Introduction

WILLIAM E. MCGRATH

WHATEVER THEORY IS, it is many things to many people. In physics and other exact sciences, the meaning of theory is well understood, with much agreement. In the arts and humanities, there are as many theories (perhaps better denoted as opinions) as individuals, with universal disagreement. In Library and Information Science (LIS), there is little formal theory to agree or disagree on. Yet there is extensive reference to theory in LIS literature, whether from a well-informed intent to place LIS on a more rigorous foundation, or from a naive effort to sound more scientific. In an extensive content analysis of 1,160 articles in six LIS journals, Pettigrew & McKechnie (2001), found that 396 "incorporated theory in either the title, abstract or text" (p. 66).

In traditional librarianship, particularly in cataloging and classification, theory was often regarded as a set of rules or a prescription established by custom and convention rather than from rigorous investigation (Smiraglia, this issue), sometimes characterized as "how we should do it" rather than "as we do do it."

Any of the following have been used as the meaning of theory: a law, hypothesis, group of hypotheses, proposition, supposition, explanation, model, assumption, conjecture, construct, edifice, structure, opinion, speculation, belief, principle, rule, point of view, generalization, scheme, or idea. Perhaps the most authoritative dictionary definition is that from the *Oxford English Dictionary,* 2nd ed, def. 4. a.:

> A scheme or system of ideas or statements held as an explanation or account of a group of facts or phenomena; a hypothesis that has been confirmed or established by observation or experiment, and is propounded or accepted as accounting for the known facts; a statement

William E. McGrath, P.O. Box 534, Westford, MA 01886
LIBRARY TRENDS, Vol. 50, No. 3, Winter 2002, pp. 309–316
© 2002 The Board of Trustees, University of Illinois

> of what are held to be the general laws, principles, or causes of something known or observed.

See also Pettigrew & McKechnie (2001) and McKechnie & Pettigrew (this issue) for references to other formal definitions.

To this author, theory is an explanation for a quantifiable phenomenon. It may be a set of relationships among variables for a fixed unit of analysis (McGrath, 1996) in which one variable may be explained by others (e.g., Kerlinger & Pedhazur, 1973 on elucidating theory with multiple regression). But Glazier & Grover (this issue) go well beyond this conventional paradigm, incorporating it and other paradigms into a more comprehensive "multidimensional" framework.

For some researchers, theory does not necessarily require more than one variable, particularly for the mathematicians and statisticians who can see predictable patterns within a variable apart from any known influence. Those patterns may be empirical, with equations fitted to curves. Or the patterns may be graphed as rank distributions such as Bradford's, Zipf's, and Lotka's, in which the theory is in the regularity of the ranking. Or the theory may be based strictly on mathematical or probabilistic data distributions such as the normal, lognormal, power functions, Poisson, negative binomial, and related distributions.

Many theories cited and used by LIS researchers originate in other disciplines. Bothamley (1993) has over 4,000 entries for theories in arts, economics, history, linguistics, philosophy, psychology, sociology, statistics, geology, physics, and mathematics. Pettigrew & McKechnie (2001) list application to LIS of many theories from the sciences, social sciences, and humanities.

The oft-heard remark that something is "only a theory" seems to suggest that theory, by definition, is something less than credible, meaningful, or valid and that only something "factual" can be believed. The papers in this issue are far more than factual. All are attempts to extract or impose meaning from highly complex phenomena from the universe of information, its processing, and use.

No attempt was made to impose any definition on the authors of this issue. And indeed its authors surely have their own.

The papers in this issue can be grouped into two categories. The first contains those papers *about* theory. These papers discuss concepts, meanings, and definitions of theory. The group also contains surveys of theory and literature reviews. The distinction between them is blurry and some may contain all of these approaches. The second group, in general, contains original research. These papers bear little resemblance to each other and all are unique. They are all difficult and require careful reading to recognize their relevance to LIS or their potential for practical application. With exceptions, papers *about* theory contain a substantial number of references, while those offering specific theories contain relative-

ly fewer references. For all papers, I have tried to indicate briefly what the "theory" means for LIS.

