Explanation and Prediction: Building a Unified Theory of Librarianship, Concept and Review*

WILLIAM E. MCGRATH

ABSTRACT

AS INSPIRATION FOR DEVELOPING A comprehensive, unified, explanatory theory of librarianship, the author makes an analogy to the unification of the fundamental forces of nature, beginning with the Copernican revolution, followed by the discoveries of Kepler, Galileo, Newton, and Einstein, and the unification of electro-magnetism, light, the weak force, the electroweak force, the strong force, and the ultimate goal to include gravity, space, time, and relativity into a single grand unified theory. While the analogy may be naïve and debatable, the linking of disparate domains suggests a process for linking the broad and classical functions of librarianship into a framework for a unified theory. The unified theory might consist of functions stemming from the world of publishing: Selection and deselection, acquisitions, the structure of knowledge and classification, storage and preservation, the library collection, and circulation. The author reviews recent Library and Information Science (LIS) research of the type that could contribute to development of unified theory. Dependent and independent variables are identified when apparent, with particular emphasis on the importance of units of analysis to theory. The recent literature is dominated throughout the framework by studies involving library circulation or its surrogates.

COPERNICAN REVOLUTION

When Copernicus showed that the known planets orbited the sun, not the earth, he began a centuries-long process of linking the fundamental forces of nature. His revolutionary theory changed the course of astronomy because it explained the movements of the planets far better than the

William E. McGrath, P.O. Box 534, Westford, MA 01886
LIBRARY TRENDS, Vol. 50, No. 3, Winter 2002, pp. 350–370

orthodox Ptolemaic system did. It was advocated by Galileo, augmented by Kepler's discovery of elliptical orbits, explained by Newton's laws of gravity, and ultimately refined by Einstein's general theory of relativity.[1]

The genius of the Copernican-Galilean-Keplerian-Newtonian achievement, or "celestial mechanics," as it is now called, is in its extraordinary ability to explain and predict. The movements of the planets, moons, comets, and other bodies can be explained in terms of gravitational force and the conic sections of classical geometry—the ellipse, parabola, and hyperbola—and their exact positions relative to each other can be predicted with great accuracy.

Similarly, the power to explain and predict also improved with the nineteenth-century reconciliation of electricity and magnetism by Michael Faraday, with light by James Clerk Maxwell, and more recently with the fundamental "weak" force, to form the "electroweak" theory. Current efforts are aimed at reconciliation of the "electroweak" force with the "strong" force and, ultimately, with gravity and general relativity to form a "super unified" theory incorporating all of the fundamental forces of nature (Ferris, 1991). Hannaford (1980), in his discussion of libraries and scientific knowledge, refers to this as the hierarchical picture of explanation, "General relativity explains special relativity explains Newtonian mechanics explains observations of planetary motions" (p. 577).

Phenomena of Librarianship

What is the implication of these great achievements for libraries—apart from being repositories for the precious documents describing them? As scholars and social scientists in our own much humbler yet somewhat pretentious sphere, can we formulate theories to explain the various interacting forces of librarianship that would enable us to predict those phenomena? The answer is "perhaps," because such application is mostly by analogy, and the analogy is more inspirational than emulative. After all, library science is not natural science. Human behavior, far more complex than planetary motions, can never be described or predicted with the precision of celestial mechanics. But we should like to try, even though our theories may never be elegant or exact.

For this discussion, an informal and simple (some would say simplistic) definition of theory can be used: A set of variables that may explain and predict another variable. A unified theory is simply one that reconciles or incorporates other theories. For a more formal definition of theory in the context of librarianship, refer to the taxonomy of theory by Grover & Glazier (1986) and their broader update of the taxonomy in this issue (Glazier & Grover, 2002).

Consider some of the traditional areas of concern to librarianship: Publishing, acquisitions, storage, preservation, classification and organization of knowledge, and collections and circulation. While not necessarily

complete, few question these as basic to the profession. Most recently, Curran (2001) has reconfirmed them as those aspects of information that library and information scientists are most concerned with, adding origin, dissemination, properties, retrieval, and interpretation of information. No doubt, this list could be even further refined or expanded. Curran offers many questions pertaining to each area, the answers and alternatives to which the profession should continue to seek. His questions (how, who, what) are all valid when attempting to describe activities. We, those in the Library and Information Science (LIS) profession, should like to have a more precise, perhaps mathematical understanding of how these areas are interconnected, and how the activities or outcomes of each may be explained or predicted in terms of inputs from others. While recognizing that Curran and others may prefer the more detailed outline or one altogether different, this paper is confined to the more limited one. However, whatever the framework, it is important to note that there is a sense of flow or connectivity from one domain to another, just as there is in everyday practical processing and use of library materials.

Within the context of these activities, but beyond their mere description, what do we mean by "explanation" and "prediction?" What do we want to explain and what is there to predict in librarianship, and why should we want to predict it? One definition of explanation—a much more complex concept than can be explored here—is simply accounting for one phenomenon in terms of others. A good explanation is one that provides understanding. More specifically, it is one that, given a set of conditions, enables us to predict another with reasonable or satisfying confidence.

In every area—acquisitions, storage, preservation, classification of knowledge, collections, reference work, and so on—there is something that varies and is dependent on something else, so that we should be interested in building theories that would enable us to explain and predict those things that vary in each area. Intuitively, we know that each area is, to some extent, dependent on some other, either directly, in a linear flow, beginning to end, or in a more complex multidimensional way, in which communication, or the workflow, may take many paths.

Consider some typical activities in each of the functional areas listed above—how they might be explained by some other area, how specific theories could be built for each area, and then finally how they might ultimately be integrated into a unified theory. Figure 1, modeled after a diagram, "Explaining the Forces of Nature," published in the *New York Times* (Broad, 1984), and reproduced in McGrath (1995b, see note 1), shows these traditional areas of librarianship with hypothetical connections (dotted lines) between them to indicate relationships not firmly established in any explanatory or predictive sense.

