



Remembering William Goffman: Mathematical information science pioneer

Glynn Harmon *

The University of Texas at Austin, School of Information, 1 University Station D 7000, Austin, TX 78712-0390, United States

Received 3 December 2007; accepted 13 December 2007

Abstract

This paper reviews the career and legacy of William (Bill) Goffman, who served as a researcher, Professor, Dean and Emeritus at Case Western Reserve University, Cleveland, Ohio, from 1959 to 2000. Goffman pioneered mathematical information science broadly and in several key areas. First, he applied disease epidemiology concepts to model accurately the spread of knowledge and the formation of knowledge systems and their ecologies, including the dynamics of scientific discovery. Second, he proposed significant improvements in information retrieval through the deployment of multi-valued logic, appropriate file ordering, effective and efficient retrieval measures, and simplified retrieval approaches, including early work in citation-based searching. Third, Goffman applied Bradford-like distributions to model effective core research literature collection development and usage. Fourth, he developed original epidemiology models, and was an early contributor in biomedical informatics. His mathematical contributions have stood the test of time and will continue to be applicable indefinitely.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: William Goffman; Mathematics; Information science; Bradford distributions; Knowledge epidemiology; Information retrieval; Scientific discovery

1. Biographical introduction

Although William (Bill) Goffman died several years ago, this article is intended to serve as a timely remembrance of his career and scholarly contributions, which remain relevant to contemporary progress in information science. He is remembered today as a leading pioneer in the establishment of mathematical information science. During his career as a researcher, Professor and Dean at Case Western Reserve University, he developed general mathematical models for several areas: epidemiology models to explain and predict the generation, growth and spread of scientific and other ideas; simple but direct and efficient information storage and retrieval approaches and algorithms; measures of information relevance in the retrieval process; the use of

* Tel.: +1 512 471 3972; fax: +1 512 471 3971.

E-mail address: gharmon@ischool.utexas.edu

Bradford distributions to explain and prescribe core literature development and rank-ordered retrieval file organization.

He and his co-authors were among the very first critics of the wide spread deployment of Boolean logic, despite its sub-optimality, as the basis for information retrieval—a practice that became chronically entrenched and institutionalized, and endures in most of our present-day systems. Goffman and colleagues also extended Bradford distribution probability principles to the realms of literature organization, retrieval effectiveness and efficiency, and core collection development. They helped pioneer the emergent fields of medical informatics, including public health informatics and clinical decision-making by finding remarkable similarities between the literatures describing the etiology and epidemiology of vastly different diseases. And, they applied the emergent logic of clinical decision-making in medicine (centered on true or false positive or negative diagnoses) to illustrate corresponding similarities in information retrieval processes.

Goffman can thus be remembered in several ways. He was a versatile and creative mathematical information scientist, an imaginative and resourceful polymath, a logician, a catalyst for the later emergence of bibliometrics and informetrics, and a unique kind of epidemiologist. In addressing and resolving information science problems, he seldom failed to surprise his colleagues and students. While others would gambit about the periphery of what they thought were authentic information science problems, he would quietly and quickly reframe the problem at hand, push aside irrelevant variables, quickly isolate the critical ones, and derive a solution. He was truly a master of convergent thinking, but would surprise others by diverging from the topic under discussion to bring in seemingly irrelevant but appropriate metaphors.

Bill Goffman was born in Cleveland, Ohio on January 28, 1924. He was the fourth of four children and the second son born to Sam and Mollie Goffman. Bill's father, Sam Goffman, was born in 1889 in Kuschenev, Russia and migrated to Montreal, Canada and then to Cleveland, Ohio in 1907, where he lived until his death in 1975. Sam spent most of his life in Cleveland, where he worked as a stonemason, contractor and employer of many construction workers. Sam's work in stonemasonry and construction gave him an affinity for numbers and measurement, and motivated him to study mathematics. He encouraged his children to do likewise. Bill once remarked that the intricate nature of his father's stone measurement and construction problems provided an intuitive foundation for such topics as algebra, plane and solid geometry, and set theory. This intuitive foundation was conveyed especially to sons Casper, who pursued a career in theoretical mathematics, and Bill, who became an applied mathematician.

Bill's mother, Mollie (Stein) Goffman, was born in Ryegrud, Poland, near the border with Russia, in 1895. In 1907, she migrated to Baltimore, Maryland, a point of entry, and then to Cleveland, where she wedded Sam and lived until her death in 1976. Four children were born to Sam and Mollie Goffman: Casper, Ruth, Sarah, and Bill (the youngest by 11 years).

Bill attended local schools and graduated from Glenview High School in 1943. During World War II, he joined the US Army, went through flight training in Texas, and became a commissioned officer and pilot in the US Army Air Corps. He flew combat missions in the Pacific theater and attained the rank of Major.

He attended the University of Michigan at Ann Arbor and received a B.S. degree in mathematics in 1950 and a Ph.D. degree in mathematics in 1954. During and after his graduate studies, he served as a part-time pilot for Pan American World Airways, flying Lockheed Constellations out of New York's Idlewild Airport (now John F. Kennedy International Airport). He also worked as a mathematical consultant following his graduate studies and continued to reside in Cleveland.

In 1959, during one of his frequent trips to the New York area, Bill had the good fortune of being introduced by a research colleague to Patricia Ann Mc Loughlin, a registered nurse, who was working at the Jersey City Medical Center, New Jersey. They continued their relationship, and Bill and Pat were married on February 7, 1964 in New York City. Pat then joined Bill in Cleveland in 1964.

He also joined the staff of the Center for Documentation and Communications Research (CDCR) in 1959 as a staff mathematician and later a research associate. The CDCR had been established in 1955 as a research arm of the School of Library Science at Western Reserve University (Western Reserve University was federated with Case Institute of Technology in 1967 to become Case Western Reserve University). As a staff mathematician and associate at CDCR, he joined such notable information retrieval pioneers as James W. Perry, Allan Kent, Jesse H. Shera, Tefko Saracevic, Jack Belzer, Allan Rees, and Al Goldwyn. Bill immersed himself in mathematical modeling for information retrieval system design, effective file organization, machine

searching heuristics and algorithms, and the application of epidemiology concepts to the spread of scientific ideas. Through the years, he also developed good working relations with notable colleagues outside of CDCR, such as A.D. Booth, Eugene Garfield, Vaun A. Newill, Tomas G. Morris, and Ken S. Warren.

