process of development of a given scientific subject.

It is assumed that the scientific problems analysed in science are characterized by a certain number N of 'sub-problems' such that solving each of them exhausts the subject.

The simpler of the two models (model I) deals only with prediction of number of publications concerning a given subject in successive years. This model has been verified on the basis of data for Wechsler tests.

Model II allows, using richer empirical data (among others, numbers of names of new authors in publications in successive years) not only the prediction of number of publications, but also the inference on variables which are difficult to observe, or whose direct measurement would be subject to heavy errors (such as, for instance, the numbers of scientists who work on the given subject, but have not solved any of the sub-problems). Moreover, this model supplies the theoretical data for control of the process of development of the given scientific object.

Model II allows us to answer the following question: what is the least number of scientists who ought to be added to the group already working on the subject (and in which moments of time they ought to be added) so that the subject will be exhausted not later than on a given year?

In the conclusions further possibilities of empirical approach to metascience are discussed.

Polish Academy of Sciences
Institute of Praxiology, Warszawa

BIBLIOGRAPHY

[1] Bellman, R., 'On the Theory of Dynamic Programming', *Proc. of the National Acad. of Sciences U.S.A.* **38** (1952).
[2] Blackwell, D. and Girshick, M. A., *Theory of Games and Statistical Decisions*, Wiley, New York, 1954.
[3] Bridgman, P. W., *Dimensional Analysis*, Yale Univ. Press, New Haven, 1931.
[4] Campbell, N. R., *Physics: the Elements*, Cambridge Univ. Press, Cambridge, 1920.
[5] Campbell, N. R., *An Account of the Principles of Measurement*, Longmans and Green, London, 1928.
[6] Cattell, R. B., *Description and Measurement of Personality*, World Book Co., Yonkers-on-Hudson, 1946.
[7] Cattell, R. B., *Personality: A Systematic Theoretical and Factual Study*, McGraw-Hill, New York, 1950.
[8] Cattell, R. B., *Factor Analysis*, Harper, New York, 1952.
[9] Cattell, R. B., *Personality and Motivation: Structure and Measurement*, World Book Co., Yonkers-on-Hudson, 1957.

[10] Cattell, R. B., (ed.), *Handbook of Multivariate Experimental Psychology*, Rand McNally, Chicago, 1966.
[11] Cattell, R. B. and Scheier, I. H., *The Meaning and Measurement of Neuroticism and Anxiety*, Ronald Press, New York, 1961.
[12] Daley, D. J., 'Concerning the Spread of News in a Population of Individuals who Never Forget', *Bull. of Mathematical Biophysics* **29** (1967).
[13] Ellis, B., *Basic Concepts of Measurement*, Cambridge Univ. Press, Cambridge, 1966.
[14] Eysenck, H. J., *Dimensions of Personality*, Kegan Paul, London, 1947.
[15] Eysenck, H. J., *The Scientific Study of Personality*, Kegan Paul, London, 1952.
[16] Eysenck, H. J., *The Structure of Human Personality*, Methuen, London, 1953.
[17] Eysenck, H. J., *Experiments in Personality*, Routledge and Kegan Paul, London, 1960.
[18] Eysenck, H. J. (ed.), *Experiments in Motivation*, Pergamon Press, London, 1964.
[19] Feller, W., *An Introduction to Probability Theory and Its Applications*, Volume I, 2nd ed., Wiley, New York, 1957.
[20] Gani, J., 'On Partial Differential Equation of Epidemic Theory', *Biometrics* **52** (1965) 52.
[21] Guilford, J. P., *Psychometric Methods*, McGraw-Hill, New York, 1954.
[22] Guilford, J. P., *Personality*, McGraw-Hill, New York, 1959.
[23] Gulliksen, H., *Theory of Mental Tests*, Wiley, New York, 1950.
[24] Hoskovec, J. and Kanka, Z., *Wechslerovy zkousky W-B, WAIS, WISC Bibliografia* [Bibliography of Wechsler Tests W-B, WAIS, WISC], Ceskoslovenska Psychologicka Spolecnost, Prague, 1961.
[25] Hölder, O., 'Die Axiome der Quantität und die Lehre vom Mass', *Berichte Über die Verhandlungen der königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig* (Math.-Phys. Klasse) **53** (1901).
[26] Kendall, D. G., 'Deterministic and Stochastic Epidemics in Closed Populations', *Proc. Third Berkeley Symposium on Probability and Statistics*, Berkeley 1955.
[27] Lehmann, E. L., *Testing Statistical Hypotheses*, Wiley, New York, 1959.
[28] Lord, F. M. and Novick, M. R., *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, Mass., 1968.
[29] Luce, R. D., 'A Definition of Stability for *n*-Person Games', *Annals of Mathematics* **59** (1954).
[30] Luce, R. D., 'Semiorders and a Theory of Utility Discrimination', *Econometrica* **24** (1956). (Further references to papers by Luce mentioned in the text can be found in [31] and [32].)
[31] Luce, R. D., Bush, R. R., and Galanter, E. (eds.), *Handbook of Mathematical Psychology*, Wiley, New York, 1963.
[32] Luce, R. D. and Raiffa, H., *Games and Decisions; Introduction and Critical Survey*, Wiley, New York, 1958.
[33] Nash, J. F., 'Equilibrium Points in *n*-Person Games', *Proc. National Acad. of Sciences USA* **36** (1950). (Further references to papers by Nash mentioned in the text can be found in [32].)
[34] Neyman, J. and Pearson, E. S., 'On the Testing of Statistical Hypothesis in Relation to Probability *a priori*', *Proc. Cambridge Philosophical Society* **29** (1933). (Further references to papers by Neyman and Pearson mentioned in the text may be found in [27].)
[35] Neyman, J. and Scott, E. L., 'A Stochastic Model of Epidemic', in *Stochastic Models in Medicine and Biology* (ed. by J. Gurland), Wisconsin, 1964.