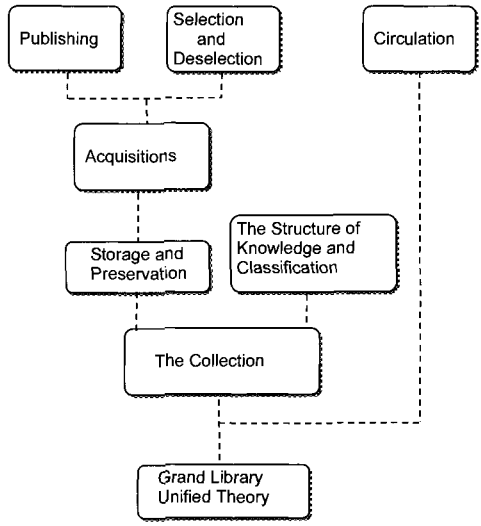

Figure 1. A Unified Theory of the Library. Dotted lines indicate hypothesized links.

PUBLISHING

To some extent, librarians want to know what societal factors contribute to the variability of publishing each year: Demand, world events, economic conditions, and so on. Knowledge of those factors is necessary to construct a theory of publishing.[2] Though such a theory is of interest for understanding the bigger picture, librarians accept information from the world of publishing as input to their considerations—the population of books, journals, and other materials, or portion thereof, to be acquired. Whereas publishing is the output of societal motivation and conditions, it is input to a theory of acquisitions.

ACQUISITIONS

Publishing is a necessary condition for acquisitions to take place. Collection-building cannot occur unless there are published items to collect.[3] The question, therefore, is "What are the conditions and criteria for selecting or not selecting specific books to add to the collection?" All of these conditions and criteria may be quantified in such a way that their affect on the number of items selected can be tested.

A proposed theory of collection building should consider—that is, should test—variables associated with publishing, selection, and censorship, as well as a host of other variables, including the education and knowledge of the selectors, the academic environment, nature of the community, size of the budget, and the required subject areas. The theory would include only those variables that significantly contribute to the variation of selec-

tion and collection building—that is, only those variables that hold up under testing. Even then, it is still a theory, because it is in the nature of science that an old theory can be modified, overthrown, or displaced, and that is certainly true in our context.

Classification

The classification scheme used by the library is a major property of the collection. The scheme reflects the librarians' perceptions of how knowledge is organized or structured. The idea of structure comes closest to our cosmological analogy: Just as there have been many theories on how astronomical bodies relate to one another in an organized system, so also have there been many classification systems. And just as some of those cosmological theories, such as the Ptolemaic system, failed in their ability to predict, so have our classification systems failed to optimize accessibility. Just as the Ptolemaic system was taken on authority for fourteen centuries or more until Copernicus put it to a test, so have librarians taken most classification systems on authority and rarely, if ever, put them to the test of predictability. Human systems can never be deterministic in the sense that, for example, orbiting bodies depend on the force of gravity. Because society is mutable, no classification theory can ever be enduring. Nevertheless, we can still look for structure in knowledge. And even though structure may not be permanent, principles are permanent and are reason enough to look for more enduring structure. Buckland, in defining theory, says that "structure *is* theory" (1988, p. 37). From that, it follows that classification and the structure of knowledge is necessary for the development of a unified library theory.

The structure of knowledge is due in no small part to what is published. For any given library, it may depend on the portion of published knowledge acquired. It may also be due to other variables in the local environment—including demand, the nature of the community, and the library's users.

Storage and Preservation

Storage and preservation are major functional concerns of every library. Storage problems involve available square footage, linear stack space, stack maintenance, retrieval and reshelving of materials, scheduling and training of stack personnel, shelf-reading, inventory, and much else. Preservation comprises the condition of materials, environmental questions, humidity, chemicals, temperature, lighting, acidity of paper, dust, protection against fire and moisture, and so on. All of these variables can be quantified, controlled, or otherwise described and are important properties of the library collection.

The Collection

Now we can see that a description of the library's collection must include everything discussed up to this point: Publishing, acquisitions (which

entails selection), the classification scheme (based on some perceived structure of knowledge), and the problems of storage and preservation. How these components fit together to make a theory of collections seems obvious and trivial, but what may seem obvious may be merely a reflection of what we actually do in practice, the current way of doing things—which may not necessarily be the best way. After all, the Ptolemaic system, which Copernicus and Galileo showed to be wrong, was able to predict planetary positions with some success. Perhaps some components, such as classification, may be based on a coherent theory, but unless the theory includes all components it is not complete. Ideally, all of the variables and all of the components must be described, quantified, tested, and retested as a complete system before we should be satisfied. Hannaford (1980), equating "theoretical" and "scientific," believes that collection development can be scientific. In two earlier papers, McGrath discussed the theory of collections in terms of the relationship between circulation and collections and the units in which data could be collected (McGrath, 1980), and in terms of the relationship between parts of collections, who uses them, and between other collections (McGrath, 1985).

Circulation

Let us now look at circulation, perhaps the ultimate first and last reason for the very existence of the modern library. Success of the library depends on its circulation.[4] Conversely, circulation depends on the library and its collection, classification, and organization of materials.

The high volume of circulation requires that library administrators maintain appropriate records, reshelve returned books promptly, keep bookshelves orderly, and so on. But library circulation varies from hour to hour and day to day. The library administrator would like to be able to anticipate (to predict) this variation in order to allocate sufficient funds to pay shelvers and to schedule them when needed. If the conditions or variables that make circulation fluctuate were known, the administrator could provide better service. What makes circulation fluctuate? We do not know until we can test the variables we think may be correlated with circulation—that is, by formulating a theory of circulation and then testing it.

Circulation may be dependent on variables both internal and external to the library. In either case, we should like to know what they are. If internal, then we would need to examine all functional areas, such as acquisitions and cataloging, for conditions that might make circulation fluctuate. If external, then circulation becomes part of a larger sociological theory.

Grand Unified Library Theory

In a very broad and nonspecific framework, this essay outlines one possible approach to the development of a grand unified library theory in which the library is an integrated system where outcomes are describable

in terms of measurable relationships, regularities, and laws. The work required to uncover these relationships—the work of intellectual design and computation—might be prodigious and challenging, but the computations should be relatively trivial once the design is formulated.