The CDCR was an especially exciting place to be in the 1960s, since many seminal information retrieval concepts were framed there, along with a significant amount of the US effort to develop computer-based retrieval during the Cold War with its military mobilization and the US-Russian space race. CDCR developed the Rapid Selector, an early machine-based retrieval system. The Center also pioneered new approaches to abstracting, indexing and current feedback systems. And the Center established the Comparative Systems Laboratory, headed by Tefko Saracevic, for the innovative testing of different retrieval systems and approaches. It was there that Bill Goffman developed a reputation for resolving the toughest and most complex information retrieval problems. He served as a sort of “Supreme Court” for handling relatively intractable cases through his ability to restate retrieval problems clearly and precisely through his mathematical postulations, rigorous proofs, and mobilization of multi-disciplinary metaphors.

Bill also served as an exemplary teacher and role model for students. Over the years, he developed and offered such graduate courses as information retrieval theory, mathematics for information scientists, general communications theory, systems analysis, bibliometrics, and seminars for doctoral students. He also supervised numerous doctoral committees and dissertations. In 1968, he became a full professor in the School of Library Science, where he continued to specialize in communication theory and information retrieval.

He was appointed Dean of the School of Library Science at Case Western Reserve University in 1971 and served in that role until 1977. Despite heavy administrative duties, he continued to be an active researcher, teacher and dissertation supervisor. As Dean, he promoted multidisciplinary problem solving and education. He was principal investigator for National Institute of Health and National Science Foundation grants. He obtained foundation funding to establish a University Complex Systems Institute, and served as its Director from 1972 to 1975. In 1978, he returned to full-time teaching and research and from 1981 through 1988 served as Director of the Training Program on Computer Applications in Health Sciences at the Case Western Reserve School of Medicine. He also served as a Professor in the Department of Biometry of the School of Medicine from 1980 to 1986.

Bill received a number of honors during his career. He was elected Fellow of the American Association for the Advancement of Science in 1982. He had the honor of serving three times as an invited, resident scholar at the Rockefeller Foundation Bellagio Study and Conference Center, Bellagio, Italy during the 1970s and early 1980s. Bill is listed in *Who's Who in the World* (4th ed., 1978/1979), *Who's Who in America* (34th ed., 1976/1977), and *Who's Who in Science and Engineering* (1st ed., 1992/1993). In the 1960s and 1970s, he also held visiting professorships at the University of Montreal, Georgia Institute of Technology, the City University of London, and the University of Denver.

He retired from active teaching in 1986, and was named Professor Emeritus and Dean Emeritus of the School of Library and Information Science. In retirement, he continued to do research, part-time teaching and consulting, and pursued his hobbies. Following his father's craftsmanship propensity, he enjoyed refurbishing his beautiful home in nearby Bratenahl on the shores of Lake Erie, a suburb of Cleveland. Bill had many other hobbies. He enjoyed classical music, such as string quartets—particularly those played by the Budapest String Quartet; he had a *huge* collection of tapes, records and discs. He was an avid baseball fan, and would buy season tickets to watch the Cleveland Indians. His strong reading interests involved ancient to modern history, numerous classics, and various mystery novels, particularly those by Dick Francis, Rex Stout, Dashiell Hammett, Raymond Chandler, and Agatha Christie. Bill especially liked to travel with Pat on cruise ships, particularly on the Queen Elizabeth. On one trip to Panama with colleague Donald Booth and wife, they rode out an Atlantic hurricane, which Bill found tremendously exhilarating. This experience prompted more cruise trips. Bill and Pat also enjoyed air travel, including yearly flights on the Concorde between New York and London. He and Pat took several other trips throughout and around the world before and after his retirement. Bill died of lung cancer at his home on February 29, 2000, a month after his 76th birthday. His wife Pat survived him, along with Casper Goffman, his very close older brother.

Casper Goffman, incidentally, was also a brilliant mathematician; he spent most of his career teaching and conducting pure mathematical research at Purdue University. But Bill's frequent communications with him

tended to focus on their strong mutual interests in baseball players and teams, or on various classical music works and artists.

The following sections summarize the key thrusts of Bill Goffman's research and briefly discuss its significance. Coverage is necessarily selective; works cited represent most but not all of his published works.

2. Early contributions in information storage and retrieval

When Goffman started working at the CDCR in 1959, there existed a very general and intense interest in furthering, refining and even perfecting earlier documentation practices for information storage and retrieval, and for automating them. Efforts were typically based on the assumption that methods for the manual organization of files and document representations could and should be applied to automated methods of indexing and retrieval. This tendency was somewhat reminiscent of earlier phenomena whereby new technologies were viewed as extensions of older ones, despite the emergence of a quite different technology. It was at this juncture that skeptical mind of Bill Goffman kicked in by questioning a priori concepts. Goffman and colleagues demonstrated mathematically that the well-entrenched reliance on Boolean logic was problematic. They concluded that the use of Boolean operators for manual and automated information retrieval was inherently sub-optimal and inefficient. For example, retrieval of the intersection of subject "a" and "b" risks giving too much irrelevant information, while retrieving "a" or "b" risks leaving out relevant material (Verhoeff, Goffman, & Belzer, 1961).

Goffman challenged another basic information retrieval mantra—the notion that relevance could be used as a measure of information carried by a single document in relation to a given query. He noted that for relevance to qualify for as a measure, one must define relevance in relation to the entire set of documents, rather than to only one document. That is, relevance as a measure must be additive (based on the aggregate of documents in a set) rather than on non-additive mappings to individual documents (Goffman, 1964b). Although the information retrieval community largely ignored Goffman's challenge for many years, his paper did anticipate and pave the way for later retrieval practices that incorporated aggregate relevance judgments.

Likewise in 1964, Goffman noted that the manual and automated retrieval processes of the time tended to be remarkably inefficient and cumbersome. Information retrieval typically involved searching through the entirety of a large file of documents or document representations to obtain a relevant set of useful documents. He proposed that instead we should treat each query as a set of indicators that has a probability distribution and could be mapped onto an entire file of documents stored in a system. Thus, on the basis of trial probes, we can partition out a subset or sector of documents that might be highly relevant to our query and then search intensively within that subset. Such a direct search of high relevance documents could increase searching efficiency dramatically (Goffman, 1964a). To further efficiency, Goffman, Verhoeff, and Belzer (1964) proposed the use of abstract and powerful meta-languages in information searching to replace the use of numerous, highly specific terms being laboriously used to frame queries.

Goffman also argued against the almost exclusive use of two-valued logic and championed the use of multi-valued logic in information retrieval. Both manual and machine retrieval operations in the 1960s tended to be based on the classical, two-valued Boolean propositional calculus. Either a document was relevant to a given query (having a value of 1) or it was not relevant (having a value of 0). He demonstrated that retrieval is far too complex to be based on either-or logic and called for the use of infinitely valued logic. He was thus an early proponent of the use of fuzzy logic in information retrieval (Goffman, 1965a).