[36] Nowakowska, M., *Język motywacji: propozycja z zakresu metodologii naukoznawstwa* [Language of Motivation: A Suggestion in the Methodology of Meta-Science], *Prakseologia* **37** (1971).
[37] Popper, K., 'Epistemology Without a Knowing Subject', in *Logic, Methodology and Philosophy of Science*, Vol III (ed. by B. van Rootselaar *et al.*), North Holland Publ. Co., Amsterdam, 1968.
[38] Scott, D. and Suppes, P., 'Foundational Aspects of Theories of Measurement', *Journal of Symbolic Logic* **23** (1958).
[39] Shapley, L. S., (references to papers by Shapley mentioned in the text may be found in [32]).
[40] Spearman, C., *The Nature of 'Intelligence' and the Principles of Cognition*, Macmillan, London, 1923.
[41] Spearman, C., *The Abilities of Man*, Macmillan, London, 1927.
[42] Spearman, C., *Psychology down the Ages*, Macmillan, London, 1937.
[43] Steinhaus, H., 'Definicje dla teorii grv i pościgu' [Definitions for the Theory of Games and Pursuit], *Myśl Akademicka* **1** (1925) (available in English translation in *Naval Research Logistic Quarterly* **7** (1960)).
[44] Stevens, S. S., 'On the Theory of Scales of Measurement', *Science* **103** (1946). (References to papers by Stevens mentioned in the text can be found in [13] and [46].)
[45] Strzałecki, A., *Wybrane zagadnienia psychologii twórczości*, [Selected Topics in Psychology of Creativity], PWN, Warszawa, 1969.
[46] Suppes, P. and Zinnes, J. L., 'Basic Measurement Theory', in *Handbook of Mathematical Psychology*, Vol I, (ed. by R. D. Luce, R. R. Bush and E. Galanter), Wiley, New York, 1963.
[47] Thorndike, R. L., *Personnel Selection*, Wiley, New York, 1949.
[48] Thurstone, L. L., *A Factorial Study of Perception*, Chicago Univ. Press, Chicago, 1944.
[49] Thurstone, L. L., *Multiple Factor Analysis*, Chicago Univ. Press, Chicago, 1947.
[50] Thurstone, L. L. and Chave, E. J., *The Measurement of Attitudes*, Chicago Univ. Press, Chicago, 1929.
[51] von Neumann, J., 'Zur Theorie der Gesellschaftsspiele', *Mathematische Annalen* **100** (1928).
[52] von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behaviour*, 1st ed., Princeton 1944; 2nd ed., Princeton 1947.
[53] Wald, A., *Statistical Decision Functions*, Wiley, New York, 1950. (Further references to papers by Wald mentioned in the text may be found in [27] and [32].)
[54] Weiss, G. H., 'On the Spread of Epidemic by Carriers', *Biometrics* **21** (1965).

NOTES

[1] I.e. the total number of those who get infected up to a given moment.
[2] The following general principle regarding bibliography is accepted in this paper: books or articles are referred to by their numbers in the list of bibliography at the end of this paper. Occasionally the title is presented in full, whenever it conveys enough information relevant to the topic actually discussed. When a contribution of a given scientist is discussed or mentioned, and it is difficult or impossible to pinpoint one particular book or paper by this author as being most representative, no number is

given in the text at all. The name of this scientist appears, however, in the bibliography at the end of this paper, and references are made to other bibliographies which contain information about his publications.

[3] Thus, it is assumed that individuals removed from epidemic do not take part in it any more; of course, one can consider other versions of this model, without the above restriction (see [54]).

[4] Of course, one can also consider model in which the number of 'attracted' scientists is proportional not to x_t, but to, say, $\sum_{i=1}^{i} x_i$, which gives the joint number of papers which appeared up to year t.

[5] It is implicitely assumed in these equations that the probability of solution of a given subproblem in the given year depends only on the number of scientists working on this subproblem in this year. Thus, as no technical basis for research is taken into account, model II is more likely to be applicable to theoretical branches of science than to those requiring specialized and expensive instruments for research. One could, however, construct a more elaborate model, in which expansion of technical basis of research would be taken into account.