Combining Bibliometrics, Information Retrieval, and Relevance Theory, Part 2: Some Implications for Information Science

Howard D. White

College of Information Science and Technology, Drexel University, Philadelphia, PA 19104.
E-mail: whitehd@drexel.edu

When bibliometric data are converted to term frequency (tf) and inverse document frequency (idf) values, plotted as pennant diagrams, and interpreted according to Sperber and Wilson's relevance theory (RT), the results evoke major variables of information science (IS). These include *topicality,* in the sense of intercohesion and intercoherence among texts; *cognitive effects* of texts in response to people's questions; people's *levels of expertise* as a precondition for cognitive effects; *processing effort* as textual or other messages are received; *specificity of terms* as it affects processing effort; *relevance,* defined in RT as the effects/effort ratio; and *authority* of texts and their authors. While such concerns figure automatically in dialogues between people, they become problematic when people create or use or judge literature-based information systems. The difficulty of achieving worthwhile cognitive effects and acceptable processing effort in human-system dialogues explains why relevance is the central concern of IS. Moreover, since relevant communication with both systems and unfamiliar people is uncertain, speakers tend to seek cognitive effects that cost them the least effort. Yet hearers need greater effort, often greater specificity, from speakers if their responses are to be highly relevant in their turn. This theme of mismatch manifests itself in vague reference questions, underdeveloped online searches, uncreative judging in retrieval evaluation trials, and perfunctory indexing. Another effect of least effort is a bias toward *topical* relevance over other kinds. RT can explain these outcomes as well as more adaptive ones. Pennant diagrams, applied here to a literature search and a Bradford-style journal analysis, can model them. Given RT and the right context, bibliometrics may predict psychometrics.

Introduction

Peter and Mary, the fictional conversationalists borrowed from relevance theory in Part 1 (White, 2007), are persons,

whereas pseudo-Mary, also introduced there, is an artifact—a bibliographic system, a thing made of language. Information science (IS) begins when we attempt, in certain kinds of dialogue, to replace a person with an artifact—in the metaphor used here, to replace Mary with a system of the kind exemplified by pseudo-Mary. And it ends, I suppose, when the replacement is successful. Far from being dehumanizing, this success will require an intense concern with human cognitive processes such as one presently finds in the relevance theory (RT) of Sperber and Wilson (1986, 1995). Indeed, the thrust of information science should be toward designing and using artifacts like pseudo-Mary to model cognitive processes bibliometrically, as implied at the end of Part 1.

The enduring problem for IS is why people succeed and fail in communications with literature-based systems. Yet as more and more information scientists explain why with more and more variables, the big reviews that give their studies near-equal weight (e.g., Borlund, 2003a; Cosijn & Ingwersen, 2000; Mizzaro, 1997; Schamber, 1994) leave one with a sense of drift rather than parsimonious cumulation (cf. Chen & Xu, 2005). A way forward lies in accepting Harter (1992) as a fundamental reorientation and adapting Sperber and Wilson's relevance theory to clarify our dialogues with both systems and other persons in literature-based information work.

RT has not quantified its variables at the interpersonal level, as Saracevic (1996) points out, but two important measures from human-system dialogues in IS map neatly onto RT. Part 1 showed that the two components of relevance in RT can be operationalized by measures from information retrieval—cognitive effects by logged term frequency (tf) and ease of processing by logged inverse document frequency (idf). Pennant diagrams, introduced at length in Part 1, are scatterplots in which these measures determine the placement of the system's answers in double-logarithmic space. Such space is recognizably bibliometric—it objectively quantifies certain relationships within literatures—but it also can be viewed as a record of human sociocognitive behavior. Pseudo-Mary's "mind," that is, consists of noun

phrases that real authors, editors, and indexers have repeatedly used. In linguistic parlance, hers is a mind composed of utterances (*parole*), along with their occurrence and co-occurrence counts as tallied in databases over time. Pennant diagrams are large-scale domain analyses of these utterances (Hjørland & Albrechtsen, 1995). They capture, unobtrusively, the experts' behavior in associating terms. With these terms, sociocognitive behavior on the user side can be simulated and predicted.

Pseudo-Mary thus represents a synthesis. The ranking of terms on the tf and idf dimensions puts bibliometrics at the heart of information retrieval, and vice versa (Wolfram, 2003). The core-and-scatter distributions of bibliometrics feed into the tf*idf relevance rankings of information retrieval, giving the major subdisciplines of IS a common language. The use of tf and idf to operationalize Sperber and Wilson's two dimensions of relevance binds information science to RT, a rich and subtle account of human cognition and communication.

Studies by homegrown analysts as diverse as P. Wilson (1973), Ingwersen (1992), and Tague-Sutcliffe (1995) seem compatible with RT and can inform its more general account of how cognition and communication work. At the same time, RT can be used to constrain the variables and organize the findings of information science. For example, many of the 80 variables related to relevance in Schamber (1994, table 1) can be taken as instances of cognitive effects or processing effort on the user side, which tightens theory.

It augurs well for the synthesis that RT and IS fit together in previously unsuspected ways. In the long, free-ranging demonstration that follows, I use RT to account for behavior in reference interviews, online literature searches, retrieval evaluation trials, and professional indexing. Along the way, I focus on linguistic mismatches that lessen relevance. The theme of mismatch emerges strongly from five decades of user studies and is detectible even in "system-side" works that shaped the term-matching paradigm of IS (e.g., Lancaster, 1968; Sparck Jones, 1981).

Two small illustrative studies involve pseudo-Mary and pennants. The first shows how specificity of language affected processing effort in a real reference interview and subsequent literature search on the literary theme of "the dark tower" at the Library of Congress. The second shows how ranking journals by yield, in the manner of S. C. Bradford's (1934) classic bibliometric study of lubrication journals, can be read as an attempt to bring journals with the highest predicted cognitive effects—and thus highest relevance—to the fore.

Overall, I hope to establish that RT can answer, in a consistent fashion, some fundamental questions not usually asked in a single article. In approximate order of discussion, they include, Why is relevance central to IS? Why is disambiguation? What is special about *topical* relevance as opposed to other kinds? How can IS be defined and characterized? Why is least effort important? Why is specificity? Why is language in information transactions not always specific? What roles do people's expertise and creativity play? How can vocabulary model levels of expertise? How does pro-

cessing effort affect topical relevance? Is relevance objective or subjective? How are relevance and truth related? How might bibliometrics predict psychometrics? What are the key variables of IS?

Why Relevance Is Central

According to Sperber and Wilson, human beings are relevance-maximizing animals; evolution has attuned them to whatever in their surroundings gives the biggest cognitive bang for the cognitive buck. Thus, Peter does not lose his built-in expectations of relevance when he shifts from conversing with Mary to conversing with pseudo-Mary. But whereas the real Mary could match him step for step in implying and inferring meanings, pseudo-Mary is almost infinitely less intelligent and can perform only a few tricks of memory and display. Her one advantage over the real Mary is her ability to retain and faithfully reproduce enormous stores of messages. The messages are typically stored because of their *potential* relevance to multiple recipients; in many cases they represent published texts and jointly constitute what may be called literatures. The typical form of communication with them is a dialogue in which persons put questions and they supply answers. But while the questioners possess their full human faculties, the answering must be accomplished through various kinds of artifice with stored language. Pseudo-Mary, who answers through the language displays of pennant diagrams, symbolizes that artifice here.

Given Peter's expectations and pseudo-Mary's limits, we can understand *why* relevance is the central concept in information science (Wilson, 1968, p. 41; Schamber, 1994, p. 3). According to RT, it is central to human communication *in general,* and that is why, with Saracevic (1975, p. 324), we understand it intuitively. However, our sense of it is never stronger than when we address *artifacts*, because the art behind the artifacts is still so imperfect (cf. Fallows, 2005; Schank, 1997; Sparck Jones, 1991). Superhuman powers of memory are no substitute for the natural human ability to produce optimally relevant communications. In systems like pseudo-Mary, the relevance-maximizing intelligence of the real Mary has been imitated with inferior materials—that is, with computer programs and messages stored outside human heads (White, 1992). Only if the imitation is at least passable by the technological standards of the day will systems like pseudo-Mary be used at all.

At the interface where human questioners meet artificially responsive literatures, there is thus a mismatch: Compared to their human interlocutors, systems are idiot savants (to use an old-fashioned term)—tremendous at remembering millions of bibliographic details, but pea-brained at critically evaluating them. Even the best systems may respond poorly or not at all to people's questions (Blair, 2003, pp. 38–39; Paisley, 1968). In the words of a recruiting bulletin from Google (2004), the major search engine company:

> MYTH: Google is "just a search engine" and search is a solved problem.

REALITY: The perfect search engine would understand exactly what you mean and give back exactly what you want. We are continuously making quality improvements in our search algorithms, but there is a huge gap between where we are today and where we would like to be, and there are huge numbers of open, interesting problems.

Information science (IS) is the science of this problematic interface. It is the science of *artificial* relevance; that makes it subsumable under relevance theory, but as a very special branch: the branch where RT meets still-primitive artificial intelligence (AI), the branch where records replace persons on one side of the dialogue.

At the problematic interface, a user blindly implies interests to a system that cannot infer. The system instead shows the consequences of matching the *verbal form* of the user's message against a stock of written language—a very different thing from understanding what the user really wanted. In a sense, it is *users* who supply the system's understanding by engaging in system-augmented dialogues with themselves. Often they simply make an ignorant guess as to what will work, and systems amplify the consequences back to them. System designers hope that users will learn to communicate better through trial and error (Swanson, 1977), and some systems accept user feedback to improve the term-matching process (Ruthven & Lalmas, 2003), but nothing is guaranteed.

In RT, interpersonal communications have a guarantee of relevance, but not of success. The communicative principle of relevance, as stated by Sperber and Wilson (1995, p. 260), is that "every act of ostensive communication conveys a presumption of its own optimal relevance." An optimally relevant stimulus must be worth the audience's processing effort yet also compatible with the communicator's abilities and preferences (Sperber & Wilson, 1995, p. 270). In IS, the user's search statement (or query or question) is such a stimulus; users always presume that their current statement to the system optimally implies what they want, in the sense that, given their abilities and preferences, it is the best they can do at the moment. The system's response, as contrived by its designers, is also the best it can do. But the latitude for miscommunication remains great, as every retrieval evaluation study shows.

Why Disambiguation Is Central

The natural relevance that persons maintain in conversation depends in part on their ability to disambiguate the senses of homonyms instantly and effortlessly. Peter knows at once that Mary means a financial institution and not the side of a river when she uses *bank* in a certain context (Wilson & Sperber, 2002, pp. 262–267). Systems of pseudo-Mary's sort cannot duplicate such behavior without elaborate artifice. The continuing problem for systems that search large files with rapid term-matching technology is how to disentangle the intended sense of the user's input from unintended senses when all are denoted by the same words or letters. I, for example, want my search term *idf* to stand for

"inverse document frequencies" and not "Israeli Defense Force." Only when a basic match on topic has been achieved can one begin to measure the actual worth of responses to a question. Indeed, the inability of bibliographic systems to disambiguate word senses properly has been the central engineering problem in information science since the 1950s, when failures to disambiguate were first called "false drops." Controlled vocabularies were designed to fix this problem (among others), but it recurs with a vengeance now that natural-language (i.e., *uncontrolled*-vocabulary) retrieval prevails in major systems. Moreover, the many projects to disambiguate word senses algorithmically—relevance feedback, latent semantic indexing, Scatter-Gather, Alta Vista's Prisma, Northern Light's Custom Search Folders, and so on—have not yet produced a clear winner.

Nothing reveals the idiot (as opposed to the savant) side of information retrieval systems like the persistence of this problem. In 2004 I searched Google with *shoulder of mutton*, by which I meant the piece of meat sold by a butcher (cf. Wilson, 1968, p. 46). The topmost relevance-ranked returns were English inns and pubs called "The Shoulder of Mutton," a football club and a biker club in England by that name, and one recipe for a meat dish from the 17th century. If Mary routinely mixed up all the senses of *shoulder of mutton* (or *bank* or any other term) in conversation, Peter would consider her speech not irrelevant but deranged. Pseudo-Mary, on the other hand, cannot begin to guarantee that her own responses are topically relevant to Peter's words, even though, as I noted, they are the best she can do.

Why Topical Relevance Is Central

In classic IS, the relation of topical relevance is taken to hold when a system-supplied document and a user-supplied query are, to some satisfactory degree, on the same topic or about the same thing (Barry, 1994, pp. 149–150). Many writers have assumed that topical relevance is the most fundamental kind (e.g., Froehlich, 1994, pp. 129–131), and it does seem to be the most common basis for declaring "hits" when document retrievals are evaluated by human judges (Bateman, 1998; Chen & Yu, 2005). Other writers (e.g., Barry, 1994, p. 157; Green, 1995) have countered that topical relevance is not the sole basis for judging documents to be, in P. Wilson's phrase (1978, p. 17), "retrieval-worthy." For instance, Harter (1992) and Janes (1994) rightly argue that topical relevance is neither a necessary nor a sufficient reason for accepting a document. It is not necessary because items called "hits" need not match a query in topic; it is not sufficient because items that do match can still be discarded as "junk" or "noise."

Nevertheless, topical relevance *is* necessary in the sense that homographic words and phrases (e.g., *bank*, *shoulder of mutton*, *idf*) must not depart utterly from what the user has in mind. Topical relevance thus puts a floor under retrieval-worthiness. In natural-language retrieval, users typically underspecify what they want, and this tendency combines with obtuseness like pseudo-Mary's to yield two kinds of

documents: those that are truly "off topic" and those that are "on topic" but of varying degrees of acceptability. The latter can be irrelevant in the RT sense that they have little or no cognitive effect on the user. Recall Goatley (1997, p. 139) in Part 1: "..if there is no Contextual Effect there will be no relevance, no matter how little the Processing Effort involved."