About Theory—Definitions, Surveys, and Reviews

Glazier & Grover attempt a broad, all-inclusive, and general definition of theory, an update on their earlier paper Grover & Glazier (1986). Whereas in their earlier paper they outlined what appeared to be a hierarchical approach to theory, from observation of phenomena to definition, concept, proposition, hypothesis, theory, paradigm, and world view, they place this scheme into a more comprehensive one they call "Circuits of Theory." The broader scheme incorporates this familiar objective and deterministic view of science into a world view where phenomena and explanation are inevitably tempered by the self, personal knowledge, social knowledge, or society in a cyclic process of change and evolution. Thus, their revised model is both objective and subjective, both deterministic and subjectivistic. Their intent is to encourage an inclusive and creative approach to research.

Smiraglia traces the history and progress of theory in knowledge organization from early rationalism based on reasoned principles and rules pertaining to cataloging and classification, to pragmatism based on observation of knowledge entities, to modern logical-positivism and qualitative methods based on empirical research. He states that no single, formal theory of knowledge organization exists. He posits, however, that (1) Lotka's Law (most names occur few times, and a few names occur many times) underlies the structure of databases, (2) Lotka's Law holds for copies, editions, translations, and other bibliographic entities, and (3) the Law holds from one collection to another, which he calls external validity. He concludes that "rationalism and historicism can help us to uncover the ineluctable truths of the natural order of knowledge entities" (p. 346).

McGrath takes the position that theory is explanatory and predictive, basically positivist. He likens the need for theory in LIS to the development and unification of fundamental forces in physics and astronomy: From Copernicus's description of solar orbits overthrowing the Ptolemaic system to Kepler's discovery of elliptical orbits, to Newton's Laws explaining the principles of gravitational attraction, to Faraday's linking of electricity and magnetism to the unification of electromagnetism with the weak force, and the current effort to link the electroweak force to the strong force and ultimately with Einstein's theory of relativity into a grand unification theory. In the far more modest field of librarianship, McGrath suggests that theories of individual functions of publishing, acquisitions, storage and preservation, structure of knowledge, library collections, and circulation can be integrated into a grand unified library theory. He then reviews recent explanatory and predictive research in each of these areas of librarianship, citing them as examples of the kind of research that could be used to build a unified theory of librarianship.

Methods used in research are critical to the building of theory. Bar-Ilan and Peritz provide an extensive survey and review of informetric methods used to study the Internet. Many of the methods they cite are from mathematical and statistical theory. An understanding of these methods and their application to the Internet, they believe, is appropriate for establishing a sound theory of the Internet. Their survey begins with literature on data collection methods. These include surveys, monitoring, and logging; crawling (retrieval of Web pages); retrieval by sampling; and exhaustive retrieval from databases, search engines, and other retrieval tools. They then review the literature of informetric methods, models, and laws used to analyze the Internet. These include citation analysis (a popular subject in informetric research in general), cocitation and coword analysis, content analysis (the method used for data collection by McKechnie and Pettigrew in this issue), evaluation using existing and new methods, identifying and calculating indicators (Web impact factor, or WIF, for example), and various models (hubs and authorities, for example, as well as fractals). They continue with a review of fitting models to the literature of scientific topics; a review of power laws and Zipf-type laws, both common in the general informetric literature but here applied to the Internet; and finally a review of the literature of obsolescence applied to the Web where documents are changed, removed, or relocated. They provide a summary table of characteristics and measurements of the informetric literature reviewed according to various categories.

In mathematics and statistics, "theory" is often used to describe a group of procedures or tools that otherwise would be thought of as "method." Examples are probability theory, game theory, information theory, chaos theory, queuing theory, catastrophe theory. They are methods used to model certain kinds of data. Hood and Wilson note, that "Some aspect of the real world may be modeled by a mathematical theory." "How useful this is," they continue, "depends on how well the mathematical model captures the essence of the reality" (p. 394). The better the fit, the better the model or theory. They review the literature of one such method that has been used in LIS, Fuzzy Set Theory (FST), a tool employed to analyze data that do not fall readily into discrete categories. FST can be used in information retrieval where, for example, relevance in a "set of relevant documents" is fuzzy rather than dichotomous as in the traditional Boolean approach. Other examples are when and if to bind periodicals, when and if to insert detection strips in periodicals, expert systems, document retrieval, relational databases, thesauri and catalogs. They cite additional literature of applications in LIS as well as the literature of the theory itself. They conclude that, despite its theoretical appeal, FST has not yet found widespread application in LIS.