The unified theory is sketched only in the broadest and briefest outline. There is much not addressed—the psychology of users and librarians, attitudinal studies, organizational behavior, interaction with other disciplines, scientometrics and informetrics, individual scholarly productivity, citation analysis, LIS education, welfare and status of librarians (including tenure, salaries, and prestige), and so on. To some critics, the most glaring omission might be inattention to the digital revolution. To this author, however, while the production of electronic databases, the World Wide Web, and the Internet is technology, their use can be described in terms of traditional library functions.

At a more mundane level, the need for bridging domains, whether it is called unified theory or something else, is recognized by the familiar sardonic complaint in libraries that acquisitions librarians do not talk with catalogers, who do not talk with reference librarians, who do not talk with circulation librarians, and so on. "No one talks with anybody," yet the need for reconciliation, cooperation, and system integration is obvious and incontrovertible. A unified theory might provide the basis and incentive both for understanding and quantifying the flow of materials between the domains and for establishing firmer communication as well.

Methods

Modern mathematical and computational tools, far more powerful than the pencil and paper used by Copernicus and Kepler three centuries and more ago, can measure the relationship between output or dependent variables and input or independent variables. Probabilistic statistical tools, such as canonical correlation, discriminant analysis, path analysis, the general linear model, multiple regression, and analysis of variance, are routinely used for testing and building theories in many scientific domains. The general idea of these tools is that they allow us to account for the variance in the dependent variables in terms of the variance of independent variables.

Other methods may be used to describe the inevitable cyclic nature of information access. After all, the Laws of Newton and Kepler were derived from pure and accurate description of orbital motion and were held to be precise and deterministic. Mathematical tools, such as time series and spectral analysis—fundamental to the understanding of celestial signals and orbital mechanics—can be applied to these cycles (McGrath, 1996a).

Building a theory, of course, entails much more than application of quantitative methods. An understanding of the entire process is essential. Scriven (1968) provides just such an understanding in an essay on the concerns of science: Observation, description, definition, classification, measure-

ment, experimentation, generalization, explanation, prediction, evaluation, and control of the environment. McGrath (1986) showed how these concerns might apply to research in LIS as a coherent and continuous process.

A Review of Recent Explanatory Research that Could Contribute to a Unified Theory

Theory in LIS is something more than just an esoteric and abstract realm out of touch with the practical problems of day-to-day professional work, as may be inferred from the extensive review by Pettigrew & McKechnie (2001). They found that of 1,160 articles in six LIS journals for the years 1993 to 1998, 397 discussed or employed theory while characterizing "the vast majority of information science" since 1950 as "atheoretical." Earlier surveys reached similar conclusions (Peritz, 1980). Nevertheless, these reviews show that, despite the failure of much research to address theory, there is considerable recognition among grass-roots researchers that theory would help to strengthen our understanding of LIS relationships.

Following is a brief review of recent papers that exemplify the sort that can contribute to theory in each of the traditional categories outlined above. The journals are replete with studies of library and information activity, but relatively little—as Pettigrew & McKechnie and others have found— cast in theory, and less that lend themselves to theory building. No attempt was made here to review all of the literature that might otherwise be considered relevant. In particular, the vast literature on digital libraries and online retrieval is left to other reviewers (e.g., see Bar-Ilan & Peritz, 2002).

There is much literature on the philosophy of LIS containing provocative and stimulating ideas that always seem on the verge of offering testable theories or of challenging empirical researchers to operationalize abstract themes. One such piece is the comprehensive and thoughtful treatise on metalibrarianship by Nitecki (1993), a tour de force, in which he explores not only the interdisciplinary character of librarianship, but the "relationships between the essential, minimal and basic elements in the communication of any recorded data, information, or knowledge" (Part 1, p. 2). In Chapter 11, "The Theory of Metalibrarianship," Nitecki explores theory, metatheory, methodology, evolution of concepts, the "multiplicity of metalibrary relations," and other ideas detailing a relational approach to librarianship.

A paper by Zwadlo (1997) similarly challenges LIS to apply "philosophical" ideas to "useful things." Many more such papers can be found in both Nitecki's and Zwadlo's citations, as well as in others. However, as interesting as it might be, unless the philosophy of librarianship tells us how to develop an explanatory theory of librarianship, it has limited value to this review.

Criteria for inclusion in this review are papers published (approximately) within the last ten years that include (1) the use of quantitative methods, such as multiple regression and the analysis of variance, that enable

researchers to test independent variables that might account for variance in dependent variables or (2) correlation methods applied to two or more variables for which dependence or independence may or may not be identified by the researcher but which are potentially one or the other or (3) studies that do not necessarily apply quantitative methods, but express a research hypothesis or objective or model that may ultimately be tested by quantitative methods and thus have the potential for building theory.

An enormous number of studies have been devoted to frequency distributions of single variables. While these are always highly mathematical and interesting and theoretical, and while there are examples even among the papers in this issue of *Library Trends* (e.g., Rousseau), their authors are mostly concerned with the ability of a frequency distribution to forecast itself. These distributions are theoretical in that researchers attempt to fit a model to actual data. They are often highly successful and accurate, but are limited in their application to explanatory theory. The relationships between them and other variables are rarely analyzed. Other than to note their importance when considering normality and homogeneity of variance, important properties of distributions used in parametric testing, they contribute very little to the explanatory relationships of concern to this review. With a few exceptions, that genre is not included among the studies reviewed herein.

Many other interesting studies, some that used an explanatory approach with dependent and independent variables, were excluded from this review because they were outside of its main thread or failed to find significant relationships. Attitudinal studies, user satisfaction studies, and psychological studies in general were excluded, as were studies on librarians' status, job satisfaction, and salaries. Thus, there is a bias toward what libraries, librarians, and users do instead of what they think or feel.

Papers about citation theory, except where citations correlated with other relevant variables, have been excluded. The literature of citation theory focuses primarily on the communication relationships among scholars and scientists or between and among disciplines—highly interesting but of indirect interest here.

There may well be studies that could have been included—papers published in the seventies, eighties, and earlier, for example. However, the purpose of this paper is not to provide an exhaustive review of all possible relevant papers or a history of theory development, but rather to provide examples of recent papers that might help to build theory.