His observation squares with what many information scientists have said—that documents that are on one's topic might still be quickly rejected because they are, e.g., already known, or from suspect sources, or at an inappropriate reading level, or the wrong length—the equivalent of talking about the right topic with the wrong persons. But that is very different from rejecting documents that represent a complete misunderstanding—the equivalent of correcting someone who gets even the topic wrong, such as hearing *Eunuchs* for *Unix* (cf. Franzen, 1992, p. 20). In a Google search on *idf*, I had to reject numerous items because I could not use everything that discussed inverse document frequencies, but that was not like my rejection of items that discussed the Israeli Defense Force. In the first case, the documents had at least some chance of acceptance; in the second, none at all. (The two kinds of rejection should not be lumped together in the *Retrieved but Not Relevant* cell of the classic table for figuring recall and precision ratios in evaluation studies, yet who knows how often they have been?)

Discussions that follow will reveal how topical relevance, as opposed to other kinds, emerges from considerations of least effort in several key areas of IS.

Defining Information Science

Pennant diagrams can be used to illustrate key beliefs from this line of argument—namely, that the central communicative exchange in information science is between a person and a system (rather than directly between persons); that the nature of systems in IS is to provide answers to persons' questions within certain domains (i.e., to behave like the staff at a desk bearing the international information symbol); that the answers are not, as a rule, from someone's personal knowledge, but from published or otherwise available writings (e.g., books, journals, Web sites, "public knowledge"); that, nevertheless, dialogue with the system involves the same expectations as dialogues with another human being (i.e., positive cognitive effects and acceptable ease of processing); and that word-sense disambiguation is still a major technical challenge at the human-system interface. There is nothing new in all this; pennants merely exemplify what many information scientists would affirm. But if information science is chiefly concerned with artificial relevance—with systems that answer questions from writings—then pennants also serve as checks on some oft-repeated but overgeneral definitions of IS as a discipline (e.g., Borko, 1968; Reitz, 2004).

Information science can be concisely defined as *the study of literature-based answering*. The point about where the answers originate—writings, records, documents—is central to an accurate definition. IS does not deal with "information"

in every sense; in fact, informative processes about which it has nothing at all to say occur constantly(cf. Klinkenborg, 2003). IS might ultimately be connected with some general science of these processes, of course, but its frequent association with the more limited world of librarianship—as when it is called "library and information science"—is highly indicative of its true nature, both past and present.

A way of appreciating this claim is to recognize the similarities between most of what information scientists call "information systems" and reference desks in libraries, even in their precomputerized state. Pseudo-Mary can be seen as a mechanical extension of the powers of a human reference librarian, as can such real-world information systems as Google, Dialog, Scirus, and Ask. If search engines like these eventually become more like persons, they might resemble Vox, the holographic reference librarian in the 2002 film version of H. G. Wells's *The Time Machine.* Even so, such systems are not designed to pass as fully human; unlike, say, the replicants in *Blade Runner,* they cannot respond appropriately to any question in any situation. But creating that ability is not the foreseeable goal of IS. For example, IS does not seek to create humanoids capable of the following question and answer, although the exchange undeniably involves relevant information:

Peter: What are you doing after work?

Mary: I'm going to have a drink at the Shoulder of Mutton.

Peter's question here is not a reference question (and he would not put it to pseudo-Mary); nor is Mary's answer an answer from writings, arrived at through indexes. In contrast, the information science portrayed here aims at extracting relevant answers from popular, scholarly, professional, and scientific writings, both on and off the Web. Sometimes a system will exist to provide answers *about* these writings, and sometimes it will exist to provide answers *from* them. If a person acts as a mouthpiece for the system, the answers will nevertheless be from the writings rather than from the personal knowledge of the mouthpiece, as at library reference desks. Systems with human intermediaries are still systems, not persons.

It therefore seems clear that *explanatory* information science should be able to account for regularities of behavior at what are literally or figuratively reference desks—that is, settings in which people, including librarians, must search literatures. As will be seen, pennant diagrams can cast light on these regularities, and others as well. Let me preface my examples with some key points from Sperber and Wilson.

RT and Least Effort

RT often deals with the way a hearer (e.g., Peter) infers meaning from something said by a speaker (e.g., Mary) in conversation. This is done through the "Relevance-theoretic comprehension procedure" (Wilson & Sperber, 2002, p. 259):

a. Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.

b. Stop when your expectations of relevance are satisfied.

This procedure is glossed in Sperber and Wilson (1996, p. 531):

> At any given moment in one's cognitive life, there is a wide range of new information being monitored in the environment, and there is an even wider range of information in memory, bits of which might be activated and would provide a context in which to process the information from the environment (or other pieces of information from memory). Only some of the possible combinations of new and contextual information would yield relevance, and this to a greater or lesser degree. There is no way for the mind to review all possible combinations of new and contextual information in order to find out which would maximize relevance. Even if there was a way, the effort involved in such a review would so lower the overall cognitive utility of the process as to defeat the whole enterprise. So how should the mind proceed? Since it cannot have foreknowledge of relevance, how can the mind have, at least, non-arbitrary *expectations* of relevance?
>
> To begin with, when expectations of effect are wholly indeterminate, the mind should base itself on considerations of effort: pick up from the environment the most easily attended stimulus, and process it in the context of what comes most readily to mind. *Ceteris paribus,* what is easier is more relevant, if it is relevant at all.

Here we see that least effort operates to constrain the hearer's relevance-seeking so that it terminates in real time. However, to introduce RT into IS, we need a complementary procedure for speakers—especially speakers who are initiating relatively impersonal or anonymous communications (unlike those of Peter and Mary, who are intimates). Examples of the latter would be conversations with an unfamiliar reference librarian or a bibliographic system.

A general guide for speakers appears in Wilson and Sperber's (2002, p. 257) discussion of *optimal relevance*:

> The communicator wants to be understood. It is therefore in her interest—within the limits of her own capabilities and preferences—to make her ostensive stimulus as easy as possible for the audience to understand, and to provide evidence not just for the cognitive effects she aims to achieve in her audience but also for further cognitive effects which, by holding his attention, will help her achieve her goal.

Sperber and Wilson also express "capabilities and preferences" as what the speaker is *willing and able* to produce in the way of ostensive stimuli.

Thus, shaped by her own experience as a hearer, the speaker can assume that a least-effort filter also operates in whoever receives her communication. But when communicating with nonintimates, she will not have an extensive history of exchanges on which to base her expectations of cognitive effect. In this situation, "when expectations of effect are wholly indeterminate," the natural course for her is to economize on her own effort: to offer some minimal, easy-to-comprehend stimulus toward being understood and see whether the response has a satisfactory payoff. (The folkloric advice is "KISS: Keep It Simple, Stupid.") The problem here, of course, is that, because of the speaker's relative ignorance of the *hearer's* capabilities and preferences, the actual words of her stimulus may be an inferior way to achieve her goal. What she *thinks* is least effort for the hearer could be very different from what *is*.

RT at the Reference Desk

Miscommunication of this sort is exemplified at library reference desks when inquirers underspecify what they want. The reference interview is the means by which librarians rework a vague or underspecified question until it represents what the inquirer really wants in a form appropriate for searching. We can use RT and pennant diagrams to elucidate this exchange.

The problem of underspecification at the reference desk occurs because inquirers seek a least-effort way of being understood by reference librarians. Since they are uncertain about the cognitive effects of their questions (and are ignorant about indexes), inquirers anticipate that librarians will understand them most easily if they "keep it simple"—that is, express what they want in highly generic terms, like those that appear in sector C of pennants. This is called "the label effect" by Ingwersen (1992, pp. 116–118). Such terms are simple in the sense that processing them requires relatively little cultural literacy. (One needs to know common phrases or famous names, but not the actual titles of writings.) The terms may in fact correspond to broad areas of the library stacks (e.g., "English literature," "the psychology journals," "your books of literary criticism"), and this converts questions that can be answered only by experts in library sources into directional questions that can be answered by anyone familiar with the library floorplan. (The American Library Association uses this very difference, knowledge of sources vs. knowledge of the floorplan, to distinguish professional from nonprofessional reference service.) Harking back to Part 1, it is as if Peter asked a reference librarian, "Where are your books on American literature?" when his real question is one for a reader's advisory service: "What should I read to follow up on *Moby Dick*?" And indeed, *American literature* (as a journal title) is one of the terms in sector C of the *Moby Dick* pennant.

The similarity to typical shopping behavior is plain: The anonymous reference librarian is taken to resemble a store clerk who can be asked for directions (for example, "Where do I find lamps and glassware?") but not for exact information about the item to be purchased. That item (for example, a 2-inch hexagonal pendant for a crystal chandelier) one may expect to *recognize,* mutely, after being directed to the right place. By this logic, one reveals one's true interest at the reference desk only gradually, if at all, letting the librarian's responses govern how much more one says. Since it would be embarrassing to make a long, candid disclosure

only to find that one has been talking to the wrong person (Eichmann, 1978), a least-effort approach allows one to back out, if need be, at any stage of the dialogue. It may even seem *polite,* in the sense of preventing embarrassment for the librarian, to assume that his or her ignorance largely mirrors one's own (Bates, 1998, pp. 1192–1193). At the same time, disclosure as a kind of unveiling game may serve another major purpose for Sperber and Wilson's speaker—to hold the librarian's attention.

Of course, this kind of least effort by inquirers is just what many reference librarians do *not* want. For them as hearers, least effort and greatest cognitive effects would result from a precise statement of the inquirer's true question—the actual title of a work to be consulted or an indication of an interest couched in specific, searchable terms. The whole point of the reference interview, when it occurs, is to arrive at this statement, because, given large collections of writings, it costs *great* effort to seek the highly specific in the highly generic by way of browsing and recognition alone. Several of these points are illustrated in *The Gold Bug Variations* (Powers, 1991, pp. 26–32)*,* a novel whose narrator is a reference librarian.

Generic Versus Specific

In the language of pennant diagrams, librarians want a proper seed term with which to search. They know that the generic terms typical of sector C are hard to process in relation to the true seed term (*measurably* hard to process, the pennant shows) and make poor substitutes for it. If Peter's real question is about *Moby Dick,* librarians know that *Moby Dick* should be used as the seed term to imply American literature (among other things), and not the other way around, because the implications of the latter are far too wide. Were *American literature* entered as a seed term in Arts and Humanities Search, answers to Peter's real question would not be brought forward on the cognitive effects scale; they would be jumbled with countless false drops far back on that scale, if they appeared at all. It is certain, moreover, that sectors A and B would not concentrate titles explicitly or implicitly relevant to *Moby Dick,* as they did in several figures in Part 1. Only seeds that represent the true question make for good pennants.

Pennant diagrams are, of course, simply a new way of talking about an old canon of library service—that reference librarians should not leap to answer an inquirer's question as first stated, because the real question is often quite different. It might be argued that, for the librarian, simply taking the inquirer's question at face value and answering it as posed would be a least-effort response: "American literature? Those are the PS books on the fourth floor. And the PS reference books are over there behind that table." However, unless the inquirer then knows exactly what to look for in a particular work, such an answer would soon prove to have high processing costs and few worthwhile cognitive effects—RT's very definition of a largely irrelevant reply. That is why reference librarians are taught to elicit and respond to the inquirer's real question. A librarian who answered presenting

TABLE 1. Speaker-hearer communications by relevance at library reference desks.

Speakers	Hearers	Less relevant for *hearers:* lower contextual effects, higher processing effort	More relevant for *hearers:* higher contextual effects, lower processing effort
Customers	Librarians	Generic questions	Specific questions
Librarians	Customers	Generic answers	Specific answers

questions as just described and never interviewed—there are some—would seem to be communicating in the same ironic mode as union members when they "work to rule," subverting management's directives by overliteral compliance with them. While this kind of literalism at the reference desk indeed costs less effort than interviewing, it also betrays a level of alienation that most would find unseemly in professional librarians.

Some librarians would justify such responses on the ground that customers (especially students) should learn to serve themselves and not be "spoonfed" answers. For many other librarians, however, least *satisfactory* effort will consist in driving toward substantive answers that they, not their customers, supply. Some of the customers will know how to advance this process with specifically worded questions or search strategies. A presenting question can thus be generic or specific; the one that costs the customer least effort will depend on his or her knowledge of reference work. And a least-effort response can be perfunctory or probing, depending on the librarian's service ethic (cf. Wilson, 1986).

In her book introducing RT, Blakemore (1992, p. 130) notes that hearers' interpretations of what they hear are always constrained by the search for optimal relevance. "What varies," Blakemore writes, "is the *degree of specificity* of the speaker's expectations about the way that optimal relevance will be achieved" (italics mine). Table 1 illustrates this argument in the setting of the library reference desk. Two levels of relevance appear to capture the major division in the way specificity affects *hearers.* RT predicts that, on the dimensions developed here, specific questions will be more relevant for librarians, and specific answers will be more relevant for their customers. At the same time, RT can plausibly explain why many of their exchanges involve generic questions and/or generic answers. It is because they cost less effort for *speakers,* as opposed to hearers. Within linguistic pragmatics, Horn (1984), following Zipf (1949), has noted that effort for speakers varies inversely with effort for hearers. Although my examples are very different from Horn's, that seems to be occurring here also.

Table 2 illustrates *generic* and *specific* when they are applied to questions and answers. The dialogue is between a librarian in a public library and a customer who wants to know Katherine Hepburn's astrological sign. *Generic* and *specific* take their meaning from comparisons of the search terms that are extracted from the questions. The customer's

TABLE 2. Examples of generic and specific communications at library reference desks.

Customer:	Generic question	Do you have biographies of actresses?
Librarian:	Generic answer	Yes, they're in the 900s, marked with a "B."
Customer:	Specific question	Can you by any chance tell me Katherine Hepburn's sign?
Librarian:	Generic answer	You might look for a biography of her in the 900s.
Customer:	Generic question	Do you have biographies of actresses?
Librarian:	Specific answer	[After interview and lookup] Katherine Hepburn's birthday is May 12, so she's a Taurus.
Customer:	Specific question	Can you by any chance tell me Katherine Hepburn's sign?
Librarian:	Specific answer	[After lookup] She's a Taurus, born May 12.

generic question term is *biographies*, and the librarian's generic answer term simply echoes that, with additional information on how to find biographies in the Dewey classification. The librarian's reply, while not wholly irrelevant, is much less relevant than a direct answer to the customer's real question would be. The customer's specific question terms are *Katherine Hepburn* and, implicitly, *birthday* and *sign of the zodiac in which this birthday falls*. Given these, any decent librarian could find the specific answer to the customer's question in less than a minute through the *World Almanac,* Google, Ask, Yahoo! or other sources. For the customer, this direct answer would surely have greater cognitive effects and be easier to process than the alternative of being sent off to the stacks with the question still hanging.