McKechnie and Pettigrew, continuing their earlier work (Pettigrew & McKechnie, 2001), cross tabulate applications of theory published in six LIS journals. Their tabulations include topics in humanities, social sciences, or

science, as well as by affiliation of author (private sector, government, LIS, humanities, social sciences, sciences), type of article (descriptive, empirical research, historical, modeling argument, review, method, theory), and sources of theory (from LIS, humanities, social sciences, sciences). The result is a useful picture (counts and percentages) of how and in what context theory is used in the current literature of LIS. They discuss the implications of differences in the number of theory articles associated with humanities, social sciences, and science, and the "surprising" finding that "many non-LIS scholars are publishing in LIS journals" and the "disappointing" finding that "LIS theories had not made substantial inroads in other disciplines" (p. 414). The large number of theories found in the surveyed articles, they conclude, is enriching LIS, but they encourage authors to list primary sources of theory, and to provide better explanations of theory and how it has been used.

For periodical collections in libraries, there is hardly an issue more critical than deciding which titles to keep, which to buy, and which to terminate in times of budget restraints. Evaluation of periodical titles is a difficult process. While many methods of evaluation have been used, most libraries probably still use old-fashioned rules of thumb and subjective criteria. More often than not, a journal's subscription price will dictate a decision. On the other hand, information scientists have devised a number of mathematical methods or indicators based on usage and citations to journal titles. Perhaps the best-known and most widely studied indicator is the impact factor (ratio of citations to articles published) and its variations. Rousseau, in his article, discusses several methods for evaluation, focusing on the mathematical issues associated with the calculation of impact factors and related measures such as the immediacy index (how quickly a journal is cited after publication). Rather than expressing these indicators in terms of formal theory, he confines theoretical issues to their "precise [mathematical] formulation," with no input-output or explanatory model. This approach, he suggests, may help fellow scientists to construct just such an overall model.

Original Theories

Each of the following papers presents a unique and original theory. Nothing ties them together in any thematic way, although there are some coincidences. They *are* theory rather than *about* theory.

An interesting phenomenon found in the use of scholarly and scientific literature that has implications for libraries is the Matthew Effect (first discovered by Robert Merton, the well-known sociologist of science), named after the Biblical passage in St. Matthew. According to this theory, a large number of citations to scientists' publications generate even more citations to the detriment of scientists who receive few citations. The effect also holds for individual articles as well as for journals. Bonitz, continuing earlier work

on the Matthew Effect, examined a large number of journal citations and found that the effect also holds for countries, the MEC. He cites a clarification of the Matthew Effect as not "the rich become richer and the poor poorer," but "the employment of your given talents is rewarded and their neglect punished" (p. 444). He then describes two aspects of the MEC. First, ranking of nations based on the effect can be regarded as a measure of the overall efficiency of scientific performance of a country. Second, the effect is concentrated in a small number of journals he calls Matthew core journals. Scientists should endeavor to publish in those journals, and libraries would profit by holding them.

What has coauthorship to do with librarianship other than to record it in catalogs and indexes? Glänzel studied trends in coauthorship for the period 1980–1998 in three scientific fields, biomedical research, chemistry, and mathematics. Using data from the Science Citation Index, he found that (1) coauthorship of individual papers has increased, but declined to offer a theoretical explanation for the increase; (2) the theory of coauthorship affecting individual author productivity was not supported; (3) the theory that coauthored papers are cited more frequently than single-authored papers was strongly supported. (The latter relationship apparently holds only for individual papers and not the journals in which they are published.) Glänzel suggested that coauthorship should be added to bibliographic coupling, coword, and cocitation analysis when studying the network of science communication and in the design of information retrieval strategies, particularly for identifying core documents. Glänzel refers to these relationships as theories supported or not supported by the findings, which suggests that further research may or may not support these theories.

In another study on coauthorship, Kretschmer applies gestalt theory from psychology to the similarities and dissimilarities of authors to each other based on counts of the number of papers coauthored. (The method used in gestalt theory may be an alternative to cluster analysis and multidimensional scaling.) The more papers two scientists coauthor, the more similar they are to each other in their research concentration. Conversely, the fewer they coauthor, the more dissimilar to each other. Ketschmer likens similarity to "birds of a feather flock together" and dissimilarity to "opposites attract," and to the Yin and Yang of Chinese philosophy. Readers familiar with gestalt theory should have no difficulty with this paper. Readers unfamiliar with it will be helped by the definition of a gestalt as a holistic configuration of the parts of a relationship. Each gestalt can be graphed as a 3-dimensional array of coauthor relationships. Though the interrelationships may vary, they can always be represented in a single holistic graph that, when stable, exemplifies the conciseness principle. This principle could be used in the design of search algorithms in databases, as in Glänzel's paper.

