(6)
PSD

GENERALIZATION OF EPIDEMIC THEORY

AN APPLICATION TO THE TRANSMISSION OF IDEAS

By Dr. WILLIAM GOFFMAN
Center of Documentation and Communication Research, School of Library Science, Western Reserve University

AND

Dr. VAUN A. NEWILL
School of Medicine, Western Reserve University, Cleveland, Ohio

ONE of the most fundamental problems in the field of information retrieval is that of determining the circumstances under which it might be necessary to introduce an information retrieval system as an aid to a given population of scientists. It is proposed that this problem be examined in terms of the transmission and development of ideas within a population. Specifically, the transmission of ideas within a population will be treated as if it were the transmission of an infectious disease, that is, in terms of an epidemic process. An attempt will be made to indicate the role of information retrieval in the development of such a process.

The Epidemic Model

Since the spread of disease in a population is to be our model for the transmission of ideas, it is appropriate to discuss the essential principles pertinent to this issue. These principles are a part of epidemiology. The necessary elements involved in the process of the spread of an infectious disease are those of : (1) a specified population ; (2) an exposure to infectious material. The members of the population belong to 1 of 3 basic classes at a given instant in time : (a) *infectives*, those members who are host to the infectious material ; (b) *susceptibles*, those who can become infectives given contact with infectious material ; (c) *removals*, those who have been removed for one of a variety of reasons such as death, immunity, hospitalization, etc. These latter members may have been either susceptibles or infectives at the time of their removal.

The entire process is time-dependent, that is, a sequence of events occurs which describes the process. An individual is exposed to infectious material either by direct contact with an infective or through some intermediary host or vector. The exposed person may either be resistant to the incoming organism, in which case the organism is rejected, or may be infected by it, in which case the invading organism proceeds on a course of development. The time interval in which the development takes place is called the 'latency period' (or incubation period), that is, the interval of time necessary for a susceptible to be transformed into an infective. In other words, the latency period is the time between the receipt of infectious material by a susceptible and the time at which he is in position to transmit infectious material to another susceptible, thus repeating the process. An epidemic occurs when the process of susceptibles being transformed to infectives crosses a certain threshold. (The American Public Health Association (1960) defined an epidemic as "the occurrence in a community . . . of a group of illnesses of similar nature, clearly in excess of normal expectation.")

The process already described here does not take into account the almost endless number of complexities which actually arise. It is a model, that is, an idealization of the real situation in which the complex process is reduced to its essential properties.

Transmission of Ideas as an 'Epidemic' Process

In general, the 'epidemic' process can be characterized as one of transition from one state (susceptible) to another (infective) where the transition is caused by exposure to some phenomenon (infectious material). The process need not be restricted to infectious disease but is a more general abstract process that might be applied to many situations. All that is needed is the appropriate interpretation of the process elements, that is, susceptibles, infectives, removals, infectious material, intermediary host, latency period, disease, etc.

People are susceptible to certain ideas and resistant to others. Once an individual is infected with an idea he may in turn, after some period of time, transmit it to others. Such a process can result in an intellectual 'epidemic' (Table 1). For example, consider the development of psychoanalysis in the early part of this century. Freud was no less host to the infectious material of the 'disease' of psychoanalysis than the person carrying the organism capable of transmitting a cold, nor is his writing less of a 'vector' carrying the 'infectious material' than the mosquito as a carrier of malaria.

Moreover, Abraham, Ferenczi, Jung and Jones were no less 'susceptibles' who were infected by the ideas of Freud and who, after a certain latency period, themselves became 'infectives' than are those individuals infected by the droplets expressed by a cold carrier. Jung might represent an example of acquired resistance to the disease while the resistance of the medical community of Vienna could represent innate immunity. The development of the psychoanalytic movement in the early part of the twentieth century was in its way no less an 'epidemic' than was the outbreak of influenza in 1917 and 1918.

One can argue similarly that Darwin and evolution, Cantor and set theory, Newton and mechanics, and so on, were examples of 'epidemics' in the world of scientific thought which were instigated by the introduction of a single infective into a population. The analogy is not restricted to science; for examples such as Christ, Buddha, Moses and Mohammed can be cited in the

Table 1. ANALOGY BETWEEN INFECTIOUS DISEASE AND INTELLECTUAL EPIDEMICS

Elements of the epidemic process	Elements interpreted in terms of: Infectious disease epidemic	Intellectual epidemic
Host		
Agent	Infectious material	Idea
Infective	Case of disease	Author of paper
Susceptible	Person who will be infected given effective contact	Reader of paper who will be infected given effective contact
Removal	Death or immunity	Death or loss of interest
Vector		
Agent	Infectious material (as for host)	Idea (as for host)
Infective	Vector harbouring the agent	Paper containing useful ideas
Susceptible	Vector not harbouring the agent	All papers containing potentially useful ideas
Removal	Death	Deletion or loss

religious field and other examples can be given almost endlessly.