The Dark Tower

My discussion of least effort derives in part from Mann's (1993, pp. 3–7) account of one of his customers at the reference desk of the Library of Congress—a woman who believed that the only way to find materials on the literary theme of "the dark tower" (as in Robert Browning's "Childe Roland to the Dark Tower Came") was to ask for the books on English literature and then drill downward. She had no inkling that one can, and should, search on the most specific terms expressing one's interest—a mistake that even highly educated people make (cf. Mann, 1993, pp. 79–82). Through her unawareness of Mann's capabilities and preferences, this woman personifies the hypothetical speaker about whom I wrote: "What she *thinks* is least effort for the hearer could be very different from what *is.*" Bit by bit Mann overcame her reluctance to reveal what she wanted and was able to propose several good sources, including ISI's *Arts and Humanities Citation Index* and the Modern Language Association's *MLA Bibliography,* both of which can be searched with ordered keywords like *dark tower.*

Figure 1 is a pennant diagram made from the *MLA Bibliography* (as offered by First Search) in response to the woman's real question. The diagram reproduces a large selection of descriptors from the bibliographic records of some 43 articles matching the phrase "dark *with* tower." Searches on descriptor phrases by themselves produced the counts that were used to compute idf values. The pennant has been drawn to enclose terms relevant to the seed at a certain level of specificity—2 or higher on the ease of processing scale. These terms exhibit MLA's distinctive style of indexing, which also produced the pennant's lower fringe—a handful of highly generic terms that evoke the woman's presenting question (which at one point was "English literature in the 19th century").

Consistently with everything said thus far, these generic fringe terms have low scores on the ease of processing scale. That is because they add no more to the "dark tower" theme than they would to thousands of others. (Recall the principle that the more widely indexing terms are applied, the less informative they become.) So why should several of them place high on Figure 1's cognitive effects scale, indicating high relevance to the seed term? It is because the *MLA Bibliography* always adds broad categories from a faceted classification scheme to its indexing: that is, generic terms that unite articles (e.g., the national literature, time period, and literary genre they belong to) are always added to the specific terms that differentiate articles. *English literature*, for example, is high on the cognitive effects scale because it occurs first in the indexing of 38 of the 43 "dark tower" articles. The following nested string is typical:

> English literature–1800-1899–Browning, Robert–"Childe Roland to the Dark Tower Came"–poetry–allusion–treatment of the quest–triumph–sources in Shakespeare, William–*King Lear*

In the pennant diagrams of Part 1, generic terms would have placed similarly high on the effects scale if the data had been expanded to include, e.g., terms identifying the parent literature of every novel cited with *Moby Dick* or the parent discipline of every author and paper cited with Harter's article.

The phrases seen in Figure 1 are detached from nested strings like the one above and make full sense only when read in their original contexts. Even so, those inside the pennant clearly enrich the "dark tower" theme. The work most highly relevant to it is indeed Browning's "Childe Roland" (a childe is a young noble intent on knighthood). Mann's customer probably had Browning's poem in mind when she made her vague initial request; in any case, the pennant data confirm its importance. Also brought forward is "The Dark Tower," the title of a science fiction story by C. S. Lewis and of a radio play by Louis MacNeice; as a descriptor, the phrase retrieves writings on both of these 20th-century British works.

Particular treatments of themes related to the "dark tower" can be seen (e.g., the quest, voyage, despair, discovery, the grotesque), as can various perspectival approaches (e.g., archetypal, psychoanalytic, biographical). "Childe
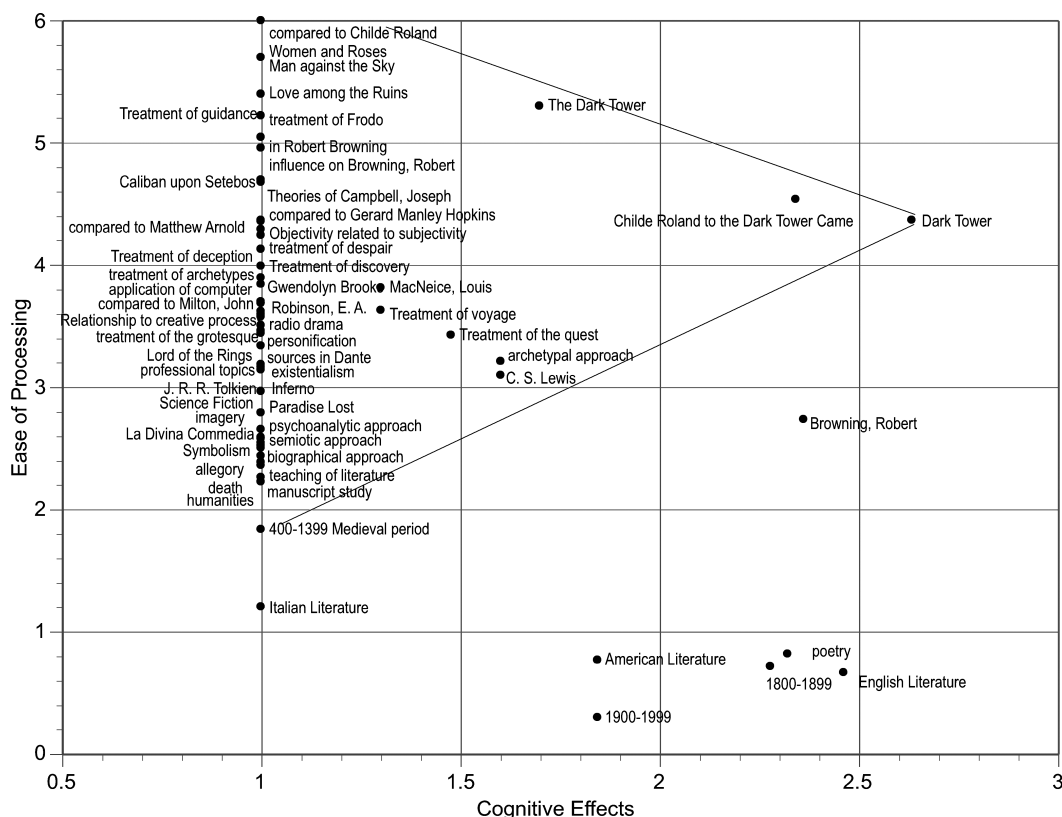
FIG. 1. Pennant diagram of *MLA Bibliography* indexing co-occurring with *Dark Tower*.

Roland" is compared to other poems by Browning (e.g., "Women and Roses," "Love among the Ruins," "Caliban upon Setebos") and to works by other authors (E. A. Robinson's "The Man against the Sky," Tolkien's *The Lord of the Rings,* Milton's *Paradise Lost*). In other articles, Browning's sources for "Childe Roland" in Dante (the *Inferno*) and Shakespeare (*King Lear*) are discussed.

Mann guided his customer to the point where specific answers like these became possible. Many of them presumably are relevant to her interest and worth her attention. Yet none of them would have emerged (without long delay) had she persisted in asking her original question and Mann not drawn her out. The dislocation of the generic terms from the "true pennant" in Figure 1 is evidence that the original question was not likely to be productive. As asked, it did not begin to suggest the relevant materials that appear in the pennant once the proper search term is used.

Such claims can be generalized. Pennant diagrams can be created, at least in principle, for any query put to writings, as at a reference desk. In keeping with RT, pennants present various answers by their degrees of relevance. More particularly, they show the degrees of relevance that are attained when specific, well-focused terms are chosen over broad, vague alternatives in a literature search. They show that some terms will be much harder to relate to the seed than others, and they imply that the difficult terms, while not *wholly* irrelevant, should not be preferred. (By hypothesis, pennants could show this not only for literature

searches but for questions with specific factual answers, like Katherine Hepburn's sun sign—a conjecture for another day.)

Pennant diagrams thus reveal the difference between good and bad search terms in a single, algorithmic graphic. People search badly when they use broad, vague terms from their own vocabularies, thinking these will somehow call up specific titles that they can then recognize as being right for retrieval. They do not realize that these terms will produce matches that are low on the ease of processing scale. Describing such searchers, Bates (1998, p. 1186) wrote, "The user, in effect, describes the fringes of a gap in knowledge, and can only guess what the 'filler' for the gap would look like" (cf. Belkin, 2005). A diagram like Figure 1 shows both the "filler" for the gap (the good terms inside the pennant) and the fringes of the gap (the user's original bad terms, which in this case resemble a ragged fringe below the pennant). It is, moreover, based on statistical evidence rather than anecdote.

Returning to Mann, I would add that the woman's question as first posed was not dumb; it simply reveals the underdeveloped knowledge that many people have of reference librarianship and especially of indexing. As noted previously, she chose her wording to place what she thought were acceptably low demands on the hearer, an unfamiliar librarian. Moreover, her question was not as generic as can be imagined; it at least limited her search to English literature (as opposed to the literatures of other countries and other

disciplines). It even agreed with the classification and indexing schemes she expected to encounter. In the Library of Congress classification scheme, for example, the top class of many of the books and journals that deal with her chosen topic is "PR," which stands for "English literature." In the *MLA Bibliography,* the first entry in the indexing of many of the articles that might interest her is, again, "English literature." According to Bates, Wilde, and Siegfried (1993), numerous scholars in the humanities find combinations of broad facets—such as *literature*: English, *period*: 1800–1899, *genre*: poetry, *author*: Browning, Robert—the most congenial way to define their interests. That is how they think, and the Modern Language Association (MLA) tailors its indexing accordingly. So at worst Mann's customer can be faulted for not knowing how much deeper standard indexing goes, and that is a fault shared by a great many people. At one time or another, almost everyone finds a "shopping model" (locate the right department, then browse till you recognize what you want) easier to follow than a "specification model" (put an exact description of what you want into words).

Pinpointing a False Drop

In another of Mann's sources, ISI's *Arts and Humanities Citation Index* (AHCI) (Arts and Humanities Search online), the *dark tower* phrase retrieves nine articles, and a pennant made from the works cited in them corroborates several earlier findings—e.g., the importance of Browning's Childe Roland poem and the possible relevance of works by C. S. Lewis, J. R. R. Tolkien, and Louis MacNeice. (Four of the nine AHCI articles are the same as those from the *MLA Bibliography*.) Mann's customer knew something of English literature, and, in going over the cited works in the AHCI pennant, she presumably could recognize the relevance to her question of writings by, e.g., Malory, Keats, Shelley, Christina Rossetti, Tennyson, and the American scholar Harold Bloom. In various ways these relate to the theme of a knightly quest that ends in a desolate, darkly significant place. (Stephen King's Dark Tower novel series is also related.)

What might be unexpected are references to "the dark tower" in African American literature and specifically to one of its writers, the poet Gwendolyn Brooks. Brooks also appears midway down the left column in the *MLA Bibliography* pennant in Figure 1. *These* references use the seed phrase in a sense quite different from what the woman intended by it and, in the context of her question, represent false drops. In this second sense, "the dark tower" is associated with the Harlem Renaissance—with a poem and magazine column by Countee Cullen and with A'Lelia Walker's nightclub and literary salon. It also has this sense in the title of a 1974 book, *From the Dark Tower: Afro-American Writers 1900–1960,* in which Arthur P. Davis discusses, among others, Gwendolyn Brooks. It is the Davis-Brooks connection that turns up in Figure 1. As far as I know, no one has linked Cullen's or Brooks's work thematically with Browning's. In terms of RT, Browning's poem is easier to relate to the woman's presenting question than any African American work; it is thus more relevant to her question than any of the latter.

If we attribute the MLA and the AHCI pennants to pseudo-Mary, we can see at once the difference between a system like her and a person like Mann. Were Mann sifting through the various retrievals on his customer's behalf, he could hardly fail to notice the very different contexts in which the phrase *dark tower* appears in records from (a) Browning studies and (b) African American studies. He would almost certainly ask his customer whether the latter connection was meaningful to her and whether she wanted to pursue it. Pseudo-Mary, in contrast, has no way of differentiating retrieval (a) from retrieval (b). She has no way of interviewing a customer to see whether one interpretation of search input is preferable to another. She even lacks what some systems have, at least in primitive form—a module that places retrieved records in separate bundles on the basis of textual similarities, in an effort to achieve topical relevance. Recall that Sperber and Wilson's relevance-theoretic comprehension procedure calls for rapid disambiguation of word senses as one way of economizing on effort in conversation. If pseudo-Mary could disambiguate the senses of *dark tower* for Mann's customer, she would separate the Davis-Brooks retrieval from documents that were at least topically relevant to the Browning retrieval, however acceptable they were otherwise. Since she cannot, the Davis-Brooks retrieval is indistinguishable in the pennant from documents much more likely to be hits.

RT and Online Searching

Let me not, however, be too hard on pseudo-Mary. I said earlier that, lacking human inferential powers and common-sense, she and other bibliographic systems resemble idiot savants, whereas the people who use them possess full human intelligence. But there are actually two mismatches: In the limited areas where systems are savants, people, as we have seen, routinely underspecify what they want. This asymmetry of strengths and weaknesses is perhaps the most important finding in several decades of user studies in IS. It is well described in Bates's (1998) synthesis of key user studies. RT can illuminate this finding, just as it could the dialogues between reference librarians and their customers.

In classic IS accounts of the human-literature interface, writes Bates (1998, p. 1186), "it is commonly assumed that indexing and searching are mirror images of each other." On behalf of the system, indexers supposedly use a certain vocabulary to describe documents, users supposedly adopt the same vocabulary to describe their needs, and the system's job is to match the descriptions and retrieve relevant materials (Lancaster, 1968, pp. 1–5). In fact, as Bates makes clear, this account is highly idealized. Human or automated indexers describe documents with the documents before them, so to speak, and it is their informed descriptions of known items that underlie the "savant side" of systems. By contrast, users must simultaneously describe both their interests and

documents they have never seen, and many have little clue as to what terms, in what combinations, would best serve them (Belkin, 2005; Taylor, 1968). They may even be unaware of their own ignorance. When searching online library catalogs, for example, few people read directions or look up controlled vocabulary; they simply plunge in with whatever language comes to mind (Bates, 1998, p. 1187; Drabenstott, 1991, p. 68; Drabenstott & Vizine-Goetz, 1994, p. 152). Not knowing what lies behind the veil of the system, they often put questions too broadly or with too little contextual detail for systems to give their best answers. Proposed remedies are a staple of every textbook on online searching.