The following studies, then, are illustrative of types that have the potential for building a comprehensive, grand theory. One could call these studies "normal" science after Kuhn (1962)—filling in the gaps of existing theory—except that existing theory is much more elemental or primitive, and LIS has far to go to build good explanatory theory.

Dependent Variables, Independent Variables, and Units of Analysis

When apparent, the author has tried to list the dependent variable, significant independent variables, and the units of analysis (the things described by variables) for each paper reviewed. Whereas the meaning and importance of dependent and independent variables in theory is understood by most researchers, the importance of units of analysis in research design is not always appreciated. Understanding the unit of analysis is crucial in building theory (McGrath, 1996b). The difference between variables and units of analysis can be quite confusing. A variable at one level, for example, might be a unit of analysis at another level. In some studies, the units of analysis were not always specified by their authors and had to be inferred.

Saxton (1997), using meta-analysis to evaluate consistency of findings and standards for reporting findings across independent studies—in this case, correlations with accuracy of reference service—makes several important observations, one of which is also critical to the development of theory. "Studies cannot be compared," he says, "if they use different units of analysis (for example, libraries, librarians, reference transactions)" (p. 282). McGrath (1996b) also makes this argument but adds that, in the development of a unified theory, different units of analysis can be related to each other at different levels. For example, *number of libraries* can be a variable in a study where country is the unit of analysis, whereas in another study, *number of books held by a library* may be a variable, while library would be the unit of analysis.

For each study, where identifiable, the independent variables are *italicized,* the dependent variables are in uppercase and, when not otherwise indicated, the units of analysis are followed by the abbreviation "u.a." in parentheses. Thus, in a study using demographic variables to predict the number of books checked out by users of a library, independent variables are, for example, *age, sex, marital status, educational level;* the dependent variable is NUMBER OF BOOKS CHECKED OUT; and the units of analysis are library users (u.a.).

Publishing

It has been said that, as pharmacies are the dispensers of the drug industry's productivity, so are libraries the dispensers of the publishing industry and scholarly output. Such a limited perspective interferes, perhaps, with our ability to perceive the whole world of knowledge and to understand how best to use it. Much research can be found on the commercial and marketing aspects of publishing, but other than pricing and availability, relatively little—in the explanatory sense—on the interaction with libraries. Not reviewed here are the multitudinous studies on factors affecting the productivity of individual faculty, scientists, and scholars in general.

Petersen (1992), using multiple regression to find the most significant correlates of journal (u.a.) PRICES, found that *for-profit publishers, those originating in Europe,* and the journal's *impact factor* were the best determinants.

Chressanthis & Chressanthis (1994), also using regression analysis, found that *the exchange rate between countries, the existence of illustrations, the number of pages, a composite of citation measures, journal age, economies of scale created by higher circulation,* and *the existence of "nonprofit motivation of publishers"* all have an effect on journal (u.a.) PRICES.

Kishida and Matsui (1997) developed a regression model in which they found that *population* and the *number of people attaining a university education* best explained THE NUMBER OF MONOGRAPHS PUBLISHED in each country (u.a.).

Quandt (1996) used an iterative simulation model to describe the evolution of library subscriptions in which cancellations inevitably cause publishers to raise their prices. Though not about the determinants of price nor the number of subscriptions, his article may be helpful in designing such a study in two respects: One in which *price* and *profit* are predictors of THE NUMBER OF LIBRARY SUBSCRIPTIONS to journals (u.a.); and another in which *cost* and *importance* to libraries are predictors of library (u.a.) decisions to SUBSCRIBE OR NOT. Thus, his article is an example of bridging more than one level of our theoretical context: Publishing and acquisitions.

Acquisitions (Book and Journal Selection)

Whereas several studies on the predictors of price as a dependent variable were cited in the section on publishing above, *price* becomes an independent (determining) variable when considering the purchase or deselection of books and journals. For example, McCain (1992) found that *price,* as well as *mathematical content* and *cocitation rate,* were significant predictors of THE NUMBER OF LIBRARIES HOLDING economics journals (u.a.). *Longevity* and *cocitation rate* were significant predictors in genetics.

Shaw (1991) found a significant correlation between the *number of reviews* of BOOKS and the *number of libraries holding them.* Likewise, Serebnick (1992) found a significant relationship between the *number of reviews* of book titles (u.a.) and THE NUMBER OF LIBRARIES HOLDING THEM. Similarly, in a sampling of books (u.a.) reviewed in *Choice* magazine, Calhoun (1998) found a positive correlation between the *number of reviewed books appearing in vendor approval plans* and *those books subsequently purchased by libraries.* Either of these could be regarded as the dependent variable, but were not so indicated in the study.

Kreider (1999) found a significant correlation between *local citation frequency* and the *global citation frequency* of journals (u.a.) appearing in Journal Citation Reports (JCR), suggesting that libraries should consider JCR data when evaluating their journal collections. Either local or global citation frequency could be regarded as the dependent variable, depending on purpose.

Tsay (1998) found significant correlations between *frequency of journal use* and *citation frequency* and between *frequency of use* and *impact factor* for some medical disciplines (u.a.) but not for others. To comment, since Tsay

did not indicate which comes first, *circulation* or *citation* (that is, which is dependent and which independent), librarians could use published *citation* data to predict CIRCULATION when selecting and, conversely, *circulation* to predict CITATION when deselecting holdings.

Crotts (1999) "develops" a model for allocating monograph budgets to SUBJECT AREAS based on *circulation.* Budget allocation for subject areas is an issue for which there is voluminous literature and many reviews going back to the seventies and eighties and earlier. His paper is cited simply to document the continued interest in and timelessness of a classic model, as an example of bridging the two domains (collection development and circulation), and as continued potential for further development of theory. His design was not conceptualized in terms of dependent or independent variables, although it is reasonable to regard CIRCULATION as the dependent variable and subject areas (u.a.) as the units of analysis. For earlier contribution of circulation to theory and collection development, see McGrath (1980, 1985).