Although biological and intellectual epidemics can be considered as special cases of a general process, there are important differences. Intellectual epidemics are often desirable while biological ones usually are not. This difference, however, does not pertain to the structure of the processes but only to external factors which might be introduced. In one case, one usually wishes to enhance the process, while in the other to eliminate it.

Another important difference relates to the infectious material. In the biological case the infective individual produces infectious material that closely resembles that which started the process ; there has been little mutation or change. Such is not the case with the intellectual epidemic, for here some mutation or change of the original idea (infectious material) is prerequisite to placing it in print.

Ideas are transmitted by personal communication or by publication. An epidemic cannot develop within a given population unless there is effective contact between the susceptibles and infectious material. The purpose of an information retrieval system is to provide such effective contact where it does not already exist. If such is the case, an information retrieval system might be introduced even though such systems are limited to the transmission of ideas contained in published material.

The essential question is : when is it desirable to introduce a formal information retrieval system into the intellectual activity of a population of scientists? Here the epidemic model may be of value for it permits the rephrasing of the question in terms of epidemiological problems such as : (1) the transmission and spread of the disease ; (2) the prediction of the epidemic course ; (3) the discovery of the threshold densities of the population that might be passed before an epidemic can develop. The answer to these epidemiological questions might help to decide when an information retrieval system should be introduced.

Mathematical Models

There has been extensive development of abstract mathematical models in the theory of epidemics which can be put to general use. There are two types of models, namely, deterministic and stochastic. The deterministic approach represents the process as a system of differential equations while the stochastic approach describes the process as a finite state Markov process of either a discrete or continuous parameter (the discrete parameter is applicable where the latency period is constant, in which case the infectives occur in generations) depending on the physical situation. The stochastic representation, though generally more realistic, is mathematically more sophisticated. An excellent detailed treatment of this subject as related to infectious disease can be found in Bailey[1].

A Deterministic Model

The most natural avenue for exploring the process of the transmission of ideas within a population would seem to be by way of the literature produced by the members of that population. Although the published article is not the only way in which ideas can be transmitted, it does play an important part. We are interested in studying this role, for it is only in this respect that an information retrieval system can be justified at present. In terms of the epidemic process we are therefore concerned with the transference of infectious material, that is, ideas, between human hosts by means of an intermediary host or vector, namely, the written article.

We wish to study the spread of the 'disease' D within the population N of the discipline F. Consider a certain field (F) such as medicine and a certain subfield (D of F) such as medical epidemiology. The field (F) need not be as broad as medicine, but might be epidemiology, and mathematical epidemiology might be the subfield (D). N must be specified because D might develop into an 'epidemic' in some population N_1 but not in some other N_2. For example, the 'disease' of mathematical epidemiology will probably follow different courses in the fields of epidemiology, where it might become an 'epidemic', and in the field of mathematics, where it probably will never become an 'epidemic'. An additional example from the field of medicine might be the present epidemic of cancer of the lung among smokers that is certainly different from the pattern of development of the disease among non-smokers.

Let N_0 be the set of authors who have written papers in F at time t_0, where t_0 represents the starting time of the process. Let I_0 be the set of authors who have written papers in D at time t_0. Then I_0 represents the infective population at the beginning of the process. If it is assumed that there are no removals at time t_0, then $S_0 = N_0 - I_0$ represents the susceptible population at time t_0. (Removals constitute those members of N who for some reason or other are no longer producing papers in F.)

Let I'_0 be the set of articles produced in D by the members of I_0 at time t_0 and S'_0 be the set of articles produced by the members of S_0 at t_0. Then $N'_0 = S'_0 + I'_0$ constitutes the vector population at time t_0.

In general as the process develops we have $N = S + I + R$ and $N' = S' + I' + R'$, where R and R' represent removals. At time t_0, $R_0 = 0$ and $R'_0 = 0$.

The process might be described as follows : A member s of S is infected by a member of I'. After a certain length of time (latency period) s becomes an infective and gives birth to new infections in N' which in turn may infect other members of S and so forth. In other words s produces a paper in D in which he cites a paper in I'. (The member of I' might contain more than one idea (more than one type of 'infectious material') with respect to D. This same member of I' may contain ideas that could also infect members of a sub-field other than D. Our only interest in the member of I' is as a carrier of the 'infectious material' of D.) By producing this paper he has introduced a new infective in N'. He may also have cited papers in S' as well, so that he has infected or associated members of S' with the 'disease' D. The appearance of the paper written by s in D indicates that he has made the transition from the susceptible to the infective state.