The "underspecifier side" of users is suggested in le Roux's (2003) *Search Engine Dictionary*. After defining a *False Drop* as "a web page displayed in the SERP [search engine results pages] that is not clearly relevant to the query," he goes on to say, "The most common cause of false drops is words with multiple meanings. If the query gives no indication of context, the search engine has no way of predicting which of the possible meanings the user has in mind." What he says about Web pages applies to document retrieval in general: it is the query that constrains false drops, and the query is the user's lookout, not the system's (even when the latter provides help with the query). For example, just as Mann's customer invited untold false drops by asking simply for "English literature," it is clear that my own relatively specific searches on *shoulder of mutton*, *idf*, and *dark tower* could have used more disambiguating detail (that is, additional qualifying search terms) if many off-topic items were to be avoided. The reason I did not add these qualifiers is that, like most searchers, I could not foresee they were needed. In searching, good results are usually arrived at through successive approximations.

The approximations begin with the choice of an information source or system to be searched. In Part 1, I sidestepped the choice-of-system problem by showing Peter and pseudo-Mary already in contact (like Peter and the real Mary); furthermore, pseudo-Mary had already been given data on *Moby Dick* from Arts & Humanities Search to display. In life, people who turn to literature-based systems may know a source worth consulting, but this cannot be assumed. What *can* be assumed is that, at least initially, they will tend to exhibit labor-saving incuriosity about what will best answer their questions. (The social equivalent occurs when people ask the *nearest* person for help rather than trying to find the *right* person, if that is not immediately obvious.) This tendency to prize convenience over quality, which Paisley (1968, p. 6) long ago called "information nonchalance," accords with RT's least-effort principle.

Sperber and Wilson state this principle for hearers as we saw: "when expectations of effect are wholly indeterminate, the mind should base itself on considerations of effort: pick up from the environment the most easily attended stimulus, and process it in the context of what comes most readily to mind." But hearers are also speakers, and the same economizing principle extends from their hearing and understanding to their speech. That is, given indeterminate expectations of effect, the speaker will *seek to engage the most easily attended audience with whatever most readily comes to mind,* a formulation that covers both the choice of whom to talk to—or the information source to consult—and the choice of what is said. Note that one does not choose to say just anything. According to RT, speakers can engage hearers successfully only by saying what is attention-worthy. So the speaker's practical purpose and any foreseeable audience resistance will constrain "what most readily comes to mind."

When artifacts replace persons as sources of answers, the principle might read, *Put to whatever source is closest the first words that express your interest.* ("Closeness" can be physical or psychological or both.) User studies show that people tend to follow this principle. In human-system dialogues, they try to elicit what they want by uttering as little as possible—a form of speaker's economy (cf. Jansen, Spink, & Saracevic, 2000; Spink, Wolfram, Jansen, & Saracevic, 2001). They do this partly because they are ignorant of optimal terms and partly because systems resemble anonymous persons (such as clerks) more than they do intimates. But when the speaker supplies too little context by using imprecise language and/or too few terms, the system as hearer cannot discriminate and responds with everything that could possibly fit—a glut of documents, many in need of rejection. Thus, RT accords with standard IS notions of low precision in document retrieval. We have seen the same thing happen with reference librarians, except that many of them know enough to counteract the glut through interviewing.

At the same time, speaker's economy finds opposition in hearer's economy. In human-system dialogues no less than dialogues between persons, hearer's economy consists in language that precisely specifies the interest and that contains enough terms to set a reasonably unambiguous context. That is what reference librarians want so as to be most helpful and what systems respond to with fewer and better documents. Unfortunately, better input for "hearing" systems places additional demands on "speaking" users, who tend to be uncertain of what terms to choose and how best to connect them to disambiguate sense. Such uncertainty takes effort to correct. Persons addressing systems therefore face a trade-off: Less effort for them at input time means greater effort for them at output time, and vice versa.

It would seem that the least-effort strategy used by most people—*Try the first source and the first words that come to you, then browse the output, editing as you go*—is all that can reasonably be expected. Table 3 has this majority strategy down the column at left. For speakers, it involves taking little enough trouble with input that any halfway acceptable output seems relevant. One succeeds, in other words, by intuitively decreasing the denominator in the equation from Part 1: Relevance = Cognitive Effects/Processing Effort. When these speakers in turn receive ("hear") the system's response, they find that their nonchalance means output that is harder to judge, in the sense of being more time-consuming to edit for false drops. However, the predicted effort in

TABLE 3. Majority strategy at left; minority strategy at right.

For users as speakers

Easier to choose	*Harder to choose*
Most readily available source.	Best source among critically evaluated sources.
Generic terms.	Specific terms. Term combinations. Search tactics that disambiguate sense.
Ambiguous single terms.	

For users as hearers

Harder to judge	*Easier to judge*
Output from readily available source.	Output from critically evaluated sources.
Output from generic terms or ambiguous single terms.	Output from specific terms, term combinations, and search tactics that disambiguate sense.

judging can itself be lessened if they accept whatever first produces cognitive effects and ignore the rest.

A great many Internet searchers apparently follow this "satisficing" strategy. They routinely act as if all useful documents are now on the Web, that everything can be found through a single search engine such as Google or Yahoo!, that a word or two in the search box will produce what is wanted without use of advanced features (e.g., Boolean operators, relevance feedback), and that anything not in the top-ranked returns is not worth looking at. As Jansen (2000) writes, "The behavior of Web searchers adheres to the principle of least effort (Zipf, 1948 [*sic,* 1949]), which postulates that there are 'useful' behaviors that are quick and easy to perform. The very existence of these quick, easy behavior patterns then causes individuals to choose them, even when they are not necessarily the best behavior from a functional point of view." Yang's (2005) review confirms an impression of least effort by these searchers.

The column at right in Table 3 has a contrasting minority strategy, used, for example, by professional searchers or highly motivated scholars. The minority strategy is more carefully crafted. It requires knowledge that, if not initially available, may accumulate through trial-and-error searching. In either case, it is the strategy of people who are seriously pursuing a topic, perhaps because their reputation may be affected by the result (cf. the accounts by Hölscher & Strube, 2000, Jansen & Pooch, 2001).

Given the same initial problem, what varies across the two columns of Table 3 are the information sources people choose and the phrasing of their questions. Both are expressions of individual differences in degree of expertise—in what and how much people know. Such individual differences have been found to be a major variable in user studies (Borgman, 1989; Drabenstott, 2003; Harter, 1996; Hirsch, 2004; Saracevic, 1991; Schamber, 1994, pp. 20–21). When individual differences in any domain are simplified as much as possible, the final division lies between lay publics (people not immersed in a subject, possibly including the young) and subject experts (professionals, professors, and the like, although

experts need not be formally trained; millions of teenagers, for example, know hip-hop language and culture better than most Ph.D.s). This division has persisted in librarians' writings from the 19th century to the present day; it is surely a permanent feature of user studies.

Pennants and Expertise

As we have seen, the degrees of expertise of many different people can be modeled by a single pennant diagram. Expertise is defined in large part by *specificity of knowledge,* which the pennant diagram's idf weighting renders as different levels of specificity in their vocabularies. As Sparck Jones and Willett (1997, p. 307) noted in Part 1, "The basis for IDF weighting is the observation that people tend to express their information needs using rather broadly defined, frequently occurring terms, whereas it is the more specific, i.e., low-frequency terms that are likely to be of particular importance in identifying relevant material." These authors ascribe knowledge of "broadly defined, frequently occurring terms" to people in general but do not say who knows the "more specific, i.e., low-frequency terms." They nonetheless imply the division in expertise just mentioned. That is, for any subject domain, the great majority of persons will know only the broad terms (if those), while the specific vocabulary and proper names of particular topics will be known mainly to subject experts and indexers. Pennant diagrams capture the knowledge of all three groups and serve as visual links to certain findings of user studies. To elaborate briefly:

1. *The forms of terms in pennants capture indexers' knowledge.* Indexing terms are linguistic types—i.e., distinctively formed words, phrases, and abbreviations. Human and automated indexers are instructed in permissible and impermissible types for describing writings. Types approved by ISI were seen in the pennants of Part 1 (for example, "Harter SJ" for Steven J. Harter and *"Studies Classic Am L"* for *Studies in Classic American Literature*); types approved by MLA are seen here in Figure 1. But all indexing schemes have their distinctive types (cf. "A history of French art" vs. "Art, French, History of" vs. "Art—France—History"). Indexers need not have mastered the content of writings to which the schemes refer. Rather, they specialize in the terms themselves—in subject headings, fashionable natural-language phrases, preferred forms of author names, and so on. This knowledge, while shallow, is not picayune; it radically affects all searching with term-matching technologies. Nor is such knowledge given to everyone; exact term forms are precisely what many people are ignorant of, as Bates (1998) points out. Pennants make these forms visible.

2. *The placement of pennant terms captures subject experts' knowledge.* Indexing terms multiply across texts as tokens of linguistic types—tokens whose occurrences and co-occurrences can be counted. After the counts are processed to yield pennants, sectors A and B contain relatively specific terms that are known to experts but frequently *not* known to the general public. This is not to say that the experts will always know the authorized forms of the terms in advance

(they are not indexers)—merely that, regardless of form, they can interpret their meaning, including the meaning of the broader, more polysemic terms in sector C.

Pennants also display indexing terms by their degrees of relevance to a query. They thereby model expert question-answering abilities. In other words, pennants answer certain questions that only subject experts, not indexers or the general public, could answer. A test of this claim would be to ask various persons to explain *why* a pennant shows certain answers as more relevant. Presumably, experts alone could give detailed accounts that would be borne out by other evidence. Pennants do not reveal the deeper levels of expert subject knowledge, but they convey its surface effects. Pseudo-Mary apes Mary.

3. *Generic terms in pennants capture the lay public's knowledge.* Finally, pennants model what people know when they lack specialized knowledge of indexing (as even experts do outside their own domains). The more frequently subject terms are used (probably because they have multiple meanings), the more they coincide with the ordinary language of lay publics, and the more likely they are to appear with other relatively generic terminology in sector C.

The boundary between sector B and sector C thus divides expert from lay vocabulary. When lay users do subject searches, they often use sector C type terms "off the top of their heads," when more specific terms would better describe what they really want. Even when they *combine* relatively generic terms to narrow the scope of a search (e.g., *God* AND *Monarchy*), they may still be far from expressing their real interest (e.g., *Divine Right of Kings*; see Mann, 1993, pp. 79–81). They make this mistake because they are unaware of specific indexing at sector A and B levels. (They may also believe that anything but generic terms would stump the system.) They thereby suffer many largely irrelevant responses, as pennants can indicate.

Where knowledge of titles is concerned, these people might well benefit from articles rather than books, or books rather than entire runs of journals, or works that are precisely on topic rather than works that are all things to all men. However, if they know no works by name, their guesses tend to match titles of the sort that appear in sector C—that is, oft-used noun phrases—rather than the less guessable titles of sectors A or B. Experts, in contrast, could have searched for exact titles from all sectors before ever seeing the pennant. The same holds true for pennants comprising authors: Experts know authors in all sectors; nonexperts, only the famous ones (if those). So, in a sense, pennants model the ignorance of lay publics as well as their knowledge.

Knowing titles in pennants is different from judging their relevance. Part 1 made the point that many sector A titles are topically relevant to the seed term. Their relevance is thus comparatively easy to judge, even for nonexperts, *once they are seen.* But only experts and a few indexers will know them from earlier encounters; to the public they will be new. The opposite is true of titles or authors in sector C. These may be known even to nonexperts, but, because of their breadth of implication, their relevance to the seed term will be harder for everyone to judge.

Expertise and Creative Relevance

RT makes relevance depend on the degree to which a new assumption and a preexisting cognitive context interact to yield an output different from either input—Sperber and Wilson's cognitive effect. The assumption and its context, when expressed, may be on the same topic—that is, share key phrases with the same sense and reference—and this commonality may well contribute to the effect. But they do not *have* to be on the same topic; anything that produces an effect qualifies as relevant (Sperber & Wilson, 1986, 1995, pp. 216–217). Harter (1992) found RT's openness in this regard attractive for information science, because it allows relevance judgments to be based on more than topical matches, as they are in real life. A minute's reflection will bear Harter out, as will countless exchanges from everyday conversation. For example, if I tell you how much my child's taste in music resembles my own, and you say, "The fruit doesn't fall far from the tree," you are not matching me in topic, yet your remark is certainly not irrelevant; I recognize that you are agreeing with me and are stating a belief about family influence.

The fruit-from-tree metaphor is hackneyed now, but it was once a creative reply to certain observations about children's characteristics. The larger point is that individual creativity continually affects relevance relations as people make claims (or, in the language of RT, generate new assumptions) within a context. Claims are not mere noun phrases. They employ verbs to state propositions, and propositions need not match or paraphrase the noun phrases of their context to be relevant to it. Some ways of being relevant beyond the topical are listed in Table 4. All can add assumptions to an existing context, fusing matters hitherto separate (cf. Koestler's 1964 "bisociations").

Table 4 reminds us, for example, that one way to create relevance relations is to posit a new *causal* connection between previously unrelated phenomena. Swanson and Smalheiser (1997, 1999) are known in information science for having done this in several areas of medicine. The novelist Patricia Cornwell (2002) has done it recently for a historical event, arguing that Walter Sickert, the English painter, caused the deaths of several prostitutes in Whitechapel, London, in 1888—i.e., that Sickert was Jack the Ripper. Cornwell's claim not only makes Sickert studies relevant to Ripperology; it also implicates such disparate topics as mitochondrial deoxyribonucleic acid (DNA), watermarks in letter paper, and English Channel ferry service, because documents on them can be used to strengthen or weaken her case for the Ripper's identity.

Such documents would be relevant in the *evidentiary* sense seen in Table 4. RT's definition of relevance has at its heart the idea of strengthening or weakening assumptions (and claims). Within IS, Patrick Wilson (1968, pp. 41–54) lucidly distinguished evidentiary from topical relevance in *Two Kinds of Power* almost 40 years ago. The evidentiary

TABLE 4. Topical relevance coexists with a variety of other kinds.