In 1966, Goffman and his Western Reserve University Medical School colleague Vaun A. Newill evaluated retrieval systems through the combined use of *effectiveness* measures (the system's ability to perform its designated purpose) and *efficiency* measures (the relative cost of task performance). They further incorporated analogous medical diagnostic test criteria into information retrieval. Medical test *sensitivity*, a test's ability to detect and identify the actual presence of a disease or disorder, could be applied to the retrieval of highly relevant documents, and these retrieved documents could be regarded as "true positives". Conversely, a medical test's tendency to sometimes inaccurately indicate the presence of a disease when it was actually not present is regarded as a false positive. A retrieval system's identification of non-relevant documents as ones that were relevant could be referred to as "false positive" retrieval. In contrast, *specificity* was defined as the

system's ability to identify non-relevant documents (true negatives). Both specificity and sensitivity were deemed to be essential to a system's effectiveness. But Goffman and Newill also noted that some user queries do not represent the users' actual, underlying information needs. They thus went on to differentiate between the concept of relevance, (the relation between a user query and a document set) and pertinence (the relation between a need and a document set). They then proposed a remarkable, holistic approach to retrieval system testing and evaluation that would incorporate the above notions of effectiveness, efficiency, sensitivity, specificity, relevance, and pertinence into a unified model (Goffman & Newill, 1966).

In 1969, Goffman followed through on his earlier, direct method of information retrieval (isolating potentially relevant subsets and searching intensively therein) by proposing a counterpart concept, an *indirect* method of retrieval. Because searches were based on what was known about a subject, the documents that were retrieved and judged to be relevant were based on what was already known, and could be used to direct further search probes. He proposed exploiting this retrieval practice by treating a given scholarly paper as a query itself and then using the paper's cited references as relevant query "answers". In turn, the citations in these cited references could be further pursued as answers of a second order. When put to the test, this indirect method of information retrieval yielded far better results than the earlier direct method of identifying and probing file subsets (Goffman, 1969). In doing this, Goffman was proposing the formalization of the then relatively new method of citation-based retrieval that Eugene Garfield and his associates were developing at the Institute for Scientific Information in Philadelphia.

In conclusion, Goffman's early contributions in information retrieval were based on a healthy skepticism about some of the leading assumptions of the time. He observed the sub-optimal results obtained from the use of Boolean logic, particularly the severe tradeoffs between being too inclusive or too exclusive. He advocated the use of relevance judgments based on holistic rather than singular analyses of information representations, and the use of highly selective information retrieval probes. He developed sophisticated ways to evaluate information retrieval performance through the use of multiple measures, such as effectiveness, efficiency, sensitivity, and specificity. And he was an early proponent of citation-based information retrieval. It is probably fair to say that those who worked with Bill Goffman regarded him as a subtle but highly effective iconoclast. He would deftly challenge basic information retrieval premises and protocols of the time and quickly replace them with more accurate, holistic and powerful models.

3. Bradfordian distributions and core literature identification

Goffman was intrigued with laws of literature dispersion, particularly Bradford's law, and took great delight in pointing out to students and colleagues its broad implications. Bradford's law states basically that there are predictable regularities in the way that a very few journals in a given topical area contain about a third of the available articles in that area. It requires a significantly larger number of journals to produce the next set of a third of all related articles, and a still much greater number of peripheral journals to produce the remaining third of all available articles on the topic. Bradford expressed these ratios as $1:n:n^2$.

Goffman and his close colleague, Kenneth Warren, reported in 1969 on their extensive National Library of Medicine-funded efforts to isolate general laws related to the dispersion of medical literature. It was and continues to be important to sift through a vast output of medical literature to identify essential, core literature related to basic and clinical research topics (Goffman & Warren, 1969). They selected two very unlike literatures (on mast cell research and schistosomiasis) to search for consistent relationships that would identify core concentrations of literature in specific areas of health science and medicine. Despite differences in content and size of these two different literatures, they found first that the ratio of journals to articles in both fields was nearly identical; second, the average number of papers by authors in each area did not vary significantly within each 5-year period through eight decades; and third, the paper/author ratio was about equal to the minimal constant of Bradford's law of dispersion. Goffman and Warren thus adapted Bradford's law to identify core medical literatures, a potentially invaluable guide for formulating journal acquisition policies for research groups and libraries. Such core literatures could be identified through citation analysis or actual journal usage data to designate a minimum nucleus, and successively larger groups of journals identified to add secondary and tertiary collections of journals around the core. Each grouping could be optimized to fulfill needs, conserve finances, and adjust to other circumstances.

To further test these findings, Goffman joined with Tomas Morris, Cleveland Medical Association Library director, to conduct a detailed analysis of periodical circulation and to designate a list of medical journals rank-ordered into successively larger zones according to apparent usage. Here too, the most highly circulated journals represented a minimal but essential, user-defined core for that setting at that time period. It required a dramatically larger number of journals to produce the number of generally used articles equal to the number in the minimal core set. They proposed that journals should thus be acquired according to their relative information usage value, budgetary constraints, and other local factors (Goffman & Morris, 1970).

Goffman and Warren then extended their application of Bradford's law to a broad analysis of the ecology of medical literatures and related Bradfordian principles. They analyzed how knowledge is communicated in ways resembling the epidemic spread of diseases. They debunked the then-current notion of the existence of a literature explosion; while there was indeed an exponential growth of literatures in various medical fields, the ratios of papers to authors and that of high quality papers to the total number of papers (as indicated by citation analysis) remained constant over long periods of time. They developed a case against the professionally popular calls of the time to alter the basic ecology of medical literature to deal with a supposed literature explosion. They argued that altering the overall ecology of medical literature would probably destroy mechanisms essential to the natural selection of high quality literature along with the destruction of literature quality. Ecology alteration would also interfere with the cyclical output of literature in various areas of investigation. Instead, both literature dissemination services and individual scientists should apply Bradford's law to identify core nuclei of highly cited and otherwise significant literature, and then identify secondary and tertiary cores according to need. These insights could be particularly useful for financially restricted medical research groups and libraries here or abroad (Warren & Goffman, 1972).