Classification and Organization of Knowledge

Few recent explanatory studies on classification and organization of knowledge were found. This does not necessarily suggest a research oversight, because there is a great deal of literature, including whole journals, devoted to classification and organization of knowledge. Nevertheless, there does seem to be a gap in the explanatory literature of classification.

Satija (2000) provides numerous definitions—one of Scriven's (1968) concerns—about classification concepts that would be necessary when operationalizing hypotheses in an explanatory design.

Leazer & Smiraglia (1999) perform a qualitative analysis of "bibliographic families," families of related works in the library catalog, intended to produce an explanation of some pattern [a dependent variable?]. Cataloger-generated maps of these families, they conclude, are inadequate to explain the pattern, and user behavior studies are needed to determine which maps are preferable. Smiraglia discusses the need for explanatory studies in this issue of *Library Trends.*

Losee (1993), in a study on the influence of classification and location on circulation, used a regression approach to predict the AVERAGE NUMBER OF BOOKS a patron (u.a.) circulates from the *relative location of books, relationships among the number of areas in which books are found* (measured by the number of stops a patron makes when browsing), and *the distances across a cluster.* Patrons made more stops than books found at a stop.

Rodman (2000) discusses the connection between call numbers and browsability on the shelf, or in an online catalog, when call numbers are not changed to fit into shelf list sequence. Though not an explanatory study, it does suggest a design in which "*number of screens between like items*" in an online catalog could be regressed against the NUMBER OF HITS during a search session (u.a.) or time period (u.a.).

Storage, Preservation, and Collection Management

Storage, preservation, and collection management are crucial components of a comprehensive theory, obviously because existence of a library collection (whether hard copy or digital) is a necessary condition for its use or circulation.

As a means for identifying low-demand titles (u.a.) for remote storage, Silverstein & Shieber (1996) looked at individual titles (u.a.) to see HOW MANY TIMES THEY CIRCULATED (0, 1, 2, . . . , *n* times circulated). They concluded that "past use [is still] the best single predictor of future use" (p. 289). Though theirs was a frequency distribution study, not intended as explanatory, their data might be submitted to explanatory methods, such as analysis of variance. Independent variables were categorical: *Last use, Library of Congress classification, publication date, language,* and *country.*

For a similar purpose, Hayes (1992) fit an exponential J-curve equation to book (u.a.) circulation frequency data, and developed a cost-allocation model to levels of access and storage. The units of analysis were books (u.a.). The dependent variable was, variously, CIRCULATION and IN-HOUSE USE. As with Silverstein & Shieber, his study was not intended to be explanatory.

Lee (1993) addresses the problem of storage space, citing past research on remote storage, weeding, and rarely used material as possible solutions. As an aid to determine the most economic approach to storage, Lee proposes a model that incorporates both prediction of DEMAND and cost analysis into a single model.

Two surveys of book deterioration (Bennett, 1992; O'Neill & Boomgaarden, 1995) were not in themselves explanatory studies but may be helpful in defining variables such as the *brittleness* and *acidity* of book paper (u.a.) and other conditions that may be helpful in eventual correlation with other components of a unified theory.

Collections

What the library collection contains and how it is organized and used is an essential component of a unified theory (McGrath, 1985).

Exon & Punch (1997), replicating a 1981 study, tested the assumption of self-sufficient library collections by performing a correlation analysis between *collection size* of a library (u.a.) and *interlibrary loan requests* of other libraries. From the strong positive correlation found, they conclude that self-sufficiency is a fallacy. This can be interpreted to mean that libraries need each other and that their interdependence may be incorporated into the larger theory. In an explanatory study, NUMBER OF INTERLIBRARY LOAN REQUESTS could be the dependent variable.

Circulation (Includes Catalog Access, Online Access, and Reference Service)

Circulation and usage may be the most studied function in libraries. The literature is voluminous, going back many decades, and has been ex-

tensively reviewed by many authors. These reviews can readily be found in the literature.

A perennial question is whether *in-house use* can be employed as a measure of circulation—that is, *books officially charged out.* The unit of analysis may be some unit of time, such as day, week, or month; or some other unit, such as subject or discipline; or type of material, such as book or journal.

Blecic (1999), investigating journal (u.a.) use in a medical library, found a significant correlation between *in-house use* of journals and their *circulation,* as well as between those two variables and journal (u.a.) *citation by faculty.* Similarly Walter & Darling (1996) showed an apparent correspondence between *circulation* of journals (u.a.), *in-house use, interlibrary loan,* and *frequency of publication.*

Lochstet & Lehman (1999) correlated *reference question counts* with *door counts* in which the units of analysis were weeks (u.a.). One would expect to find as high a correlation between these two variables (either of which could be dependent on the other), as one would expect between *circulation* and *in-house use.* There was indeed a very high correlation, 0.96, suggesting, at worst, a meaningless comparison or, at best, an error somewhere. At the very least, what is apparent is that virtually all gross counts of library use—whether *in-house,* official *checkouts, reference counts,* or *door counts*—are necessarily highly correlated, because the same users who are counted as they walk through the door (turnstile counts), are counted again when they ask reference questions, and still again when they check out books.[5] What is needed to build theory are correlation studies between library use and variables that are truly independent of use. Circulation, in-house, and other measures of use are not independent of each other.

One such study, in the public library context, is that by Yilmaz (1998), who regressed CIRCULATION against *age, sex, marital status, educational level, occupation, income level* and "*geographic past,*" as well as *social status* and *social role,* in three *socio-economic strata.* Regardless of whether these variables were significant or not, they appear to be truly independent. Users (u.a.) were the units of analysis.

Cooper & Chen (2001) used a logistic regression approach to predict "relevance" of a catalog search session (u.a.), where relevance was defined as a discrete result—that is, WHETHER OR NOT A USER SAVES, PRINTS, MAILS, OR DOWNLOADS A CITATION. The prediction is based on "the time spent performing tasks during the session, and the counts, relative frequencies, and proportions of actions taken during the session," which the authors call "surrogates for user behavior" (p. 826). The unit of analysis was search session (u.a.) rather than individual user (u.a.) because, presumably, individual users made repeated searches.