Types of relevance	Glosses	Examples
Topical	Writings share a subject matter (at the same level or different levels in a hierarchy of subjects).	*Moby Dick* and other books on whaling; *The Scarlet Letter* and other books on Puritanism in New England
Analogical	Parallels are drawn between concepts.	Ishmael and Queequeg = light-skinned male and dark-skinned male = Huck and Jim
Causal	Causal links are predicted between or among variables.	Fish oil, blood viscosity, and Raynaud's syndrome; magnesium deficiency and migraine
Methodological	A method is invoked in the study of a subject.	Multidimensional scaling and author cocitation analysis
Evidentiary	Examples or data are brought to bear on a claim so as to strengthen or weaken it.	Jack the Ripper = Walter Sickert —> mitochondrial DNA and watermarks in stationery
Poetic	An apt metaphor or allusion helps to make a point.	"The fruit doesn't fall far from the tree."
Terminological	Writings comment metalinguistically on language as used in other writings.	Borko (1968) defines *information science* more broadly than the present article.

sense is prominent in dictionaries; my *Webster's New Collegiate,* for instance, defines relevance as "affording evidence tending to prove or disprove the matter at issue or under discussion." And despite the Yiddish saying that *"For example" is no proof,* examples that are far from settling a matter can be relevant as evidence; e.g., the movie *Norma Rae* might be used to illustrate a claim about alienation in the American workforce, even though alienation is not an explicit topic in the movie, and the linkage requires a small creative leap.

This line of argument bears directly on judging the relevance of documents to a query. In an extended example, Harter (1992, p. 607) imagines a request for documents that deal with "online searching of bibliographic databases … the dynamics of the search process … how people do online searches … theoretical models related to the online search process that have been tested with empirical data." (This request is the context in which retrieved documents will be the retrieval system's assumptions.) He then asserts that a seemingly perfect hit—Raya Fidel's article "Online Searching Styles: A Case-Study-Based Model of Searching Behavior"— is initially only weakly relevant to his interest because it is not certain to produce a cognitive effect when read. Moreover, if its effect when read does turn out to be negligible, then it is not RT-relevant even though it matches his topic about as well as can be imagined. In contrast, half a dozen other documents that Harter describes are not, strictly speaking, on his topic of interest, yet in each case he generates fresh claims that *make* them potentially relevant—a clear example of creativity at work in a dynamic retrieval situation (Schamber, Eisenberg, & Nilan, 1990).

Of course, the creativity that individuals can contribute to discussions of particular topics will vary sharply because of differences in what they know beforehand—their expertise. If a pennant showed *Walter Sickert*, *DNA*, and *Watermarks* pulled toward the seed term *Jack the Ripper*, only those steeped in recent Ripperology would be able to interpret what is going on behind the verbal surface. In Harter's case, every document he considers is taken from a field he already knows well. Thus, he can make new interpretive assumptions with relative ease. Were he to consider the relevance of documents to an interest outside library and information science—e.g., *broken chiral symmetries* from physics or *renal brush border membrane vesicles* from biomedicine— he would probably have much greater difficulty in making creative claims and would have to fall back on topical phrase matching like every other nonexpert. His situation in both instances is typical: Expertise counts (Harter, 1996; Saracevic, 1975, p. 341).

The notions of verbal intercohesion and intercoherence, introduced in Part 1, help to explain relevance judgments like Harter's. Intercohesion occurs with the explicit repetition of words and phrases across different texts; it lends itself to computer discovery. Intercoherence is the connectibility of texts through meanings that can be read into them, regardless of whether they share vocabulary; it is less discoverable by computer. Intercohesion and intercoherence bind documents into literatures; they also bind documents and users' search statements (White, 2002).

While retrieval is actually based on intercohesion, what users really want is intercoherence—that is, documents that match what they *mean* by their requests. But what users mean evolves as they examine the output of retrieval systems (Schamber et al., 1990). Thus, in his own field, Harter can use background knowledge to make new assumptions about documents that are not strictly intercohesive with his request and thereby *construct* intercoherence. But with documents from fields in which he lacks expertise, he cannot readily enrich the context through creative assumptions and must act more as a computer does in seeking to match phrases. Creative assumptions might lead to any of the kinds of relevance seen in Table 4; phrase matching leads pretty much to the first kind, topical relevance. Swanson (1977, pp. 139–143) similarly contrasted relevance that is "created or constructed by the requester" of a literature search with relevance that is merely "a judgment by the requester that a given document does deal with the topic that he requested."

RT and Retrieval Evaluations

Although the straightforwardness of topical relevance (compared to the other kinds) has made it the classic basis for evaluating retrieval systems, many critics in IS have found this basis too narrow. Among them, Harter (1992) points to findings by McCain (1989) and Pao (1989) that, within the same subject area, retrievals based on citations have little overlap with retrievals based on descriptors, presumably because citation retrievals tap into multiple kinds of relevance (such as those seen in Table 4) while descriptor retrievals tap into topical relevance only (a point ISI has always made in promoting its indexes).

As we saw, the pennants made from ISI citation data in Part 1 confirm Harter's intuition. For example, certain novels very different in subject matter from *Moby Dick* (e.g., *The Scarlet Letter, Huckleberry Finn, The Great Gatsby*) were pulled toward Melville's novel in part because of their *analogical* relevance to it in discussions of myth and symbolism in American literature. Other cocited works in that pennant provide the aforementioned *evidence for claims* not topically related to *Moby Dick*—for example, claims about the lives of writers with whom Melville is compared. In the pennant that shows the authors cocited with White HD, some authors appear because their work is *methodologically* relevant to mine, such as Kruskal JB, one of the creators of multidimensional scaling. In general, pennants drawn from the ISI databases can supply massive data for testing hypotheses about multiple kinds of relevance in any learned or popular field.

We may conclude with Harter (1992) that citers tell us more about relevance than do judges in retrieval evaluation trials. This should not be surprising, because the two judgment situations are quite different. Citers refer to documents found over time through many different means, not just through a single literature search, and they put those documents to explicit uses that can be analyzed for motive. Judges, on the other hand, evaluate documents found by a single query (usually topical), and if they envision how those documents could be used in new contexts, it leaves no trace.

Even so, one might ask whether retrieval evaluation trials *could* capture the different kinds of relevance that authors perceive when they cite. Could the trials be made more realistic? Borlund (2003b) describes a new kind of trials that are intended to take interactive retrieval systems, evolving queries, and dynamic relevance judgments into account. But could judges routinely be asked to find at least as many kinds of relevance as appear in Table 4? The answer is almost certainly no, for reasons drawn from RT as developed here.

The first reason, foreshadowed previously, is that authors cite documents in connection with *claims* made in their prose, and the documents they cite are by and large relevant to those claims. But researchers conducting retrieval evaluation trials do not usually give judges claims (i.e., propositions, statements) to work with; they give them or let them pick *topics* (i.e., noun phrases) so that phrase-matching searches can be carried out. Nor do they instruct judges to

create claims against which to evaluate documents. Without claims, all the judges have to go on are topics, and judgments of relevance made under these two formulations will differ markedly.

Take, for example, the Fiedleresque claim that "Adult heterosexual relations are largely absent from the 19th-century American novel." It is easy to see how *Moby Dick, Huckleberry Finn, The Scarlet Letter,* and even (for purposes of contrast) a 20th-century American novel like *The Great Gatsby* are jointly relevant to this statement as various kinds of evidence. Any suitably literate judges would feel their impact as assumptions in this context. Now imagine instead that the context is simply the novel *Moby Dick,* with no particular claim about it (or with the claim being only "I want documents relevant to *Moby Dick*"). Then most retrieval-trial judges, even literary experts, would not readily link *Huckleberry Finn, The Scarlet Letter,* or *The Great Gatsby* to Melville's novel as relevant documents. (The citation-based pennant on *Moby Dick* in Part 1 revealed both kinds of outcome, with topical relevance dominating sector A and other kinds dominating sectors B and C.)

But suppose judges in document retrieval trials *were* instructed to create claims as the basis for their judgments. Could we not then go beyond topical matching?

Again no, because the instructions would be too hard to follow. Consider the kinds of relevance listed in Table 4, and imagine that researchers (as speakers) tried to give instructions so that judges (as hearers) would take all of them into account as they evaluated each document. For researchers, the task of explaining and giving examples of all these kinds of relevance would be a burdensome contradiction of speaker's economy, yet it would have to be done. Even so, the researchers might fear that their instructions were both too complex and not complex enough. And they would be right, because such instructions would surely overburden hearer's economy as well: Unless the documents to be evaluated were few indeed, it would be almost impossible for judges to create claims reflecting several different kinds of relevance and apply them to each document in a limited period. Too much expertise would be needed, too many imponderable questions would arise, and the judges would rebel.

It would not do, either, for researchers to urge the making of claims through imperatives like "Be creative! Have new insights! Connect the dots!" Instructions of this sort would strike hearers as simply fatuous, again because they cannot reliably be carried out. Like the reference librarian who is given a highly generic question to work with, the judges would have to *interview* the researchers for definitions and concrete examples of what was meant, and even then the desired actions would not necessarily be performable.

A retrieval evaluation trial resembles a reference interview in that it is a piece of business whose time to completion cannot exceed certain limits; it is seldom a long, leisurely conversation. Under these conditions, researchers must instruct judges so as to produce the desired effects with acceptable effort for themselves as speakers and for the judges as hearers. They are thus likely to give brief, broad

instructions or scenarios and hope that the judges will improvise their own specific criteria of acceptance and rejection. One way researchers imply complex criteria without spelling them out is to provide scales that make relevance a matter of degree rather than a yes-no judgment. But researchers cannot put too fine a point on what they mean by relevance if judges are to have full leeway in accepting documents on *various* grounds, as authors do in life. Consequently, the researchers will leave open the possibility of nontopical relevance judgments, but they will not require them. That is, they will *allow* claims (like the one about sex and the 19th-century American novel) if the judges happen to make them, but they will not usually *call for* perception of the different kinds of relevance one sees in Table 4.

Explaining Cuadra and Katter

The one (partial) exception I know of occurred in the classic study of Cuadra and Katter (1967). They took nine abstracts from the psychological literature and asked college seniors in psychology to judge their relevance to eight requirements statements—72 judgments on a 9-point scale. After obtaining scores, they instructed the students to pretend to be agents for persons whose "implicit use orientations" might affect relevance judgments—orientations such as an interest in the *hard data* or the *methodology* or the *practical utility* of articles. (Implicit uses are not explicit claims, but they tend in that direction.) Some 14 different orientations were described in brief scenarios, and each student was given one scenario and asked to make the 72 judgments again. All 14 orientations produced shifts in the students' original scores. The students also rated their revised judgments for difficulty. Interestingly, the scores shifted least and the revised judgments were easiest when the scenarios suggested choices based on *topical* relevance—for example, in the preparation of a selective review or an exhaustive bibliography. In contrast, the scores shifted most and the judgments were most difficult when the scenarios called for consideration of the abstracts' *methodological* relevance to the queries or their *strict terminological consistency*. This is as RT would predict, because these latter orientations require drastic rethinking of the data, which increases processing effort.

Cuadra and Katter–style reconsiderations, however, are not usually imposed on judges. To get through trials in a timely fashion, the judges tend to limit their own effort as they follow the researchers' instructions (hearer's economy). They do this by intuitively opting for the easiest criterion in Table 4, which is topical relevance. This kind of relevance does not require claims to be made (beyond "I am looking for documents on this topic"). It does not require creativity or extraordinary expertise to detect—merely some familiarity with the vocabulary in the domain being searched. It lets judges evaluate the fit between a query and a retrieved document immediately, with only the title and abstract of the document in hand. (Judgments of nontopical relevance would often require reading a document in full.) It makes the whole task performable in real time.

We have seen this conversion from broad and generic to narrow and specific before. I wrote earlier that when speakers seek nonpersonal information from someone unfamiliar, they tend to open with a generic request, thinking what is easiest for them (speaker's economy) will also place the fewest demands on the hearer. For their part, hearers often counter with a request for specifics, in order to clarify what the speaker really wants (hearer's economy). My intent was to explain why customers put overbroad questions to reference librarians and why reference librarians would prefer specific questions. But the customer-librarian exchange is in fact paradigmatic for dialogues in information science, including dialogues between researchers and judges. Both of the latter groups continually seek what is *easier for them* in their communications, as if they were automatically dividing expected effects by expected effort in their heads, as RT would predict. That is why retrieval evaluation trials so often reveal the importance of topical relevance rather than the other kinds seen in Table 4. The other kinds reflect creative associations that authors make over extended periods—unpredictable associations that researchers studying retrieval systems cannot induce on demand, especially under time pressure. Few researchers have even tried (Harter, 1996).

Objectivity of Relevance

If someone *did* spontaneously make a creative association between two documents not topically related—say, *Huckleberry Finn* and *Moby Dick*—it would be taken by many in IS as demonstrating the notorious "subjectivity" of relevance.

In fact it does not, but the point needs clarification, because confusions about whether relevance is *subjective* or *objective* have muddled IS for decades (cf. Borlund, 2003a; Schamber, 1994). Summarily put, it is not *relevance* that is subjective; it is people's *claims* that are subjective, in the sense that they may vary from person to person. Relevance to claims is always objective, although, as we have seen, it can vary in degree of specificity (cf. Wilson, 1968, p. 44). The reason so many information scientists have considered relevance subjective is that judges in retrieval evaluation trials almost never reveal the claims that underlie their relevance judgments. (Again, these claims need not be fresh or bold; most are probably on the order of "This document is [or is not] on the stated topic of interest.") Since it is quite possible for judges to assess the same document with different implicit claims (or for the same judge to assess the same document with different implicit claims at different times), different verdicts on relevance are not only possible but likely, and these different verdicts give the illusion that relevance is highly subjective in nature. The illusion also appears when a judge's implicit claim is *mistaken*: Were such a claim brought into the open, it would be seen that a document is indeed objectively relevant to it, but that other judges differ because they have not made the same error (such as misinterpreting the topic of interest).