One of Goffman's key applications of Bradford's law is illustrated in a significant, jointly authored article. The article illustrated how medical libraries and other research organizations in the world's developing countries could afford to acquire essential, core medical literature in the face of rapidly increasing numbers of journals being published and their escalating costs (Bruer, Goffman, & Warren, 1981). In the late 1970s, for example, Latin American medical school libraries each held from fewer than 300 up to about 1200 journals, but the great majority (about 90%) each held fewer than 300 journals. In contrast, major US medical libraries held 2000–3000 journal titles each. Goffman and colleagues used Bradford's law and bibliometric analysis to identify minimal core collection lists consisting of 35 general nucleus journals and 92 specialized journals that could be portioned into 31 medical specialty classes. These minimal and specialty cores could be adjusted to local user demands and other circumstances to conserve funds, bring about substantial savings and improve access to higher quality journals. Goffman and colleagues noted that a geometric increase in the quantity of literature results in only a small arithmetic increase in utilization, and that it was neither practical nor cost-effective to acquire more than a small threshold number of journals, since utilization capacity could easily be surpassed. But larger collections could be developed incrementally by lowering the required citation frequency of candidate journals to acquire to build needed specialty collections, and to participate in regional resource sharing. Thus, journal supply and demand relationships could economically be optimized to satisfy educational, clinical and research needs, especially in developing nations and regions.

Goffman extended these concepts by proposing an additional approach to literature selection (Goffman & Warren, 1980). He was critical of the then-typical library approaches that sought to acquire and preserve everything printed rather than to make key contributions available to advance personal or scientific knowledge. He likewise objected to anecdotal approaches to journal and other document acquisitions based on tradition, consensus, and casual relevance assessments. He also criticized information retrieval approaches that divorced retrieval from their broader knowledge communications contexts and their uncritical use of Boolean logic. He again proposed that instead we try to build highly relevant Bradfordian core sets of literature geared to local user populations and conditions. Regional resource sharing could be used to augment such core sets. Last, he advocated going beyond the rigidities of Boolean retrieval by developing relevance networks based on citation linkages among related papers, especially for the strategic advancement of knowledge.

While the Internet and web have obviously changed knowledge access patterns, Goffman's proposed adaptation of Bradfordian principles, citation networking, and going beyond Boolean retrieval logic all remain applicable to this day.

In September, 1993 Goffman applied for a US patent that would apply his application of Bradford' law and his bibliometric methodology for selecting core collections of journals and other objects in various settings. The patent (US Patent Office number 5,594,897) was granted on January 14, 1997. Thus one might add the word “inventor” to describe Goffman's extensive realm of talents.

4. Epidemic spread of knowledge

Goffman is probably best known for his mathematical work in the realm of communicable disease epidemiology and its application to the growth and spread of knowledge. He and his colleagues studied similarities between the genesis, spread and decline of various diseases across populations over time. They then drew parallels with how human individuals and populations become “infected” with ideas and knowledge fashions, and successfully modeled these processes against empirical data.

In the early 1960s, Goffman and Newill started drawing parallels between disease and knowledge epidemics. For a disease to originate and spread, there must exist infective material and a vulnerable population. In turn, the population is made up of *infectives* (hosts to the infective material), *susceptibles* (those who can be infected), and *removals* (those who are susceptible or infected but leave the population because of immunity, hospitalization, or death). Exposed individuals are either immune to the disease or can become infected through contact with a disease host or vector. The time required for an infective to develop the disease after contact with infectious material is known as a *latency* or incubation period. An epidemic may be said to occur when a disease outbreak involves an abnormally high number of individuals. Goffman and Newill extended the above epidemic concepts by discussing various knowledge epidemics, such as the spread of major religious ideas and such scientific concepts as those of Newton, Darwin and Freud. But they noted that while most idea epidemics might be regarded as desirable, disease epidemics are generally not desired. Vulnerable members of a scientific or lay population could, for example, be infected with a concept published in a book or journal (a vector), while other members, being immune, might reject that concept. Or, immune individuals might transmit the concept without being infected (the “Typhoid Mary” phenomenon). Once an abnormally large number of population members become infected with the concept, a knowledge epidemic can be said to occur.

But could such knowledge epidemics be successfully modeled to explain or predict the actual spread of scientific or other concepts based on historical data? Goffman and Newill attacked this question by developing a general deterministic model based on differential equations to show how infectives are involved in the spread of mathematical or medical concepts, especially in large populations. They briefly explored a contrasting stochastic modeling approach, based on finite state Markov processes and particularly applicable to smaller populations, to derive the probability of new levels of infective cases. Mathematical epidemic approaches could thus be used to depict the dynamics of knowledge epidemics, despite their necessary limitations (Goffman & Newill, 1964).

Goffman elaborated on these epidemic models by specifying epidemic processes in *open* populations. If epidemic processes are treated stochastically, they can be represented with finite state Markov models using either discrete or continuous parameters. Discrete parameters are applicable where the latency period is constant, in which case infectives occur through multiple generations. Continuous parameters are applicable when latency periods are variable. Past deterministic and stochastic treatments of epidemic processes, he noted, tended to deal with *closed* populations, wherein population N remains constant throughout the epidemic. But in actual epidemics the total number N tends to vary considerably, since the number of susceptibles, infectives, and removals change throughout the epidemic process. New supplies of infectives tend to be introduced continuously. An epidemic will peak and decline if and only if the rate of change of removals is constant. (Goffman, 1965b). It is remarkable that in one short article in the journal *Nature*, he demonstrated that most epidemic modeling in the 1960s addressed only the special case of closed populations. In effect, Goffman became an epidemiologist per se, without reference to knowledge epidemiology, by introducing a more general epidemic model pertinent to open populations.

Next, he looked at the relative stability of epidemic processes by extending Russian mathematician A.M. Lyapunov's theory related to differential equations. Sub-population numbers in open epidemic populations tend to remain in flux or in state transitions, owing to different degrees of potency of infectious material, exposure, immunity, etc. Consequently, populations can be either in an increasing, stable or declining disease

state. While closed populations tend toward a relatively predictable stable or equilibrium states, the size of open populations is variable, given the differential rate of introduction of new susceptibles and infectives. Thus, outcomes are less predictable in open populations, and there is no guarantee that the epidemic will reach a maximum and then transition to a stable or declining state. To clarify the conditions under which an open population epidemic might stabilize or decline, Goffman extended Lyapunov's differential equation stability theorem and a uniqueness theorem to specify points along a growth or decline trajectory at which epidemic stability might occur (Goffman, 1966a). Again, he produced an original contribution to mathematical epidemiology.

He then returned to the problem of exploring the epidemic-like spread of knowledge. But he was then interested in explaining the more general problem of the growth of science itself. As a case in point, he explored the history of mast cell research. Goffman felt that it was fortunate that the acclaimed stress theorist, Hans Selye, had assembled a complete bibliography on the topic (which covered the period from Erlich's discovery of mast cells in 1877–1963). Selye also made his personal library available to Goffman. Bill Goffman then defined authors of mast cell works as "infectives", who would later become "removals", a year following their last publication on the topic. From 1877 through World War II, mast cell publication output was minimal and sporadic, but it reached epidemic proportions following World War II and could be projected to increase well into the future. Goffman noted that this epidemic model was "completely general" and could be applied in any scientific area to assess the comparative importance of existing and emergent lines of inquiry and anticipate new ones (Goffman, 1966b).