Let β be the rate at which the members of S become members of I. Hence, β is the rate of infection in the population N. Let β' be the rate of infection in the population N'. Then β' represents the rate at which papers in S' are being cited by members of N who are producing papers in D.

As the process develops with time, a certain number of susceptibles and infectives in the host and vector populations are removed. In the case of the host population this removal might be loss of interest, death, etc. In the case of the vector population it might be lack of citation due to deletion, inaccessibility, etc. Let γ (γ') be the rate of removal of infectives I (I') and let δ (δ') be the rate of removal of susceptibles S (S').

Another aspect of the process as it develops is that new supplies of susceptibles and infectives are introduced into the populations N and N'. Let μ (μ') be the rate at which new susceptibles are introduced and let υ (υ') be the rate at which new infectives are introduced. That is, μ represents the rate at which new authors are introduced in F but not in D, while μ' is the rate at which new papers are introduced in F but not in D. υ represents the rate at which new authors are introduced in D and υ' the rate at which new papers are introduced into D.

The mathematical model to be presented includes the following states of transition of the host and vector populations within a short time interval Δt : the susceptible can remain a susceptible or become an infective

Table 2. STATES OF TRANSITION OF THE HOST AND VECTOR POPULATIONS*

State to which an individual belongs	
At time t	At time $t+\Delta t$
(1) Susceptible	Susceptible, infective or removal
(2) Infective	Infective or removal
(3) Removal	Removal

* In general the host and vector populations need not follow the same transitional patterns as in this case.

or removal, the infective can remain an infective or become a removal, and the removal must remain a removal (Table 2). (To keep the mathematical model simple in this presentation, we are not permitting immunes to lose their immunity and to return to the susceptible population.) The latency period is assumed to be zero so that infectives occur continuously with time. If we assume homogeneous mixing among the members of the populations, the total process can be deterministically described by the following systems of differential equations:

$$\frac{dS}{dt} = -\beta SI' - \delta S + \mu \quad : \quad \frac{dS'}{dt} = -\beta' S' I - \delta S' + \mu'$$

$$\frac{dI}{dt} = \beta SI' - \gamma I + \upsilon \quad : \quad \frac{dI'}{dt} = \beta' S' I - \gamma' I' + \upsilon'$$

$$\frac{dR}{dt} = \gamma I + \delta S \quad : \quad \frac{dR'}{dt} = \gamma' I' + \delta' S'$$

A threshold density of susceptibles can be obtained if we assume initial conditions of a single infective introduced into the host (N_0) and vector (N'_0) populations at time t_0, that is, $N_0 = (S_0, 1, 0)$ and $N'_0 = (S'_0, 1, 0)$. For an epidemic to develop from time t_0, the change in the number of infectives in both the host and vector populations must be positive. Hence, from the above differential equations, we obtain the threshold $\rho\rho'$ as follows:

$$\beta SI' > \gamma I - \upsilon$$

$$\beta' S' I' > \gamma' I' - \upsilon'$$

$$\text{or } \beta\beta' S_0 S'_0 > \gamma\gamma' - \gamma\upsilon' - \gamma\upsilon + \upsilon\upsilon'$$

$$\text{and } \quad S_0 S'_0 > \left(\frac{(\gamma-\upsilon)}{\beta}\right)\left(\frac{(\gamma'-\upsilon')}{\beta'}\right) = \rho\rho'$$

$$\text{where } \quad \frac{\gamma-\upsilon}{\beta} = \rho \; ; \text{ and } \frac{\gamma'-\upsilon'}{\beta'} = \rho'$$

The epidemic curve which traces the development of the epidemic in time is given by the differential equation $\frac{dI}{dt} = F(t)$ with peaks at the points in time where $\frac{d^2I}{dt^2} = 0$. The size of the epidemic is given by the limit of I as t goes to infinity, that is, the size of the infective population I after a long period of time. This discussion indicates that an epidemic can only develop from time t_0 provided that the product of the host and vector susceptible population exceeds the threshold $\rho\rho'$. When a potential epidemic condition is present, the quantity of scientists and information is sufficiently large that given effective contact an epidemic can develop. It is at this time that an information retrieval system might be introduced.