Whatever the case, if the judges were to explain the claims behind their judgments, the illusion of subjectivity would vanish. Any qualified referee could then be asked to decide whether, given the claim being made, the document were relevant to the query. This is possible only if relevance is objective—that is, has an easily processed cognitive effect across all qualified referees, not just the judge making the initial evaluation.

I say "qualified referees" because, in learned fields, relevance judgments often require possession of specialized culture and vocabulary. By contrast, in the dialogues of ordinary life, anyone who knows the language can judge the relevance of anything. Were it otherwise, we could not correctly infer meanings from our own or other people's conversations, or follow movies and novels (e.g., McEwan, 1998, p. 79), or understand Sperber and Wilson's examples using Peter and Mary. If relevance were not objective, judges in *real* trials—the fateful kind in law courts, as opposed to document retrieval trials—could never rule on a challenge to an argument on grounds of irrelevance:

Prosecution lawyer: Objection, Your Honor. Where is counsel going with this?

Defense lawyer: Your Honor, I think the way my client has paid his gambling debts is material to the charge of embezzlement.

Judge: I'll allow it; overruled. But keep to that point.

Or suppose someone in a retrieval evaluation trial claims, as earlier, that *Huckleberry Finn* is relevant to *Moby Dick*. If challenged by a critic to explain how, the person could relate the Huck/Jim pair to the Ishmael/Queequeg pair as archetypes in American literature (á la Fiedler, 1966), and anyone could perceive that, on that level, they strengthened the claim. The relevance of the explanation would be objective if it produced appropriate cognitive effects in anyone, as presumably it would. But note again that, in a typical retrieval evaluation trial, there are no challenges to relevance judgments and no third-party checks on them; the judges simply turn in their scoring sheets without saying anything about the grounds for their evaluations. Hence, on the surface, everything seems "subjective," especially their disagreements.

A factor that complicates the subjective/objective distinction is the novelty of messages (Barry & Schamber, 1998; Chen & Yu, 2005; Furner, 2004, pp. 440–441). As noted in Part 1, people do vary on what is new to them, and newness does affect cognitive impact. Objective relevance in the sense of "strengthening or weakening claims" is a logical relationship independent of novelty (Swanson, 1986; Wilson, 1973).

RT and Indexing

To return to our main thread, least-effort arguments from RT explain not only why reference interviews, online searches, and retrieval evaluation trials are as they are, but also why indexers do what they do. Table 5 gives the correspondences among roles. Once again, what minimizes effort for speakers is not what minimizes effort for hearers. In this

TABLE 5. Pairs exemplifying least-effort effects.

Setting	Communicating pairs
Reference desk	Customers and librarians
Online retrieval task	Users and systems
Retrieval evaluation trial	Researchers and judges
Indexing service	Supervisors and indexers

final example, the generic statements of supervisors of indexing services are less relevant than they could be for indexers in those services, who want specifics. A quick sketch will show what I mean.

In naming the subject matter of documents with noun phrases, indexers work under broad instructions from their supervisors, but a great deal is left to their discretion. For example, in his influential Library of Congress (LC) guidelines for indexing books, Haykin (1951, pp. 7–11) tells subject catalogers to assign the subject heading "that the reader will seek in the catalog" and that also is "as specific as the topic it is intended to cover." How this is to be done is left open—a move that belies the true complexities of indexing (cf. Drabenstott & Vizine-Goetz, 1994, pp. 3–14). Haykin implies that there is a typical reader whose information-seeking vocabulary is known, that matching books and LC subject headings on a specificity scale is not problematical, and that the typical reader will regularly look for a book under the most specific heading available. In fact, readers regularly enter library catalogs with terms *less specific* than their interests, according to research cited in Drabenstott and Vizine-Goetz (1994, p. 13). That is what RT as developed here would predict.

The lack of elaboration, the baldness, of what Haykin says is another instance of speaker's economy. The contrasting hearer's economy allows the indexer, in this case a subject cataloger of books, to get on with a complex job in the face of guidelines that underspecify what is needed. Like the reference librarian and the retrieval evaluation judge, the indexer wants well-defined procedures that can be carried out in real time, and he or she will supply them if they are not forthcoming from supervisors. Sometimes, of course, supervisors *do* supply them, and then indexers may use them to speed work along even when the choices they lead to are suboptimal (as illustrated later with some books by William F. Buckley).

The indexer's more-or-less permanent situation is *document in hand, reader nowhere in sight.* Sperber and Wilson's words from RT once again apply: "when expectations of effect are wholly indeterminate, the mind should base itself on considerations of effort: pick up from the environment the most easily attended stimulus, and process it in the context of what comes most readily to mind." For the indexer, the most easily attended stimulus is the document in hand; the context most readily in mind is the thesaurus (or similar bibliographic tool) on the desk; and considerations of effort show themselves in thesaurus-to-document mappings that are readily justifiable to anyone, such as subject headings

that differ little from titles or subtitles. Thus, the indexers' operationalization of Haykin's guidelines might be something like "Copy or paraphrase front-matter as much as possible." If indexers routinely make generic instructions specific in this fashion, they need not constantly ask their supervisors, "How shall I handle this? What do you mean by that?" as if they were reference librarians interviewing vague customers.

An alternative suggested by Maron (1982, p. 104) is for the indexer to try to guess what proportion of readers using a query term will want a document if it is indexed by that term. But probabilistic indexing in this vein is obviously much harder to do than simply reexpressing the title or other front-matter with terms from a controlled vocabulary (cf. Blair, 1990, pp. 253–254). It is also harder to check: Indexing services do not poll readers as to their preferences for particular subject headings either before or after a document is indexed. It seems safe to conclude that "the reader" is a very diffuse force in indexing, whereas the thesaurus and in-house conventions are very strong. The one actionable rule an indexer might draw from Haykin's "reader as focus" principle is to prefer vernacular over learned headings—e.g., English over Latin names for biological species—when there is a choice.

As hearers of broad instructions, indexers will do what is easier for them while still attempting to index acceptably. When they in turn become speakers—that is, communicators whose utterances are subject headings—they will again "seek to engage the most easily attended audience with whatever most readily comes to mind." The most easily attended audience in their case is not readers, who are far away, but the supervisors responsible for index quality control, who are immediately present. If the latter approve what the indexers do with whatever indexing terms can be readily assigned, that is the approval that counts.

In practice, this means that indexers will choose terms that supervisors can easily check against the document for closeness of sense—topical relevance again. It means that indexers will follow supervisors' in-house rules for handling indexing problems rather than trying to guess what some ideal-typical reader might want. It also means that indexers will not choose terms whose bearing on the document requires creative expertise to perceive. For example, they will not read, ponder, and describe documents to bring out nonobvious properties, such as their evidentiary relevance to claims in other documents. (Only citing authors will do this.) They will *certainly* not tag documents with epigraphs from classic writers to poeticize themes, as George Eliot did with the chapters of some of her novels. Indexers are thus in the same boat with the retrieval judges who would not and could not routinely extend their criteria to include the nontopical kinds of relevance seen in Table 4.

Such an account, centered on least effort, can be extended to explain two frequent outcomes in indexing. The first is describing documents with terms that are simply copied or paraphrased from full-text (Montgomery & Swanson's, 1962 "machinelike indexing by people"). The second is skimping on what is copied (as seen, for example, in many back-of-the-book indexes). These outputs have indexers' implicit guarantees of their optimal relevance, but readers are likely to find them perfunctory at best, thereby playing out the speaker-hearer opposition once more.

The notion that indexers make supervisors rather than readers their primary audience accounts for certain idiosyncrasies in controlled-vocabulary systems. For instance, the few Library of Congress subject headings assigned to William F. Buckley's books on ocean sailing always include the names of his yachts. LC practice allows catalogers to treat sailing vessels as corporate bodies and name them in the subject field of bibliographic records. This lends itself to a passive-aggressive "working to rule" with respect to Haykin: "You want the most specific heading for Buckley's reflections at sea? Okay, what's more specific than 'Cyrano (Yacht)'?" While readily copiable from the book in hand, such least-effort headings not only fail to cover Buckley's true range of topics; they would also never occur to most readers as search terms. Nevertheless, supervisors approved them as relevant communications.

If indexers are offering "some minimal, easy-to-comprehend stimulus toward being understood," so as to "see whether the response has a satisfactory payoff," then supervisors' approval is their payoff. Further examples of dubious but approved indexing are not hard to find. Supervisors seem disinclined to impose searches for better terms, especially when it is debatable what those terms might be. Their goal, understandably enough, is to keep indexers happy and catalog copy flowing. (This encourages the "mark it and park it" mentality well known in American book classification.) As long as utterances about subject matter can be justified on *some* ground, it is presumed that "the reader" will see their relevance. Real readers hardly enter the picture.

Summing up, the exchanges in Table 5 generally take place between nonintimates or even strangers. It appears that in relatively impersonal communications relevance is often sought by minimizing effort rather than by maximizing effect. In many contexts in IS (not just those implied by Table 5), more effective communication would require more effort from all concerned. We can begin to explain why this is so—why information systems and their users so often underperform—by extending ideas from RT to the exchanges studied in IS. The most powerful idea from RT is the effect/effort definition of relevance. The present article shows how this idea can account for several types of behavior important to IS in a consistent way.

The effect/effort ratio also has an important implication for systems design: For many persons, any extra work involved in interacting with an information system will reduce its benefits and its likelihood of being used (Hardy, 1982; Mann, 1993, pp. 91–101). Without doubt, people can put prodigious effort into impersonal communications when they so choose. But these are usually communications for which they receive byline credit (e.g., writing, teaching, acting, broadcasting), whereas many of the communicative acts of interest to IS—indexing, reference interviews, relevance judgments, database searches, etc.—are essentially anonymous.

Bylines or financial rewards or both are necessary if people are to exert themselves very much as they interact with information systems.

Psychologizing Bradford

Having said so much about processing effort, the denominator in the RT ratio, I must revisit the numerator, cognitive effects. At the same time, I will be returning to human-system dialogues and pennant diagrams.

In Part 1, I equated the items at the high end of the cognitive effects scale, or toward the tip of the pennant, with the most productive sources in the core of a Bradford distribution. Using Bradford's own example of journals that yield articles on a subject, I wrote, "In Sperber and Wilson's language, the journals at the high end of the scale (the 'core' journals) produce their greatest effects in the context of a subject term and hence are most relevant to it. More precisely, they are the system's assumptions as to what is most relevant." Putting a psychological spin on bibliometric data may have seemed somewhat unusual. However, the idea is long familiar in information retrieval, where relevance rankings of documents are based on data that in another frame of reference bibliometricians would call their own. The Google Scholar service, for example, seeks to maximize relevance by ranking research documents high to low by their frequency of citation on the Web.

In general, when a system puts rank-ordered bibliometric distributions on display, they can be taken as artificially intelligent *answers to questions,* whatever other meanings they have. The top-ranked (highest-weighted) items are attempts by the system designers to estimate which replies (from among the many possible) will be most relevant to whoever caused the rankings to be made. More generally, bibliometric rankings on the system side may be seen as attempts to produce cognitive effects on the user side. This account squares with Sperber and Wilson's (1986, 1995, p. 123) claim that "at a general level, we want to compare the concept of relevance to concepts such as *productivity* or *yield,* which involve some form of cost-benefit analysis."

A literal ranking of term weights to express *productivity* or *yield* is not the way persons achieve relevance in answering. But when human designers try to give question-answering abilities to an artificial intelligence, ranking is among the best tools they have available. That is because displays of what is higher or lower on some scale (or the associated comparative and superlative statements) generally affect people cognitively if the context is right. Ranking also allows items below a certain threshold to be ignored, thus reducing processing effort, thereby increasing relevance. An AI that ranks answers to bibliographic questions may thus serve human relevance seekers better than one that gives the same answers in, say, alphabetical order, especially if the rankings are offered as replies in real time. (Alphabetizing makes for easy lookups, but ranking is better for human interest.)

In Bradford's day, the compilation of a subject bibliography took weeks, and he probably did not foresee systems like pseudo-Mary that can rank journals instantly. Acting more as a scientist than a librarian, he published his ideas on the mathematical form of the ranked yield distribution but never revealed what the core journals in lubrication and applied geography actually were. It therefore seems unlikely that he thought of his core-and-scatter zones as producing greater to lesser cognitive effects in the RT sense. Yet that is how we can now construe them.

Take, for example, the proposal in White (1981) that systems like Dialog display retrieved documents not by accession number (the default) but by journal in descending order of yield. I realize now that what I called "Bradfordizing" search results was an intuitive attempt to heighten their cognitive effects. The subsequent emergence of Dialog's Rank command (which can do what I asked for) and of comparable capabilities in other systems suggests that the cognitive effects of "Bradfordized" bibliographic data are increasingly perceived. Bates (2002) observes that Bradford's three zones of journal yield call for different search strategies; journals in the core zone, but not the others, can productively be *browsed,* and that capacity presumably is a mark of their greater cognitive payoff.

Harter (1992, p. 611) and Furner (2004, pp. 440–441) shrewdly equate relevance with *informativeness,* the experiencing of cognitive change. This accords not only with RT's notion of relevance as cognitive effects but also with standard IS notions of "information as process" (Buckland, 1991)—that is, changes in one's stock of knowledge (Brookes, 1975) or one's image of the world (Boulding, 1956; Pratt, 1982). Bateson (1972, p. 453) famously called information in this sense "a difference that makes a difference."

Consistently with these ideas, Tague-Sutcliffe (1995, pp. 50–51) notes that documents can be ranked so that their potential informativeness—their power to cause cognitive change—decreases logarithmically from higher ranks to lower. In her abstract formulation: "When records are ordered according to nonincreasing user preference (i.e., with the most informative records first but with ties in ordering permitted), the informativeness of a subsequence of records is approximately proportional to the logarithm of the number of records in the subsequence." She mentions the Bradford distribution, "a logarithmic law of diminishing returns," as one piece of evidence for her claim. Others she mentions are the logarithmic relationships between, first, the number of retrieved documents needed to provide a specified number of relevant documents in literature searches (Salton, 1975); and, second, the number of new papers and the number of new ideas in a specialty in mathematics (May, 1968). Echoing B. C. Brookes (1980a, 1980b), she goes on to say, "A logarithmic law is consistent with the Weber-Fechner law that the magnitude of a psychological response to a stimulus varies directly as the logarithm of the magnitude of the stimulus. Brookes suggests that if, as claimed by the Weber-Fechner law, our sensory mechanisms behave according to a logarithmic law, it is possible that all of our neural mechanisms

behave in that way." The similarities she notes appeared a quarter-century earlier in Fairthorne (1969).