About the same time, he developed a model to explain the conditions of disease co-occurrence. Two co-occurring diseases in a population might either be independent of one another, or one might augment or inhibit the other. Goffman put forth an extensive set of proofs to explain the probability of individuals being infected with two diseases simultaneously or sequentially (Goffman, 1966c).

Goffman and Newill resumed their investigation of knowledge epidemiology by looking beyond the rather narrow information retrieval paradigm of the time that centered on file organization and query matching. They addressed the wider problem of achieving optimal, effective and timely knowledge diffusion throughout specified populations. In a lengthy, elaborate mathematical paper published in the *Proceedings of the Royal Society of London* (Goffman & Newill, 1967), they proposed that both the transmission of diseases and of ideas are special cases of a more general communication process. In both cases, individuals are exposed and are either immune or infected, and might or might not transmit infection to others. Eventually, the epidemic might become dormant.

An information retrieval system might thus be regarded as something embedded in a population to stimulate an epidemic. The retrieval system's purpose would be to provide effective (relevant) contact between infectives (document authors) and susceptibles (searchers or inquirers) so that ideas are transmitted and knowledge is acquired. They proposed a number of definitions and theorems dealing with several topics: population stability; information relevance relations and transformations; one-to-many relevance mappings or closures between susceptibles and infectives; and the ordering of documents and exposures. For a given document set to be relevant, it must have compatible relevance with other documents in its set; in contrast, a partially relevant document in the set might inhibit the relevance of others in the set. It is noteworthy that Goffman and Newill explained these concepts also in terms of systems theory; they pointed out that systems consist of sets of elements that interact optimally to perform specific functions (procedures) to fulfill specific purposes (requirements). An information retrieval system's output consists of documents (agents) that convey relevant (infectious) information in response to queries posed by users (susceptibles). Systems theory thus augmented their eclectic use of communications theory, mathematics, epidemiology and information retrieval.

With Ken Warren, Goffman applied the Kermack–McKendrick model of epidemics, a model particularly applicable to schistosomiasis since it applies to cyclical and multi-stage diseases involving intermediate hosts. Schistosomiasis was and continues to be a serious, widespread disease outside of the US and is transmitted by parasitic worms carried by snails. Goffman and Warren determined three kinds of required threshold numbers and density levels: those above which snail populations must exist to bring about schistosomiasis epidemics in human populations; those at which epidemics would stabilize; and those below which epidemics would decline or cease to exist (Goffman & Warren, 1970). They thereby provided critical parameters for public health officials to control snail populations and consequently to control schistosomiasis epidemics.

In the 1960s, Goffman was pleased to find published, exhaustive bibliographies that encompassed the origination and development of symbolic logic (Church, 1936, 1938). To further test his epidemic theory of the growth of science, he observed that the long-term growth of symbolic logic was characterized by continual evolution and diffusion. Boole and DeMorgan had published their seminal formulation of symbolic logic in 1847. Prior to that time, works on symbolic logic were scattered, rudimentary, non-cumulative, and spurred by some persistent stimulus that originated in the remote past, Goffman observed. After 1847, in contrast, publications in symbolic logic were outgrowths of related publications preceding them. From the bibliography, it became clear that scholars would juxtapose and synthesize ideas from the past to produce new concepts. New concepts would then diffuse through a population and result in production cycles of yet newer syntheses. Thus, Goffman reasoned that epidemic theory could be applied to establish quantitatively the relative importance of past lines of inquiry, and to predict the behavior of current lines of investigation and the probable emergence of new research directions. He provided saw-tooth-like graphs of the cyclical development of symbolic logic from 1847 onward. Initial publication output was sporadic and limited, with peaks occurring in 1867, 1877 and 1892. Symbolic logic then entered successive epidemic states, with peak points occurring at 25-year intervals (1907, 1932 and 1957). Goffman observed that initial, original works in the foundations of mathematics stimulated discoveries on the paradoxes of set theory, and later stimulated developments in prepositional calculus, meta-mathematics, intuitionism, and other sub-fields. Applications of epidemic theory demonstrated that once there was a reasonably successful culmination of work in one area, new questions would be evoked and new lines of inquiry would then build to form newer periods of activity of different sizes and intensities. He then applied a stochastic (Markov chain) model to explain further the evolution of symbolic logic. Apparently, the seven sub-fields of symbolic logic that emerged from 1847 to 1932 had developed as a somewhat predictable sequence of overlapping epidemics. But he could find no predictable pattern of migration of researchers from one sub-field to another; once researchers left an area, they were unlikely to return to make further contributions (Goffman, 1971).

5. Scientific discovery and growth of knowledge

Over the years, Goffman developed a keen interest in the scientific discovery process. He viewed the growth of specific areas of knowledge as the cumulative result of a succession of related, epidemic-like discoveries. As my dissertation supervisor, he challenged me in 1968 to write a dissertation that would provide a model for predicting the occurrence of scientific discoveries. Under Goffman's guidance, this dissertation effort produced a systems model and general equation that would illustrate how previous discoveries were made: a series of key contributions in a given area would over time aggregate into a discovery (a disciplinary sub-sub-system). In turn, these discoveries would eventually aggregate into disciplinary sub-systems (sub-fields), and these sub-systems would aggregate into disciplinary systems. Disciplinary systems would form into disciplinary super-systems (such as the biological, physical or social sciences). There were chronological regularities in the making of key contributions and their manner of aggregation into discoveries. In all cases, human short-term memory limits of about seven informational chunks required that contributions be summarized and aggregated periodically; otherwise, a logjam of contributions would tend to occur and stop progress (Harmon, 1970, 1973).