Most of the variables cited in this review were measured irrespective of their change over time. That is, they were measured at points in time, wheth-

er minute, day, week, or year. A complete theory of librarianship should, of course, consider change over time. Time adds another dimension to the structure of explanatory theory.

Two kinds of past-future use studies are (1) probabilistic frequency distribution studies, which count the number of times a thing happens and where low-frequency is usually more common than high frequency, and (2) forecasting studies in which the units of analysis are sequences of points in time, a univariate framework different from correlation studies, which are usually multivariate, and in which the units of analysis are taken as snapshots in time.

Kasukabe (1990), using multiple regression to study public library use in Tokyo, found that *per capita collection* (presumably holdings), *population of community per thousand librarians, day time population,* and *proportion of administrative workers* were all predictors of PER CAPITA CIRCULATION but differed at different points in time (u.a.). Since "per capita" appears to be a component in each of their variables, one suspects that the significant correlations were due to the colinearity thus introduced.

In two papers McGrath (1995a, 1996a) examined circulation per day (u.a.) over a period of several years. In one paper (1996a), he first converted daily *circulation* in a university library from the time domain to the frequency domain using spectral analysis, and was able to show at least two distinct and pronounced frequencies: A 122-day semester period and a 7-day period.[6]

In the other paper (1995a), he argued that *circulation* per day (u.a.) could be modeled using a combination of three sources: (a) Correlative predictor variables, (b) Normal cyclic influences (time or frequency domains), and (c) A complex or recursive process (from chaos theory) in which some part of circulation is due to previous circulation—for example, when the references in a borrowed book are later borrowed.

Naylor & Walsh (1994) fitted a time series equation to weekly (u.a.) *pick-up* data (books picked up off tables for shelving). Decroos et al (1997) also submitted two years of daily (u.a.) *circulation* data to spectral analysis. They "clearly detected" semester and weekly periodicity. These time and spectral papers suggest that they should be considered when building theory.

Kishida & Sato (1991) used the same approach as Kasukabe above, but without looking at the time component. *Library collection (holdings) per capita, annual per capita acquisitions, number of libraries* in each community (u.a.), *proportion of professional occupations,* and *daytime population* were all submitted to regression analysis as predictors of PER CAPITA BOOK CIRCULATION, but again it is not clear what effect the per capita component has.[7] The explanatory (r-square) coefficients are very high, suggesting colinearity (self-correlation) due to the *per capita component* in each variable. Nevertheless, theirs is an interesting approach to the prediction of circulation in a public library context.

Table 1 recapitulates the dependent variables and units of analysis for

Table 1. Summary of Reviewed Studies for Building A Unified Theory.

Domain/Context	Unit of Analysis*	Dependent Variable*	Authors
Publishing			
Marketing	Journals	Prices	Petersen (1992)
Marketing	Journals	Prices	Chressanthis & Chres. (1994)
National Production	Country	No. monographs published	Kishida and Matsui (1997)
Pricing of Subscriptions	Journals	No. subscriptions	Quandt (1996)
Acquisitions			
Selection	Library	Subscribe or not	Quandt (1996)
Selection/Deselection	Econ. Journals	No. of libraries	McCain (1992)
Selection	Books	No. of libraries	Shaw (1991)
Selection	Titles	No. libraries holding them	Serebnick (1992)
Selection	Reviewed Books	No. books purchased [?]	Calhoun (1998)
Evaluation	Journals	Citation frequency [?]	Kreider (1999)
Selection/Deselection	Journals	Citat. frequency or use [?]	Tsay (1998)
Budget Allocation	Subject Areas	Circulation	Crotts (1999)
Classification			
The Catalog	Bibliographic Families	[?]	Leazer & Smiraglia (1999)
Classification/Browsing	Patron	Avg. no. of books	Losee (1993)
Searching the Catalog	Search Session	Number of hits	Rodman (2000)

Table 1. Con't.

Domain/Context	Unit of Analysis*	Dependent Variable*	Authors
Storage and Preservation			
Storage	Titles	Times circulated	Silverstein & Shieber (1996)
Storage	Books	Times circ'ed; in-house use	Hayes (1992)
Hypothetical Collection	Materials	Demand for materials	Lee (1993)
Preservation	Book Paper	[Brittleness, acidity]	Bennett (1992)
Preservation	Book Paper	[Brittleness, acidity]	O'Neill & Boom. (1995)
Collections			
Library Holdings	Libraries	I.L.L. requests	Exon & Punch (1997)
Circulation			
Use	Journals	In-house/circulation/cites	Blecic (1999)
Use	Weeks	Door counts or ref. queries	Lochstet & Lehman (1999)
Usage	Users	Books checked out	Yilmaz (1998)
Catalog Access	Search Session	Prints/mails/downloads	Cooper & Chen (2001)
Use	Community, Time	Per capita circulation	Kasukabe (1990)
Pickup, Circ.	Day	[No. of books & journals]	McGrath (1993, 1995)
Pickup	Week	[No. of books & journals]	Naylor/Walsh (1994)
Pickup	Day	[No. of books & journals]	Decroos et al. (1997)
Use	Community	Per capita circulation	Kishida & Sato (1991)

* Dependent variables and units of analysis were not always specified as such by their authors.

each study reviewed in each broad domain. Not shown, for lack of space, are independent variables, which, when significant, explain the dependent variable in terms of percentage of variance accounted for using a coefficient, such as R-square or some other statistic. Two important properties of the literature are apparent. First, the table clearly shows the dominant role of circulation or its surrogates at nearly every level, with the possible exception of classification, for which there appears to be a gap in recent explanatory literature. Second, it is clear that a variable at one level can be a unit of analysis at another. Under COLLECTIONS, for example, LIBRARIES (Exon & Punch, 1997) are a unit of analysis whereas under ACQUISITIONS, *number of libraries* is a variable (McCain, 1992; Shaw, 1991). Under STORAGE AND PRESERVATION, BOOKS are the unit of analysis (Bennett, 1992; O'Neill & Boomgaarten, 1995) whereas under CIRCULATION *number of books checked out* is a variable (Yilmaz, 1998). Otherwise, the distinction between variables and units of analysis at the various levels is not always clear or straightforward. If these two important properties are indeed essential to explanatory theory, as this author believes they are, then theorists have much work to do to sort them out.