All this, of course, does not guarantee that high-ranking answers on the system side will *have* cognitive effects, merely that system designers intend or predict they will. But they evidently do have them. The popularity of Google and other relevance-ranking search engines (despite their flaws) indicates that positive effects in the RT sense frequently occur. Well-known writers like Thomas L. Friedman and Richard A. Posner nowadays gauge the psychological effects of a term of interest by noting how many hits it produces in Google (cf. Levine's, 1977, idea of "whumps"). Or consider ranked impact factors—that is, normalized counts of the citations received by journals, institutions, and authors. The term *impact* implies the ability to produce cognitive changes in individuals. If not, why are there so many discussions of impact factors in the literature and on the listservs? Why among proponents are high factors greeted with delight? Why among opponents is there so much concern that they will mislead?

Regarding the last question, a brief example. Critics of citation analysis have long warned that ranked citation counts are untrustworthy measures of quality. In recent years, after researchers reported that citation-based rankings of British university departments correlated nicely with rankings derived from the laborious peer reviews of the Research Assessment Exercises (Oppenheim, 1997; Smith & Eysenck, 2002), a political counteroffensive blocked the use of such counts except on a voluntary basis (Warner, 2003). As I write, bibliometric measures seem to have made a comeback (MacLeod, 2006). But it is precisely because the counts are seen as highly relevant—impactful—in this context that their truthfulness was challenged: many academics seem loath to admit that an automated system—some version of pseudo-Mary—could replace them as judges of departmental quality.

This brings up the interplay of relevance and truth. In Gricean pragmatics, from which RT descends, the relevance and truthfulness of utterances are independent of each other (see the famous maxims in Grice, 1975, pp. 45–46; see also Wilson, 1973, p. 462; Furner, 2004, pp. 440–441). Thus, the most dangerous utterances would be lies or errors that are highly relevant to innocent hearers. (Anticitationists presumably think citationists are wrong-headed, not lying.) Early RT folded truthfulness into relevance; that is, Sperber and Wilson (1986) argued that a true statement ultimately has greater cognitive effects, and so is more relevant than a false or misleading one. The second edition of Sperber and Wilson (1995, pp. 263–266) goes some length toward restoring independent status to truthfulness (cf. Yus Ramos, 1998, passim).

While pseudo-Mary cannot deliberately lie, there is no doubt she can mislead if her utterances are not qualified as necessary by a human interpreter. Some years ago I published a list of information scientists ranked by their contributions to *JASIST* as covered by *Library and Information Science Abstracts* (White, 1990). The top three names were Lois F. Lunin (20 items), Abraham Bookstein (10), and Gerard Salton (10). You may be sure my first thought was to explain why Lunin appeared where she did (she was a featured editor of *JASIST*'s Perspectives series, not a researcher like the other two), so that a list potentially relevant to readers in information science should not give the wrong impression. Similarly, in White (2004) I made the obvious point that Foucault is not more important than Shakespeare, despite recent citation evidence to the contrary. In Part 1, I adjusted citation counts for Melville's novels *Pierre* and *Billy Budd* and for Patrick Wilson and T. H. Wilson in an attempt to record their use histories more accurately.

These are my own small instances of what might be called "designer's conscience." Given the tendency in information science to depersonalize systems, one must stress that systems are simply fronts for people who are no less human—or conscientious—than the users on the other side. Salton did not want his SMART system to be dumb; Brin and Page do not want Google to deluge people with false drops (or even too many hits). Designers of systems like these are doing the best they can to communicate relevantly and truthfully and nonoverwhelmingly, even when they fail in the attempt. As such, their exchanges with users are as explicable by RT as conversations between Peter and Mary—probably more so, because system-user dialogues are not colored by intentions like irony, metaphor, evasiveness, and deliberate deception that complicate interpersonal talk.

The Lubrication Literature

To sum up boldly: Given the right context, bibliometrics predicts psychometrics. The higher the measure of impact on the system side, the higher the predicted cognitive effects on the user side. This can be illustrated with a pennant display of journals that publish the lubrication literature, in imitation of Bradford's (1934) study. Figure 2 presents journals from Scisearch (Dialog File 34), ranked by the number of their articles that have the word *Lubrication* in their titles. To reduce overlapping labels, only journals with at least five articles are mapped.

Lubrication is a *descriptor* in both INSPEC and Compendex EI, and the wider literatures retrieved by descriptors, as opposed to title terms, would be closer than my Scisearch retrieval to Bradford's original compilation. Unfortunately, INSPEC and Compendex EI list journal titles under multiple allonyms, and such data when ranked by Dialog make for confusing pennants.

The journals with the greatest predicted cognitive effects in Figure 2 are largely those on top in a Bradford-style ranking of the data. The major difference lies in the way the pennant's idf factor stratifies them: Those with relatively high proportions of *Lubrication* articles to total articles move up, and those for which that proportion is relatively low move down. (Examples of the latter are *Wear, Journal of Materials Processing Technology,* and especially *Abstracts of Papers of the American Chemical Society.*) The extremes of statistical
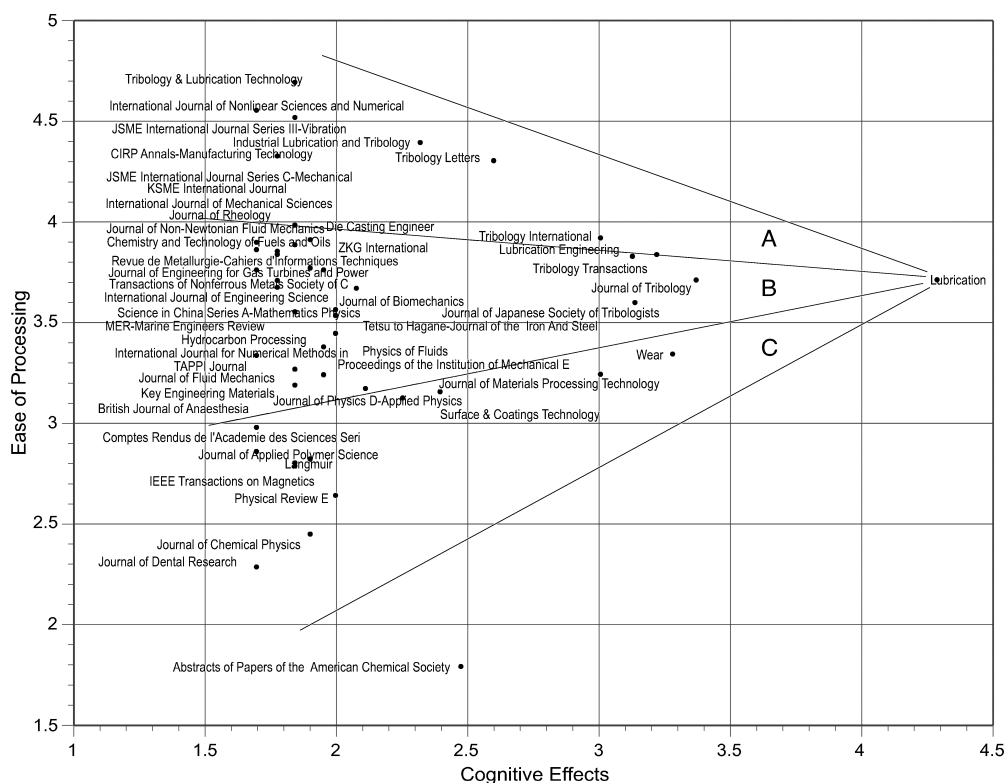
FIG. 2.   Pennant diagram of Scisearch journals that publish articles with *Lubrication* in their titles.

specificity are clear: The journals in sector A and rightmost in sector B tend to have *Lubrication* or *Tribology* in their names (*Webster's New Collegiate* defines *tribology* as "a science that deals with the design, friction, wear, and lubrication of interacting surfaces"), while the journals in sector C tend to encompass far broader sciences and technologies (physics, chemistry, mechanical engineering). The relevance of the more specific journal titles to the seed term is easier to see, and these journals would also be easier to scan for articles with *Lubrication* in their titles because they have relatively fewer total articles.

The main point here, however, is that the pennant focuses attention on the identities of the most relevant journals. In other words, pseudo-Mary makes a prediction about what will most interest the person who asked for the pennant, and this is the same as predicting what will have the greatest cognitive effect in the context of the question "What journals are most relevant to the study of lubrication?" The *Journal of Tribology* wins the horse race, so to speak. The abstract Bradford zones of journal yield are implicitly present but would be less relevant to most people than the answers explicitly advanced.

Although pseudo-Mary is not a person and cannot be held to standards of personal responsibility, she is doing her (designer's) best to give a (user's) question a reasonably relevant answer. Again, this relevance is measured on the *system* (or designer's) side; it amounts to a forecast that the user will find the results worth heeding. Given all that can go wrong, designers

might be unwilling to guarantee the "optimal relevance" of their results as called for by RT, but they could not exhibit their systems for long without some commitment to quality.

The hole in this account is the absence of measures on the user side by which pseudo-Mary's performance can be judged. Few present readers are likely to experience strong cognitive effects from Figure 2, since tribology is not part of their mental furniture. (Little expertise = wrong context.) But it is easy to imagine inquirers for whom it *would* produce strong effects—the editors of certain journals, for example, or serials librarians in certain libraries. Probably the best judges of Figure 2 would be tribologists. Since the figure would presumably be at least somewhat relevant to them, they could immediately consider how truthful they thought it. For example, one can imagine them nodding in agreement, or snorting with disbelief, or hastily explaining why the *Journal of Tribology* won (as I did with Lois Lunin).

How could the pennant in Figure 2 be evaluated? How could all such pennants be evaluated as responses to questions from users? First and foremost, pennants need to make good sense overall. My impression is that the pennants in Parts 1 and 2 could be called sensible communications. They need editing, of course, but the advice they convey seems much closer to being good than bad, and that is an important outcome. More particularly, they model various kinds of bibliographic knowledge and make concrete predictions about them. Those predictions could in fact be tested with users, with all the techniques available to IS.

One possibility: Just as pennants involve domain analysis on the system side, their evaluation could involve domain analysis on the user side (Palmer, 1999). On the system side, domain analysis reveals the terms associated with a seed term, ranked by predicted relevance. The equivalent on the user side would show the terms that a set of qualified judges associate with a seed term presented as a stimulus. For example, given *Lubrication* as a seed term, tribologists could be asked to produce a list of journals they associate with that topic. Their aggregated responses would yield a list of journals that could be rank-ordered by frequency of mention, and this could be directly compared with the ranked frequency list used to create the pennant. White (2005) proposes that term-association lists drawn from people be called *associagrams* and term-association lists drawn from literatures be called *bibliograms*. The question here is thus the degree of match between an associagram and a bibliogram.

High agreement on top-ranked terms would indicate that the pennant captures the tribologists' composite mental model of journals in their field. Low agreement would indicate not necessarily that pseudo-Mary is wrong, but that the system view differs from the human view. If the two are to converge, the criteria implicit in human judgments would have to be made explicit to the system. But the system might also be delivering information that users should allow to alter their preconceptions. In certain contexts, "Bradfordizing" systems like pseudo-Mary could be at least as expert, and as worthy of attention, as any human informant. The ultimate test is whether users value her accordingly. The fact that she, a thing, articulates a point of view not universally respected—that of the bibliometric record—makes her seem, paradoxically, a bit like a person. Call in the science fiction writers.

Key Variables of IS

The notion of pseudo-Mary as an informant is a good one on which to close, since it leads directly to a consideration of the key variables in information science. What do we want from a human informant? We want replies to our questions that are RT-relevant in a context—that is, replies that are impactful (which may mean being new and on topic), easy to process, and true. Though RT differs from Gricean pragmatics in important respects, it does no injustice to RT to say that we want replies that conform to Grice's conversational maxims (cf. Blair, 1992, pp. 204–205, 2003, pp. 27–29; Chen & Xu, 2005). Such replies are what, for example, journalists want from politicians but often do not get: replies that speak to the topic and address it in sufficient detail (hence glum reports such as "The President unveiled his health plan today but gave few specifics"). The politician-journalist mismatch is predictable; when politicians speak, least effort consists in not putting too fine a point on matters so as to deflect opposition and maximize wiggle room; when journalists hear, least effort consists in getting the juicy details immediately so that stories with high news value—high relevance to many people—almost write themselves. The project for in-

formation science is to create artificial mouthpieces for literatures that communicate not like politicians but like a cross between a highly responsive press secretary and a super-knowledgeable librarian (cf. Fallows, 2005).

This ideal level of relevance means that information science is essentially concerned with the following global variables, all discussed in this and the preceding article, as well as thousands of others:

- Topicality, in the sense of intercohesion and intercoherence among texts. The relations that bind writings into literatures and the relations that bind question texts to answer texts are both fundamentally important.
- Cognitive effects of texts in response to people's questions. Impact measures from bibliometrics are subsumable.
- People's levels of expertise, as a precondition for cognitive effects. Expertise shapes other situational variables, such as the activities in which questions are put, required response times, and depth of answers sought.
- Processing effort as textual or other messages are received. (Effort may explain *resistance* to systems.)
- Specificity vs. generality or vagueness of terms as it affects processing effort. (For systems that respond in sentences rather than noun phrases, *brevity* also affects processing effort and hence relevance. This motivates the research on summarization of texts that has long been a part of IS.)
- Relevance, defined in RT as the effects/effort ratio.
- Authority of texts and their authors, a measure separable from relevance. It may stand as a proxy for the trustworthiness or truthfulness of texts, which are much harder variables to operationalize.

These interrelated variables derive from our expectations when we seek answers from persons; the difference is that we now seek answers from writings through systems. As shown here, such systems are constrained by a user model with definite particulars from RT. Around this basic idea, much work in information science can be organized. Findings in information retrieval, bibliometrics, and user studies are all part of the research on *literature-based answering* that defines the field.

References

Barry, C.L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45, 149–159.