Following dissertation completion, I remodeled the above discovery process in terms of set theory. A discovery itself could be seen as a complete, finite and ordered set of about seven cognitive elements or information chunks formed within the discoverer's short-term memory. Such an ordered set served to provide a meaningful organization of knowledge within a realm of scientific discourse. That is, the discovery process involved a series of cognitive set transformations involving successive mental iterations through four states or stages: Possession of (I) insufficient and unordered information elements; (II) insufficient but ordered information; (III) sufficient but unordered information; (IV) sufficient and ordered informational elements—the discovery state. A two-by-two matrix represented these four states. The discovery event thus reflected a successful culmination of a series of efforts to gather the necessary information elements, to define an appropriate set-defining criterion, and to organize the set. The set was regarded as finite, consisting of about seven elements, because the number of elements had to be restricted to a cognitively manageable number to accommodate the human working (short-term) memory capacities of both the discoverer and each individual who might seek to understand or use that discovery. Once the discovery was made, the set could be further elaborated,

tested and refined. We then further modeled this discovery process as a four-state Markov chain, in which the stages or states could be represented as probabilities of movement of an incipient discovery system between the four different states. The previously mentioned Church bibliographies on the development of symbolic logic (Church, 1936, 1938) were applied to test the model. Church had marked key contributions in the bibliography with single asterisks, and significant breakthroughs with double asterisks. The Markov model showed that symbolic logic discoveries resulted from forward and backward oscillations through states I–III until a discovery emerged as state IV. The mean recurring times for states I–IV were, respectively, 2.2, 4.0, 4.4 and 12.5 years. But each discovery state apparently posed new and different questions and the probability of the discovery set remaining in discovery state IV was 0. Throughout the development of symbolic logic, a discovery could be predicted to occur every 12.5 years. The Markov transition possibilities was irreducible, since each or the four states could be reached from each of its other states. Each state was a function of its preceding state, and all states were representative of the others (ergodic), so the limited memory of the Markov model was sufficient to explain transitions. In summary, these discoveries resulted from a predictable pattern of oscillations through three states to acquire the necessary information elements and order them appropriately to arrive in state IV, the discovery state (Goffman & Harmon, 1971).

Goffman later teamed with his colleague and former doctoral student, Tefko Saracevic, in an attempt to understand the mechanisms involved in forecasting information user and use requirements for subject literatures. They did an analysis of the structure and behavior of sample scientific literatures through extended time periods. They reported their findings at the 1977 International forum for Information and Documentation, noting that Goffman and Warren's theory of epidemics and holistic literature ecology concepts could be invoked to furnish qualitative and quantitative literature developmental parameters that would serve to forecast needed capabilities for large information services. The old ritual of building exhaustive collections was no longer feasible or even possible, given the huge numbers of journals in most scientific areas and the trend for scientists to demand quality filters. They further observed that future research should explore how to provide high quality literature, but with attention to qualitative–quantitative tradeoffs (Saracevic & Goffman, 1977).

Goffman worked with another former doctoral student and colleague, Miranda Lee Pao, on several projects. For example, they sought to refine a citation impact measure to assess the quality of a large, 20-year body of schistosomiasis literature. Their statistically rigorous analysis revealed that a small core of 351 of nearly 10,000 researchers in this field (3.51%) had produced a third of the field's literature over a 15-year period. They also found that the most productive half of grant awardees over the same period produced the highest quality research, as indicated by citation impact factors and average influence weightings (the latter developed by Computer Horizons Inc.). The more productive researchers tended also to link parasitic diseases with the broader realm of modern biology. Sustained funding of schistosomiasis research over extended periods of time had served to enhance the high quality research, they concluded (Pao & Goffman, 1990). Their remarkably timeless article is worth reading today.

While mentoring a succession of doctoral students through their dissertations, Goffman frequently expressed concern that information science itself was slow in defining its problems rigorously and in developing the abstract body of knowledge required to qualify it as a true science. In his noteworthy paper, "Information Science: Discipline or Disappearance" (Goffman, 1970), he observed that both library and computer science were excessively rooted in their respective institutional and technological realms. Thus, neither area had been able to produce neither abstract modeling of key concepts nor rigorous theoretical formulations. The aim of information science, he argued, should be to develop a unified approach to the study of information phenomena throughout the biological, physical, social and technological realms. Through the deployment of such approaches as the modeling of knowledge growth patterns, including citation graphing and relevance mapping, information science might develop into an authentic scientific area. Otherwise, it would become extinct.

But Goffman did not confine himself to information science when it came to analyzing scientific discovery and knowledge growth. For example, he joined with colleague Michael Katz, to explore the preformation of ontogenetic patterns in biology. Preformation theory asserted the existence of pre-determined, prototypical genetic and embryonic models that govern the development of organisms. Modern biology had abandoned eighteenth and nineteenth century preformation doctrine in favor of epigenetics, which holds that adult organisms possess features that are not pre-existent in germ cells; such features are acquired through a process of complex, synergistic interaction during development. But Katz and Goffman argued that preformation

explanations still have validity, and can be seen in DNA formations of the genome and within cellular and multi-cellular development. They analyzed the ontogenetics and growth of fresh water snails and proposed that preformed, precursor patterns were indeed present and had been conserved during ontogeny. They applied mathematical typology and discrete element pattern analysis to show how intra-cellular generators influence the course of cellular development. They demonstrated that generator configurations in given cells serve as skeletal representations of cellular cytoplasm organization in both unicellular and multi-cellular organisms. Therefore, many final organism features could generally be traced to preformed, precursor elements through conserved typology, despite some variations in the process (Katz & Goffman, 1981). In retrospect, these findings appear to be fundamental and revolutionary in modern biology. The authors had proposed and validated a more qualified, middle ground theory between the pure extremes of preformation and epigenesis.

6. Scientific communication: A synthesis

In the late 1970s, Goffman collaborated with Ken Warren to co-author a book, *Scientific Information Systems and the Principle of Selectivity* (Goffman & Warren, 1980). Their joint effort integrates and extrapolates much of their previous work to produce a general model of the development of scientific communication systems and their ecologies.

Their book begins by evaluating a number of differing perspectives and problems related to scientific communication processes. They point out that even the comprehensive *Index Medicus* in the late 1970s covered only about an eighth of the world's extant biomedical literature. Key contributions, like Mendel's 1865 paper on heredity, were often ignored, largely because they were paradigmatically premature. Insignificant "sewage" literature often clogs scientific communication channels and obscures significant literature. Roughly only a half of the then-existing scientific papers were ever cited at all, or were only cited once, and slightly more than a half of library collections went used. Yet, the then-popular conceptualizations of a "literature explosion crisis" and frequent calls to completely revamp the scientific communication system both tended to oversimplify the problems and preclude their resolution.

In contrast, Goffman and Warren argued that formal and informal formal scientific communication systems and ecologies had evolved in a complementary and selective fashion over time and had developed their own adaptive mechanisms. Accordingly, what was needed was not to radically change these systems, but to adjust their quality filtering mechanisms through optimal literature selectivity and retrieval. They provided a comprehensive and original mathematical foundation to model the structures and dynamics of scientific communication systems and their ecology, including the synthesis of scientific findings. Their general model masterfully incorporates elements of the classical Shannon theory of communications; Bradford distribution principles; Garfield's citation indexing; knowledge epidemiology; and the mathematical axioms of identity, permutation, transitivity, and self-distribution. Their model explains how findings are grouped into conceptual syntheses.

They tested their general model by marshalling data from comprehensive bibliographies of quite different literatures: schistosomiasis, mast cell research, symbolic logic, biochemistry, nutrition, immunology, neoplasia, and other areas. Despite the diversity of these literatures, they graphed remarkably similar cyclical patterns of growth and decline of literature production for each area through extended time periods, thus supporting their basic model.