If the population is already in an epidemic state, an information retrieval system can be introduced in order to maintain or increase the frequency of effective contacts between susceptibles and infectives. Hence, the points at which an information retrieval system might be usefully introduced into a population are either at the initial point or the peak of the epidemic process or at a time when intensification of the epidemic process is desired.

Other models different from the one we have presented could be constructed that would interpret the epidemic in a different manner. For example, a model can be elaborated by introducing two interacting populations with rates of migration from one population to the other. One population might consist of biostatisticians and the other of epidemiologists. The 'disease' could be statistical epidemiology. Then one might study the development of the field in terms of the interacting host and vector populations rather than in terms of single populations.

The Stochastic Approach

In a stochastic model of the process described above the actual number of new infectives occurring in a short time-interval would be replaced with the probability of a new case occurring in that interval. This type of mathematical treatment is particularly applicable when dealing with small populations.

An epidemic process within a large population can be characterized as a collection of epidemic processes within sub-populations. For example, the outbreak of an infectious disease in a community might more specifically be studied in terms of the household populations rather than the community as a whole. Similarly the intellectual epidemic process within the population of a certain discipline could be studied in terms of its sub-populations, that is, the 'disease' D of field F might be studied in terms of the sub-fields of F, indicating, from the point of view of information retrieval, where intellectual epidemics are developing in a given discipline.

Thus, a stochastic treatment of the epidemic processes would in general be more desirable because the size of the populations being studied has been reduced. Unfortunately a stochastic model of an epidemic process with a vector population does not exist as far as we know. Such a model is in need of development and attention is being given to this problem by the authors.

Deterministic models, however, can be very useful, particularly in dealing with large populations, and experiments have been designed for the purpose of testing the model presented above.

Discussion

The construction of a mathematical model of a complex physical situation involves a certain amount of simplification. This is because the purpose of the mathematical model is the description and prediction of the essential patterns of the physical process and not the achievement of its complete analysis.

A total understanding of the process requires studies complementary to the mathematical ones. In the case of the infectious diseases this means biological, ecological, and clinical investigations of the process as well as mathematical; in the transmission of ideas, psychological, sociological, and linguistic studies.

Mathematics cannot indicate why exposure to a certain virus will lead to infection or why a certain article will infect the reader with an idea. These important questions must be answered elsewhere. However, mathematics can make important contributions toward an understanding of the epidemic process if one keeps its limitations in mind.

Of course, the usefulness of a mathematical model depends on: (1) being able to solve the model; (2) obtaining reliable measurements in order to test it. These are by no means negligible problems, since even the most simple representations present difficulties in obtaining algebraic solutions. This is particularly true of the stochastic case. Additional problems are encountered in relation to defining populations and obtaining reliable estimates of the parameters of the process.

Although implementation of the epidemic model presents certain difficulties, much useful information concerning the transmission of ideas within a population might be obtained from the mathematical approach. It could help in answering several questions that are basic for the design and operation of information retrieval systems. These questions include: (1) Where and when is activity within a given discipline developing into 'epidemic' proportions? (2) What is the expected duration of this 'epidemic' activity? (3) What is the intensity of the 'epidemic'? (4) What are the classic papers of a certain discipline, that is, the initial infectives of an outbreak?

Information retrieval systems must be dynamic and should reflect the interactions between the researcher and the literature. It seems that the 'epidemic' approach is a workable method for describing and predicting certain fundamental aspects of this interaction.

Since the process of transmitting ideas, as in the case of an infectious disease, is not a single process within a population but a collection of interacting processes within sub-populations, it would seem that the notion of an all-encompassing information retrieval system spanning the totality of knowledge should be replaced by the notion of small dynamic interrelated systems that appear when needed and disappear when not needed.

'Epidemic' theory could have a wide range of applications other than to the transmission of ideas, for example, to the study of certain chronic diseases, divorces, accidents, etc., but these will not be elaborated on in this report.

This work was supported by AFOSR grant 403-64 and U.S. Public Health Service grant *AM* 06399-03.

[1] Bailey, N. T. J., *The Mathematical Theory of Epidemics* (Griffin, 1960).

NEWS and VIEWS

Director of the Rowett Research Institute, Aberdeen: Dr. D. P. Cuthbertson, C.B.E.