Moed, Luwel, and Nederhof develop a general framework in which librarians can be called upon as bibliometric professionals for the collec-

tion of data regarding research performance and productivity in the humanities and social sciences, the dynamics of which differ significantly from the sciences. The authors develop a methodology for performance indicators, with a review of earlier studies, then provide an example from the field of Belgian Law with a critical discussion of the methodology. The framework was based on results obtained from surveying Flemish scholars regarding their perceptions of the quality or importance of books, theses, journal articles, reports, lectures, and other items of productivity to be used in judging research output. The framework contains as many as eighteen categories, but is considered preliminary, needing operationalization. It may also be regarded as structure and therefore theoretical.

Tsai develops an elaborate theory of information-generating forces and subforces and, like McGrath (this issue), invokes the vocabulary of physics but there the similarity ends. He adopts a genetic metaphor in describing the sub-forces of query, command, statement, and term-term bond denoted as Q-C-S-T or Q-T-S-C chaining which can, apparently, interact with each other in any sequence inseparably but whose symmetry can be broken, producing a need for change or readjustment. The model can be portrayed in 3-dimensional graphs reminiscent of Kretschmer's gestalts (see above), suggesting an unrecognized connection. The broken symmetry can be restored and recombined as in a Möbius strip, or re-sequenced by "clip-jointing," a simulation of cocitation. Multiple Möbius twists and repeated clip-jointing results in genetic sequencing while the original configuration is lost and must be.reconstructed by information specialists. Tsai follows with a Fuzzy Commonality Model (FCM) for describing data generated by the Q-C-S-T process. In still another analogy, he likens libraries to the Q-C-S-T process, with library administration as the command center (C), technical services providing statements (S), public services for user queries (Q), and publication of library services for term-term bonding (T). (This analogy could be regarded a component of or even an alternative to McGrath's unified theory approach.) In the final section of his paper, Tsai describes software written to analyze, mine, map, and repackage information, and lastly he describes its application to a total quality knowledge management (TQKM) system.

Modeling the growth of science has been a preoccupation of information scientists ever since the publication of Price's *Little Science, Big Science* (1963). A large literature seems to have settled on a standard model of either linear or exponential growth. Vinkler, in his paper, asserts that models based on annual cumulative or relative growth can only give a simplified picture. These models seem to hold only for short periods. "*There is no general law 'governing' the publication growth* of disciplines for longer periods," Vinkler asserts (p. 555), emphasis original). Instead, he offers the "Institutionalization of Scientific Information Model" which "integrates the production, evaluation, modification, and aging of processes of scientific in-

formation" over time through "various evaluation and modification processes toward a cognitive consensus of distinguished authors" (p. 557) The model is elaborate and detailed and is based on generated information, its evaluation, and its impact. Vinkler's intent is to model the institutionalization of science disciplines by this process. A discipline's information is fully institutionalized when it becomes common scientific knowledge. References (citations) are regarded as proof of impact when assessing research results and when making library subscription decisions.

The contents of these papers are far richer than what has been summarized here. Some require careful and patient reading to comprehend them, but the effort is rewarding for those who try.

Originally, this author had envisaged a collection of theoretical essays more representative of the broader aspects of LIS. What has been achieved is a collection of worthy papers, an international representation, albeit of narrower scope. Still needed is a deeper understanding of theory and the fundamental sociologic forces driving LIS and a volume of literature to elucidate this need.

References

Bothamley, J. (1993). *Dictionary of Theories.* London: Gale Research International.

Grover R., & Glazier, J. D. (1986). A conceptual framework for theory building in library and information science. *Library and Information Science Research, 8*(3), 227–242.

Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple Regression in Behavioral Research.* New York: Holt, Rinehart, and Winston.

McGrath, W. E. (1996). The unit of analysis (objects of study) in bibliometrics and scientometrics. *Scientometrics, 35*(2), 257–264.

Pettigrew, K. E., & McKechnie, L. (E. F.). (2001). The use of theory in information science research. *Journal of the American Society for Information Science and Technology, 52*(1), 62–73.

Price, D. J. de Solla. (1963). *Little Science, Big Science.* New York: Columbia University Press.