Summary

The achievements of Copernicus, Kepler, Galileo, Newton, Faraday, Maxwell, Einstein, and others in the reconciliation of natural forces and development of grand unified theory are cited as inspiration for attempting to build a grand unified theory in a humbler sphere, librarianship. Though some may say the vision is naïve or grandiose, the effort to describe the interrelationships of traditional functions of librarianship (i.e., selection, acquisitions, storage and retrieval, classification, collections, and circulation) as integrated and interdependent is an important and worthwhile effort.

Quantitative methods, which can relate the variability of outputs to the variability of inputs, can be used to test the variables of publishing and selection to the variables of acquisitions. The variables of acquisitions, in turn, are important input to storage and preservation which, in combination with the classification scheme, defines the dynamic and static nature of the collection, a necessary condition for its circulation and use. All functions would be tied together in a grand integrated, coherent, and logical scheme in which one functional level explains one level and is explained by another.

To illustrate the potential contribution of recent research to a unified theory, literature for the period 1990 to 2001 was reviewed. Included were studies that used explanatory and predictive statistical methods to explore relationships between variables within and between the broad areas outlined above. These studies do not in themselves constitute broad theory, although, individually, they can be said to posit theory at the narrow level, because when one tests a hypothesis (i.e., computes a correlation) one is also test-

ing theory. One need only connect these hypotheses, these mini-theories, from one level to another.

The review uncovered explanatory studies in nearly every level, with the possible exception of classification, while studies in circulation and use of the library were clearly dominant. A recapitulation showed that a variable at one level may be a unit of analysis at another, a property of explanatory research crucial to the development of theory, which has been either ignored or unrecognized in LIS literature.

It remains for researchers to tie the various levels together more formally—or to find an empirical basis for alternative levels. In a carefully designed study, a theorist might construct a broad scheme in which variables and units of analysis at each level are inevitably and necessarily embraced and follow from the highest level. To a very limited extent, the review suggests that explanatory and predictive relationships do exist and that they can be useful in constructing a comprehensive unified theory of librarianship.

NOTES

* The first half of this paper is based in part on material extracted, shortened, and revised from a paper originally published in Poland in a collection of essays on libraries and democracy (McGrath, 1995b).

1. Nicolaus Copernicus, *De Revolutionibus Orbium Coelestium;* Isaac Newton, *Principia Mathematica;* Johannes Kepler, *Astronomia Nova;* Johannes Kepler, *Epitome Astronomiae Copernicae;* Johannes Kepler, *Harmonices Mundi;* Galileo, *Dialogue on the Two Chief World Systems;* and Albert Einstein (1926), *Relativity: The Special and General Theory,* trans. Robert Lawson, New York: Crown Publishers.
2. "Publishing" refers to the production of books, journals, and printed or stored knowledge.
3. We can also define "publications" to mean any collectible information format.
4. "Circulation" is broadly defined to include not only borrowing, but also use inside the library and interlibrary loans, as well as any other type of use.
5. This author failed to recognize that simple fact in a study thirty years ago (McGrath, 1971).
6. It can be shown that any uniformly cyclic data, such as library circulation, can be graphed either as waves or as closed, elliptical orbits—curves intrinsic to celestial mechanics and a dramatic analogue to the Copernicus-Kepler-Newton context discussed at the beginning of this paper.
7. "Community" (u.a.) appears to be defined as town or city.

REFERENCES

Bar-Ilan, J., & Peritz, B. (2002). Informetric theories and methods for exploring the internet: An analytical survey of recent research literature. *Library Trends, 50*(3), 371–392.

Bennett, R. E. (1992). Paper preservation studies at the University of Manitoba libraries. *Canadian Library Journal, 49*(1), 41–48.

Blecic, D. D. (1999). Measurements of journal use; An analysis of the correlations between three methods. *Bulletin of the Medical Library Association, 87*(1), 20–25.

Broad, W. J. (1984). Is great atom smasher too costly? *New York Times,* section C, Science Times. March 27, C1, C4.

Buckland, M. K. (1988). *Library services in theory and context.* Oxford and New York: Pergamon Press.

Calhoun, J. C. (1998). Gauging the reception of *Choice* reviews through online union catalog holdings. *Library Resources and Technical Services, 42*(1), 21–43.

Chressanthis, G. A., & Chressanthis, J. D. (1994). A general econometric model of the determinants of library subscription prices of scholarly journals; The role of exchange rate risk and other factors. *Library Quarterly, 64*(3), 270–293.

Cooper M. D., & Chen, H.-M. (2001). Predicting the relevance of a library catalog search. *Journal of the American Society for Information Science and Technology, 52*(10), 813–827.

Crotts, J. (1999). Subject usage and funding of library monographs. *College and Research Libraries, 60*(3), 261–273.

Curran, C. (2001). What do librarians and information scientists do? *American Libraries, 32*(1), 56–59.

Decroos, F.; Dierckens, K.; Pollet, V.; Rousseau, R.; Tassisgnon, H.; & Verweyen, K. (1997). Spectral methods for detecting periodicity in library circulation data; A case study. *Information Processing and Management, 33*(3), 393–403.

Exon, F. C. A., & Punch, K. F. (1997). The self-sufficient library collection; A test of assumptions. *Journal of the American Society for Information Science, 48*(1), 11–16.

Ferris, T. (1991). Unified theories of physics. In T. Ferris (Ed.), *The world treasury of physics, astronomy and mathematics* (pp. 116–127). Boston: Little, Brown & Co.

Glazier, J. D., & Grover, R. (2002). A multidisciplinary framework for theory building. *Library Trends, 50*(3), 317–329.

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 .

Hannaford, W. E. (1980). Toward a theory of collection development. In R. D. Stueart & G. B. Miller, Jr. (Eds.), *Collection development in libraries; A treatise* (pp. 573–583). Greenwich, CT: JAI Press.

Hayes, R. M. (1992). Measurement of use and resulting access allocation decisions. *Library and Information Science Research, 14*(4), 361–377.