Dr. D. P. Cuthbertson, who will be retiring from the directorship of the Rowett Research Institute, Bucksburn, Aberdeen, on September 30, 1965, obtained his medical qualification in Glasgow and was for nearly twenty years lecturer in pathological chemistry and physiological chemistry in the Royal Infirmary, Glasgow, and in the University of Glasgow. He became an authority on the biochemical effects of injury, and particularly on the acute depletion of protein following injury. His work on the prevention and treatment of these changes became of especial importance during the Second World War. Cuthbertson worked for two years on the headquarters staff of the Medical Research Council before he was appointed director of the Rowett Research Institute, Aberdeen, in 1945. Under his influence the Institute grew rapidly in size and importance, and it is remarkable how Cuthbertson was able to direct so much work and produce an interesting report each year, summarizing work that had been published in 100 papers or more. He managed, however, to keep his own laboratory going and to give a number of lectures, which were published. He is an accomplished artist in water colours and a keen golfer, and will clearly be able to find something to do when he retires.

Dr. K. L. Blaxter

Dr. K. L. Blaxter, at present head of the Nutrition Department, Hannah Dairy Research Institute, has been appointed to succeed Dr. Cuthbertson as director of the Rowett Research Institute. Dr. Blaxter, who is forty-five, is a graduate of the University of Reading. He has been with the Hannah Dairy Research Institute since 1948. Before that he was a research officer at the Veterinary Laboratory of the Ministry of Agriculture, Weybridge. During 1946–47 he was a Commonwealth Fund Fellow at the University of Illinois, Division of Nutrition, under Prof. H. H. Mitchell. With the exception of a period of war service with the Royal Artillery, he was a research assistant at the National Institute for Research in Dairying from his first graduation until going to Weybridge. Dr. Blaxter was Clive Behrens Lecturer in the Faculty of Agriculture, University of Leeds, during 1958–61; 1960 Thomas Baxter Prizeman and Gold Medallist (for research on the nutrition of dairy cattle); National Research Council of Canada visiting lecturer to the Universities of British Columbia, Manitoba, Saskatchewan and Alberta in 1964; 1964 Royal Agricultural Society Gold Medallist (for research on the nutrition of farm livestock). He also gave the Fernhurst Lecture to the Royal Society of Arts in 1962, lectures to the Professors Council of the University, Uppsala, and to the Berlin Academy of Sciences, D.D.R., in 1962, and the third Samuel Brody Memorial Lecturer in the University of Missouri in 1964. Dr. Blaxter's work has been concerned with the nutritive value of home-grown foods for dairy cattle, food intake and food utilization, protein requirements and protein metabolism of dairy cattle and calves, hyperthyroidism in the ruminant and the use of iodinated protein, hypomagnesæmic tetany, vitamin E metabolism and muscular dystrophy, energy metabolism and the energy requirements of cattle, and their relation to the environment. In 1962 his monograph entitled *The Energy Metabolism of Ruminants* was published. Dr. Blaxter is a member of numerous learned societies and is a member of the joint Agricultural Research Council–Medical Research Council Development Commission sub-committee on the monitoring of radioactive fall-out and of the Working Party on Ruminants of the Agricultural Research Council's Technical Committee on Nutrient Requirements of Farm Livestock.

Ministry of Aviation: Director of Engine Research and Development (1): Mr. Dennis H. Mallinson

Mr. D. H. Mallinson has been appointed director of engine research and development (1), Ministry of Aviation, in succession to Dr. J. Remfry, who has retired from the Public Service. Born in 1921, Mr. Mallinson was educated at Prince Henry's Grammar School, Otley, before going to the University of Leeds in 1939. He graduated with honours in applied mathematics in 1942 and was posted as a junior scientific officer to the Engine Department of the Royal Aircraft Establishment. As a member of the small but growing turbine division under Hayne Constant and W. R. Hawthorne, he was employed on engine performance investigations and particularly on the analysis of results from the flight tests of the pioneer jet aircraft. Transferring with the Royal Aircraft Establishment's Turbine Division to Power Jets, Research and Development, Ltd., he remained on engine performance work and was in charge of the Performance Section when the National Gas Turbine Establishment came into being. In 1949, following a period of about a year in which he acted as technical and administrative assistant to the deputy director (Mr. Peter Lloyd, the present director-general of engine research and development), Mr. Mallinson became a member of the National Gas Turbine Establishment's ramjet research team under Mr. R. P. Probert, concentrating first on thermodynamics but later encompassing fuel control systems and combustion development and finally assuming responsibility for the full-scale ramjet test vehicle programme. This programme was successfully completed in 1957 and from then until early 1963 he was superintendent of the Engine Research Division at the National Gas Turbine Establishment in charge of work on propelling nozzles, control systems and high-speed ramjets. He transferred to Ministry of Aviation Headquarters, London, in February 1963 as assistant director responsible for engine project assessment and continued in this post until receiving his present appointment. Dennis Mallinson is a district commissioner of the