Goffman and Warren further explored the qualitative dimensions of scientific literature through the use of Bradfordian distribution analysis. They noted that small circles of elite scientists tend to produce the most significant (highly cited) literature in their respective fields, and that the contributions of the great majority of scientists tend to be relatively minor. Cost-effective filtering can be achieved by centering on the more frequently cited literature worthy of publication in highly ranked journals, state-of-art reviews and abstracts.

Last, they proposed a comprehensive policy for improving scientific communications. They noted that scientific communications systems evolve over time, and develop their own regulatory mechanisms and ecological relations. While knowledge systems in some areas might become sub-optimal or even fade away, their operation can be explained and predicted through the use of epidemiological modeling. Some communication systems might best be left to natural evolution and survival of the fittest. Others might be improved through optimal filtering, storage, and retrieval to support the much-needed synthesis of scientific findings.

Goffman and Warren's book provides a noteworthy consolidation and extension of their separate, previous works and details numerous significant findings. They also acknowledge the parallel efforts of colleagues Eugene Garfield and Vaun Newill. Overall, this book is a synergy of two prodigious minds; here, one plus one equals far more than two, which epitomizes the old observation that wholes represent more than the sum of their parts.

7. An unfolding legacy

Bill Goffman's legacy is obviously diverse and far reaching, particularly in the areas of knowledge epidemiology, recognition of patterns in the statistical distribution of literatures and direct knowledge representations, and information storage and retrieval. He also made or helped make original contributions in disease epidemiology and helped to revitalize pre-formation theory in evolutionary biology. A 2002 historiography compiled by Eugene Garfield from his *HistCite Guide* (http://garfield.library.upenn.edu/histcomp/goffman-w_citing/index-tl.html) reveals that Goffman had been cited over 550 times by colleagues throughout the world, particularly those from the US, UK, Canada, and Germany. Scholars from France, Belgium, Hungary, India, Nigeria, Japan, Israel, and other countries also cited his works. Colleagues Ken Warren, Vaun Newill, and Tom Morris cited him often, along with former students Tefko Saracevic, Miranda Pao, Glynn Harmon and others. Such information science pioneers as Eugene Garfield, B.C. Brooks, Robert A. Fairthorne and Brian C. Vickery acknowledged his works as well.

Close to a half of the citations of Goffman's works by others relate to his formulations of knowledge epidemiology or to his original contributions in disease epidemiology itself. A quarter of the citations relate to his statistical distributions of literature collections and usage; and nearly another quarter relate to his works on information storage and retrieval. But over the years, some of Goffman's ideas appear to have become indirectly absorbed or "main-streamed" into such areas as search engine technology (particularly page-ranking algorithms and relevance rankings), citation pattern analysis, disease epidemiology, the epistemology of scientific discovery, applied mathematics, and even evolutionary biology. It might be said that over time the rivers of knowledge absorb flows from many tributaries. In this sense, Goffman's legacy is ongoing, and the formation of new knowledge epidemics can serve to revitalize that legacy. As long as information is retrieved and used, or new knowledge develops, Goffman's explanations come into play.

8. Reflections of colleagues

Don Cleveland, one of Goffman's former doctoral students, submitted the following four paragraphs, which accurately describe Bill's working relations with students and his lasting impact:

In the early fall, 1970, I sat waiting in Bill Goffman's office. I was a new doctoral student and had been directed to him for my first semester advising. To be honest, I didn't know anything about him. I was at Case Western Reserve University because I knew of the reputation of Jessie Shera and, of course, about the outstanding prestige of the School. Also, I had an excellent federal fellowship. Bill strolled in, with his rather amusing way of walking, flopped into his chair, and propped his legs on the desk. He said, "So, you are the guy from Texas. What do you want to take?" No niceties. No preliminary discussion. The question was both startling and refreshing because I came from a rigid academic background where professors told you what courses to take. They didn't ask. This model later played out with my relationship with Bill. As his student and then later as a colleague on the faculty at Case, I found that this was the way he operated. He always sincerely wanted to have input on a situation. Sometimes he went along with my ideas, sometimes we melded ideas into a workable plan and sometimes he told me that my ideas were "a crock", to quote him directly. But he always respected what I had to say and he honestly considered it. He also had a knack for trying to catch you off-guard by throwing out a suggestion that sounded good, but was fraught with unworkable tenets and then he expected you to figure out why it wouldn't work and thus trigger a new line of thinking. This, of course, came from his mathematical background of demonstrating something to be right by showing that its flipside is wrong. You had to be on guard to be sure you understood where he was coming from.

It was a challenge and a delight to work with him. Sometimes we laughed, sometimes we got ticked-off with each other, but we always finished the task and were happy with the outcome. Beneath a surface that appeared laid-back and easy going, often laced with humor, was an intense integrity for a rigorous sense of scholarship, mellowed with a genuine human concern about his students and his colleagues. We developed a friendship that lasted after I left Case. He sent me copies of his books. I send him copies of mine. After all these years, I still miss him very much (personal communication, 03/19/2007).

Don Cleveland's reflections express the sentiments of many former students and colleagues, including those of Ana Cleveland, Don's spouse, who also pursued her Ph.D. degree under Goffman.

Likewise, I recall similar feelings about him and an identical respect and admiration. When I first saw Goffman, I noted that co-workers would take their most difficult, seemingly intractable information retrieval problems to him for an opinion. He would then mull problems over, step to a blackboard, and reframe the problem mathematically. He would then say to his stymied colleagues that they were addressing either the wrong problem or a special case of a more general problem, or that they might have it right. But if he looked uninterested or indifferent about the problem, he would tend to leave the scene emotionally and attend to more interesting things—an unstated form of negative feedback that signaled to the colleagues that they should redefine or rethink the problem.

When I later would hear Goffman lecture about his research findings, particularly those about Bradford distributions or parsimonious information retrieval, I sensed that he possessed a sort of uncanny brilliance. He thrived on detecting the mysteries of hidden patterns among numbers and their interrelations, and doing so would often bring a child-like gleam to his eyes, and sometimes the subtle smile of the stereotypical mad scientist. For the first time, I developed an appreciation for the clean simplicity and power of mathematical reasoning and, as he would say, the desirability of developing “elegant” solutions. While he liked to have something in writing from others before reviewing their ideas, he did not like verbiage but wanted something terse and translatable into a mathematical ontology or equation. Nevertheless, he had a keen perception for the presence of sheer intellectual excitement in others—if anyone had stumbled across a great idea and was ebullient, he would make himself available to share that excitement.