Kasukabe, C. (1990). A time-series analysis of use of public libraries in Tokyo [Abstract in English, text in Japanese]. *Library and Information Science, 28,* 121–143.

Kishida, K., & Matsui, S. (1997). International publication patterns in social sciences; A quantitative analysis of the IBSS file [Abstract]. *Scientometrics, 40*(2), 277–298.

Kishida, K., & Sato, Y. (1991). A test of the function predicting the book circulation in public libraries by multiple linear regression analysis [Abstract]. *Library and Information Science, 29,* 161–168.

Kreider, J. (1999). The correlation of local citation data with citation data from *Journal Citation Reports. Library Resources and Technical Services, 43*(2), 67–77.

Kuhn, T. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lee, H. L. (1993). The library space problem, future-demand, and collection control. *Library Resources and Technical Services, 37*(2), 147–166.

Leazer, G. H., & Smiraglia, R. P. (1999). Bibliographic families in the library catalog; A qualitative analysis and grounded theory. *Library Resources and Technical Services, 43*(4), 191–212.

Lochstet, G., & Lehman, D. H. (1999). A correlation method for collecting reference statistics. *College and Research Libraries, 60*(1), 45–53.

Losee, R. M. (1993). The relative shelf location of circulated books; A study of classification, users, and browsing. *Library Resources and Technical Services, 37*(2), 197–209.

McCain, K. W. (1992). Some determinants of journal holding patterns in academic libraries. *Library and Information Science Research, 14*(3), 223–243.

McGrath, W. E. (1971). Correlating the subjects of books taken out of and books used within an open-stack library. *College and Research Libraries, 32*(4), 280–285.

McGrath, W. E. (1980). Circulation studies and collection development; Problems of methodology, theory, and a typology for research. In R. D. Stueart & G. B. Miller, Jr. (Eds.), *Collection development in libraries; A treatise* (pp. 373–403). Greenwich, CT: JAI Press.

McGrath, W. E. (1985). Collection evaluation—Theory and the search for structure. *Library Trends, 22*(3), 241–266.

McGrath, W. E. (1986). Science and library science; A challenge to library education. *Journal of Education for Library and Information Science, 25*(3), 219–222.

McGrath, W. E. (1995a). Dynamics of chaos in library circulation; preliminary analysis. In M. E. D. Koenig & A. Bookstein (Eds.), *Proceedings of the 5th International Conference of the International Society for Scientometrics and Informetrics,* June 7–10, 1995, Rosary College, River Forest, IL (pp. 283–292). Medford, NJ: Learned Information, Inc.

McGrath, W. E. (1995b). In the tradition of Copernicus—Building a scientific theory of librarianship; The freedom to research. In M. Kocójowa & G. S. Bobinski (Eds.), *The role of li-*

braries in the democratic process (pp. 58–67). Studies in Librarianship and Information Science, 2(4). Kraków, Poland: Jagiellonian University.

McGrath, W. E. (1996a). Periodicity in academic library circulation; A spectral analysis. *Journal of the American Society for Information Science, 47*(2), 136–145.

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

Naylor, M., & Walsh, K. (1994). A time-series model for academic library data using intervention analysis. *Library and Information Science Research, 16*(4), 299–314.

Nitecki, J. Z. (1993). Metalibrarianship; A model for intellectual foundations of library information science. *Nitecki Trilogy* (Vol. 1). Retrieved August 12, 2001 from http://www.du.edu/LIS/collab/library/niteck. ERIC Document ED363 346.

O'Neill, E. T., & Boomgaarden, W. L. (1995). Book deterioration and loss; Magnitude and characteristics in Ohio libraries. *Library Resources and Technical Services, 39*(4), 394–408.

Peritz, B. (1980). The methods of library science research; Some results from a bibliometric study. *Library Research, 2*(3), 251–268.

Petersen, H. C. (1992). The economics of economics journals—A statistical analysis of pricing practices by publishers. *College and Research Libraries, 53*(2), 176–181.

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.

Quandt, R. E. (1996). Simulation model for journal subscription by libraries. *Journal of the American Society for Information Science, 47*(8), 610–617.

Rodman, R. L. (2000). Making the connection between processing and access: Do cataloging decisions affect user access? *Library Collections, Acquisitions and Technical Services, 24*(4), 443–458.

Rousseau, R. (2002). Journal evaluation: Technical and practical issues. *Library Trends, 50*(3), 418–439.

Satija, M. P. (2000). Library classification; An essay in terminology. *Knowledge Organization, 27*(4), 221–229.

Saxton, M. L. (1997). Reference service evaluation and meta-analysis; Findings and methodological issues. *Library Quarterly, 67*(3), 267–289.

Scriven, M. (1968). The philosophy of science. In *International Encyclopedia of the Social Sciences* (Vol. 14, pp. 84–92). New York: Macmillan Company and the Free Press.

Serebnick, J. (1992). Selection and holdings of small publishers books in OCLC libraries; A study of the influence of reviews, publishers and vendors. *Library Quarterly, 62*(3), 259–294.

Shaw, D. (1991). An analysis of the relationship between book reviews and fiction holdings in OCLC. *Library and Information Science Research, 13*(2), 147–154.

Silverstein, C., & Shieber, S. M. (1996). Predicting individual book use for off-site storage using decision trees. *Library Quarterly, 66*(3), 266–293.

Smiraglia, R. (2002). The progress of theory in knowledge organization. *Library Trends, 50*(3), 330–349.

Tsay, M. Y. (1998). The relationship between journal use in a medical library and citation use. *Bulletin of the Medical Library Association, 86*(1), 31–39.

Walter, P. L., & Darling, L. M. (1996). A journal use study; Checkouts and in house use. *Bulletin of the Medical Library Association, 84*(4), 461–467.

Yilmaz, B. (1998). A sociological study of public library use in Ankara, Turkey. *Journal of Librarianship and Information Science, 30*(4), 259–267.

Zwadlo, J. (1997). We don't need a philosophy of library and information science; We're confused enough already. *Library Quarterly, 67*(2), 103–121.