9. Conclusion

In summary, Goffman made substantial, pioneering contributions in mathematical information science and even contributed to the areas of disease epidemiology and evolutionary biology. His initial, early efforts focused on information storage and retrieval problems and brought about better ways to frame queries and search effectively and efficiently by escaping the severe constraints of Boolean logic and through the use of multi-valued logic. He created better file organization schemes and developed methods of searching subsets of files rather than total files. He was an early advocate of indirect, citation-based searching. A fascination with Bradford-like statistical distributions led him into developing authentic core research literature collections, particularly in the biomedical areas. His subsequent efforts brought about the creative application of disease epidemiological models to the transmission and spread of ideas and knowledge through populations, which led to accurate modeling of the growth of entire knowledge systems and their respective ecological dimensions. He was later successful in understanding the dynamics of scientific discovery and the epistemological dimensions of large bodies of scientific knowledge. As one driven by curiosity, he also delved into such areas as biometry, clinical decision-making in medicine.

Bill Goffman can thus be remembered as a gifted and creative mathematician and eclectic, multi-disciplinary scientist, teacher and administrator. To his immediate colleagues, his presence was never boring and often full of surprise. He was simultaneously a mature and wise colleague, but one who exhibited humorous and playfully child-like streaks of satire, mischief, and anti-authoritarian skepticism. He always had a keen nose for the ironies of conventional life. He could be characterized as a mathematician, artist, scientist, scholar, administrator, intellectual leader, and even as a mystic who was fascinated with the implicit dynamics of nature and society. His past students and colleagues would likely all agree that his intellectual presence is indelible and remains with them to this day, and that his mathematical explorations produced durable, significant results.

Acknowledgements

The author is most grateful to Patricia Goffman, who continually provided detailed biographical information and access to Bill Goffman's writings and other related materials. Eugene Garfield originally proposed this work, provided ongoing encouragement and kindly compiled Goffman's citation data from his *Histicite Guide* and library. Likewise, Tefko Saracevic and Chistinger Tomer provided data and encouragement. Don Cleveland's keen reflections about Bill Goffman are truly appreciated.

References

- Bruer, J. T., Goffman, W., & Warren, K. S. (1981). Selective medical libraries and library networks for developing countries. *Journal of Tropical Medicine and Hygiene*, 30(5), 1133–1140.
- Church, A. (1936). A bibliography of symbolic logic. *Journal of Symbolic Logic*, 1, 121–218.
- Church, A. (1938). Additions and corrections to a bibliography of symbolic logic. *Journal of Symbolic Logic*, 3, 178–212.
- Goffman, W. (1964a). A searching procedure for information retrieval. *Information Storage and Retrieval*, 2, 73–78.
- Goffman, W. (1964b). On relevance as a measure. *Information Storage and Retrieval*, 2, 201–203.
- Goffman, W. (1965a). On the logic of information retrieval. *Information Storage and Retrieval*, 2, 217–220.
- Goffman, W. (1965b). An epidemic process in an open population. *Nature*, 205(4973), 831–832.
- Goffman, W. (1966a). Stability of epidemic processes. *Nature*, 210(5038), 786–787.
- Goffman, W. (1966b). Mathematical approach to the spread of scientific ideas—the history of mast cell research. *Nature*, 212(5061), 449–452.
- Goffman, W. (1966c). A mathematical model for describing the compatibility of infectious diseases. *Journal of Theoretical Biology*, 11, 349–361.
- Goffman, W. (1969). An indirect method of information retrieval. *Information Storage and Retrieval*, 4, 361–373.
- Goffman, W. (1970). Information science: Discipline or disappearance. *Proceedings of the 44th ASLIB annual conference* (Vol. 22, pp. 589–596). London, UK: ASLIB.
- Goffman, W. (1971). A mathematical model for analyzing the growth of a scientific discipline. *Journal of the Association for Computing Machinery*, 18(2), 173–185.
- Goffman, W., & Harmon, G. (1971). Mathematical approach to the prediction of scientific discovery. *Nature*, 229, 103–104.
- Goffman, W., & Morris, T. G. (1970). Bradford's law and library acquisitions. *Nature*, 226(5249), 922–923.
- Goffman, W., & Newill, V. A. (1964). Generalization of epidemic theory: An application to the transmission of ideas. *Nature*, 204(4955), 225–228.
- Goffman, W., & Newill, V. A. (1966). A methodology for test and evaluation of information retrieval systems. *Information Storage and Retrieval*, 3, 19–25.
- Goffman, W., & Newill, V. A. (1967). Communication and epidemic processes. *Proceedings of the Royal Society of London, Series A, Mathematical and physical sciences* (Vol. 298, pp. 316–334). London, UK: Royal Society.
- Goffman, W., Verhoeff, J., & Belzer, J. (1964). Use of meta-language in information retrieval systems. *American Documentation*, 15(1), 14–22.
- Goffman, W., & Warren, K. S. (1969). Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. *Nature*, 221(5187), 1205–1207.
- Goffman, W., & Warren, K. S. (1970). An application of the Kermack–McKendrick theory to the epidemiology of schistosomiasis. *The American Journal of Tropical Medicine and Hygiene*, 19(2), 278–283.
- Goffman, W., & Warren, K. S. (1980). *Scientific information systems and the principle of selectivity*. New York: Praeger Publishers.
- Harmon, G. (1970). *Human memory as a factor in the formation of disciplinary systems*. PhD dissertation, Case Western Reserve University, Cleveland, Ohio.
- Harmon, G. (1973). *Human memory and knowledge: A systems approach*. Westport, CT: Greenwood Press.
- Katz, M. J., & Goffman, W. (1981). Preformation of ontogenetic patterns. *Philosophy of Science*, 48(3), 438–453.
- Pao, M. L., & Goffman, W. (1990). Quality assessment of schistosomiasis literature. In L. Egghe & R. Rousseau (Eds.). *Informetrics* (Vols. 89/90). Amsterdam: Elsevier.
- Saracevic, T., & Goffman, W. (1977). Structure and behavior of subject literature as the base for forecasting in scientific communication. *International Forum for Information and Documentation*, 2(1), 17–19.
- Verhoeff, J., Goffman, W., & Belzer, J. (1961). Inefficiency of the use of Boolean functions for information retrieval systems. *Communications of the ACM*, 4(12), 557–559.
- Warren, K. S., & Goffman, W. (1972). The ecology of the medical literatures. *American Journal of the Medical Sciences*, 263(4), 267–273.

Glynn Harmon is a Professor in the School of Information at the University of Texas at Austin, TX, USA. His research interests include the modeling of scientific discovery, cognition, and biomedical informatics. He has published numerous papers on scientific discovery, information retrieval, and development of information science, and books on discipline growth patterns and information management.