Barry, C.L., & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. Information Processing & Management, 34, 219–236.

Bateman, J. (1998). Changes in relevance criteria: A longitudinal study. In Proceedings of the 61st ASIS Annual Meeting (pp. 23–32). Medford, NJ: Information Today.

Bates, M.J. (1998). Indexing and access for digital libraries and the Internet: Human, database, and domain factors. Journal of the American Society for Information Science, 49, 1185–1205.

Bates, M.J. (2002). Speculations on browsing, directed searching, and linking in relation to the Bradford distribution. In H. Bruce, R. Fidel, P. Ingwersen, & P. Vakkari (Eds.), Emerging frameworks and methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS 4) (pp. 137–150). Greenwood Village, CO: Libraries Unlimited.

Bates, M.J., Wilde, D.N., & Siegfried, S. (1993). An analysis of search terminology used by humanities scholars: The Getty Online Searching Project report no. 1. Library Quarterly, 63, 1–39.

Bateson, G. (1972). Steps to an ecology of mind. New York: Ballantine.

Belkin, N.J. (2005). Anomalous state of knowledge. In K.E. Fisher, S. Erdelez, & L.E.F. McKechnie (Eds.), Theories of information behavior (pp. 44–48). Medford, NJ: Information Today.

Blair, D.C. (1990). Language and representation in information retrieval. Amsterdam: Elsevier.

Blair, D.C. (1992). Information retrieval and the philosophy of language. Computer Journal, 35, 200–207.

Blair, D.C. (2003). Information retrieval and the philosophy of language. Annual Review of Information Science and Technology, 37, 3–50.

Blakemore, D. (1992). Understanding utterances: An introduction to pragmatics. Oxford: Blackwell.

Borgman, C.L. (1989). All users of information retrieval systems are not created equal: An exploration into individual difference. Information Processing & Management, 25, 237–251.

Borko, H. (1968). Information science: What is it? American Documentation, 19, 3–5.

Borlund, P. (2003a). The concept of relevance in IR. Journal of the American Society for Information Science and Technology, 54, 913–925.

Borlund, P. (2003b). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. Information Research 8(3). Retrieved April 8, 2006, from http://www.informationr.net/ir/8–3/paper152.html

Boulding, K. (1956). The image: Knowledge in life and society. Ann Arbor: University of Michigan Press.

Bradford, S.C. (1934). Sources of information on specific subjects. Engineering, 137, 85–86.

Brookes, B.C. (1975). The fundamental equation of information science. In Problems of Information Science, FID 530 (pp. 115–130). Moscow, Russia: VINITI.

Brookes, B.C. (1980a). Measurement in information science: Objective and subjective metrical space. Journal of the American Society for Information Science, 31, 248–255.

Brookes, B.C. (1980b). Information space. Canadian Journal of Information Science, 5, 199–211.

Buckland, M.K. (1991). Information as thing. Journal of the American Society for Information Science 42, 351–360.

Chen, Z., & Xu, Y. (2005). User-oriented relevance judgment: A conceptual model. In Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05). IEEE Computer Society. Retrieved April 7, 2006, from http://csdl2.computer.org/comp/proceedings/hicss/2005/2268/04/22680101b.pdf

Cornwell, P. (2002). Portrait of a killer: Jack the Ripper: Case closed. New York: Putnam.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. Information Processing & Management, 31, 191–213.

Cuadra, C.A., & Katter, R.V. (1967). Opening the black box of "relevance." Journal of Documentation, 23, 291–303.

Drabenstott, K.M. (1991). Online catalog user needs and behavior. In N. Van Pulis (Ed.), Think tank on the present and future of the online catalog: Proceedings (RASD Occasional Papers No. 9, pp. 59–83). Chicago: Reference and Adult Services Division, American Library Association.

Drabenstott, K.M. (2003). Do nondomain experts enlist the strategies of domain experts? Journal of the American Society for Information Science and Technology, 54, 836–854.

Drabenstott, K.M., & Vizine-Goetz, D. (1994). Using subject headings for online retrieval: Theory, practice, and potential. San Diego, CA: Academic Press.

Eichmann, T.L. (1978). The complex nature of opening reference questions. RQ, 17, 212–222.

Fairthorne, R.A. (1969). Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction. Journal of Documentation 25, 319–343.

Fallows, J. (2005, June 12). Enough keyword searches: Just answer my question. New York Times, p. BU3.

Fiedler, L.A. (1966). Love and death in the American novel (Rev. ed.). New York: Stein and Day.

Franzen, J. (1992). Strong motion. New York: Farrar, Straus, Giroux.

Froehlich, T.J. (1994). Relevance reconsidered—towards an agenda for the 21st century: Introduction to special topic issue on relevance research. Journal of the American Society for Information Science, 45, 124–134.

Furner, J. (2004). Information studies without information. Library Trends, 52, 427–446.

Goatley, A. (1997). The language of metaphors. London and New York: Routledge.

Google. (2004). Google myths and realities. Formerly at http://www.google.com/jobs/great-people-needed.html

Green, R. (1995). Topical relevance relationships: I. Why topic matching fails. Journal of the American Society for Information Science, 46, 646–653.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and semantics: Vol. 3. Speech acts (pp. 41–58). New York: Academic Press.

Hardy, A.P. (1982). The selection of channels when seeking information: Cost/benefit vs. least effort. Information Processing & Management, 18, 289–293.

Harter, S.P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43, 602–615.

Harter, S.P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. Journal of the American Society for Information Science, 47, 37–49.

Haykin, D.J. (1951). Subject headings: A practical guide. Washington, DC: Library of Congress.

Hirsch, S.G. (2004). Domain knowledge and children's search behavior. In M.K. Chelten & C. Cool (Eds.), Youth information-seeking behavior (pp. 241–270). Lanham, MD: Scarecrow Press.

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. Journal of the American Society for Information Science, 46, 400–425.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. In Proceedings of the Ninth International WWW Conference. Retrieved April 10, 2006, from http://www9.org/w9cdrom/81/81.html

Horn, L.R. (1984). Toward a new taxonomy for pragmatic reference: Q-based and R-based implicature. In D. Schiffrin (Ed.), Meaning, form, and use in context: Linguistic applications (pp. 11–42). Washington, DC: Georgetown University Press.

Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.

Janes, J.W. (1994). Other people's judgments: A comparison of users' and others' judgments of document relevance, topicality, and utility. Journal of the American Society for Information Science, 45, 160–171.

Jansen, B.J. (2000). The effect of query complexity on Web searching results. Information Research 6(1). Retrieved April 8, 2006, from http://InformationR.net/ir/6–1/paper87.html

Jansen, B.J., & Pooch, U. (2001). A review of Web searching studies and a framework for future research. Journal of the American Society for Information Science and Technology, 52, 235–246.

Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing and Management, 36, 207–227.

Klinkenborg, V. (2003, November 12). Trying to measure the amount of information that humans create. New York Times, p. A20.

Koestler, A. (1964). The act of creation: A study of the conscious and unconscious in science and art. New York: Macmillan.

Lancaster, F.W. (1968). Information retrieval systems: Characteristics, testing, and evaluation. New York: Wiley.

le Roux, A. (2003). False drop: Search engine dictionary. Retrieved April 7, 2006, from http://www.searchenginedictionary.com/

Levine, M.M. (1977). The informative act and its aftermath: Toward a predictive science of information. Journal of the American Society for Information Science, 28, 101–106.

MacLeod, D. (2006, March 26). Research exercise to be scrapped. Guardian Unlimited. Retrieved April 7, 2006, from http://education.guardian.co.uk/RAE/story/0,,1737082,00.html

Mann, T. (1993). Library research models: A guide to classification, cataloging, and computers. New York: Oxford University Press.

Maron, M.E. (1982). Probabilistic approaches to the document retrieval problem. In Proceedings of the Fifth Annual ACM Conference on Research and Development in Information Retrieval (pp. 98–107). New York: Springer-Verlag.

May, K.O. (1968). Growth and quality of the mathematical literature. Isis, 69(199), 363–371.

McCain, K.W. (1989). Descriptor and citation retrieval in the medical behavioral sciences literature: Retrieval overlaps and novelty distribution. Journal of the American Society for Information Science, 40, 110–114.

McEwan, I. (1998). Amsterdam. New York: Anchor Books.

Mizzaro, S. (1997). Relevance: The whole history. Journal of the American Society for Information Science, 48, 810–823.

Montgomery, C., & Swanson, D.R. (1962). Machinelike indexing by people. American Documentation, 13, 359–366.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy and archaeology. Journal of Documentation, 53, 477–487.

Paisley, W.J. (1968). As we may think, information systems do not. Stanford, CA: Institute for Communication Research. (ERIC Document Reproduction Service No. ED037095)

Palmer, C.L. (1999). Aligning studies of information seeking and use with domain analysis. Journal of the American Society for Information Science, 50, 1139–1140.

Pao, M.L. (1989). Retrieval effectiveness by semantic and citation searching. Journal of the American Society for Information Science, 40, 226–235.

Powers, R. (1991). The gold bug variations. New York: Morrow.

Pratt, A.D. (1982). The information of the image. Norwood, NJ: Ablex.

Reitz, J.M. (2004). Information science: Dictionary for library and information science. Westport, CT: Libraries Unlimited. Also retrieved April 11, 2006, from http://lu.com/odlis/odlis_i.cfm

Ruthven, I., & Lalmas, M. (2003). A survey on the use of relevance feedback for information access systems. Knowledge Engineering Review, 18, 95–145.

Salton, G. (1975). Dynamic information and library processing. Englewood Cliffs, NJ: Prentice-Hall.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 26, 321–343.

Saracevic, T. (1991). Individual differences in organizing, searching, and retrieving information. Proceedings of the American Society for Information Science Annual Meeting, 28, 82–86.

Saracevic, T. (1996). Relevance reconsidered. In P. Ingwersen & N.O. Pors (Eds.), Integration in perspective: Proceedings of the Second International Conference on Conceptions in Library and Information Science (CoLIS 2) (pp. 201–218). Copenhagen: Royal School of Librarianship.

Schamber, L. (1994). Relevance and information behavior. Annual Review of Information Science and Technology, 29, 3–48.

Schamber, L., Eisenberg, M.L., & Nilan, M.S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26, 755–776.

Schank, R.C. (1997). "I'm sorry, Dave, I'm afraid I can't do that": How could HAL use language? In D.G. Stork (Ed.), HAL's legacy: 2001's computer as dream and reality (pp. 171–190). Cambridge, MA: MIT Press.

Smith, A., & Eysenck, M. (2002). The correlation between RAE ratings and citation counts in psychology. Technical Report, Department of Psychology, Royal Holloway, University of London. Retrieved April 7, 2006, from http://cogprints.ecs.soton.ac.uk/archive/00002749/

Sparck Jones, K. (1981). The Cranfield tests. In K. Sparck Jones (Ed.), Information retrieval experiment (pp. 256–284). London: Butterworths.

Sparck Jones, K. (1991). The role of artificial intelligence in information retrieval. Journal of the American Society for Information Science, 42, 558–565.

Sparck Jones, K., & Willett, P. (1997). Techniques. In K. Sparck Jones & P. Willett (Eds.), Readings in information retrieval (pp. 305–312). San Francisco: Kaufmann.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Oxford: Blackwell.

Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition (2nd ed.). Oxford: Blackwell.

Sperber, D., & Wilson, D. (1996). Fodor's frame problem and relevance theory (reply to Chiappe & Kukla). Behavioral and Brain Sciences, 19, 530–532.

Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). Searching the Web: The public and their queries. Journal of the American Society for Information Science and Technology, 52, 226–234.

Swanson, D.R. (1977). Information retrieval as a trial-and-error process. Library Quarterly, 42, 128–148.

Swanson, D.R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. Library Quarterly, 56, 389–398.

Swanson, D.R., & Smalheiser, N.R. (1997). An interactive system for finding complementary literatures: A stimulus to scientific discovery. Artificial Intelligence, 91, 183–203.

Swanson, D.R., & Smalheiser, N.R. (1999). Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. Library Trends, 48, 48–59.

Tague-Sutcliffe, J. (1995). Measuring information: An information services perspective. San Diego, CA: Academic Press.

Taylor, R.S. (1968). Question negotiation and information seeking in libraries. College & Research Libraries, 29, 178–194.

Warner, J. (2003). Citation analysis and research assessment in the United Kingdom. Bulletin of the American Society for Information Science and Technology, 30(1), 26–27.

White, H.D. (1981). "Bradfordizing" search output: How it would help online users. Online Review, 5, 47–54.

White, H.D. (1990). Profiles of authors and journals in information science: Some trials of ORBIT's Get command. In Proceedings of the 11th National Online Meeting (pp. 453–459). Medford, NJ: Learned Information.

White, H.D. (1992). External memory. In H.D. White, M.J. Bates, & P. Wilson, For information specialists: Interpretations of reference and bibliographic work (pp. 249–294). Norwood, NJ: Ablex.

White, H.D. (2002). Cross-textual cohesion and coherence. In the CHI 2002 Discourse Architectures Workshop. Retrieved April 7, 2006, from http://pliant.org/personal/Tom_Erickson/DA_White.pdf

White, H.D. (2004). Reward, persuasion, and the Sokal hoax: A study in citation identities. Scientometrics, 60, 93–120.

White, H.D. (2005). On extending informetrics: An opinion paper. In Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2005) (Vol. 2, pp. 442–449). Stockholm: Karolinska University Press.

White, H.D. (2007). Combining bibliometrics, information retrieval, and relevance theory, Part 1: First examples of a synthesis. Journal of the American Society for Information Science and Technology, 58(4), 536–559.

Wilson, D., & Sperber, D. (2002). Relevance theory. UCL Working Papers in Linguistics, 14, 249–290. Retrieved April 7, 2006, from http://www.phon.ucl.ac.uk/publications/WPL/02papers/wilson_sperber.pdf

Wilson, P. (1968). Two kinds of power: An essay on bibliographical control. Berkeley: University of California Press.

Wilson, P. (1973). Situational relevance. Information Processing & Management 9, 457–471.

Wilson, P. (1978). Some fundamental concepts of information retrieval. Drexel Library Quarterly, 14(2), 10–24.

Wilson, P. (1986). The face value rule in reference work. RQ, 25, 468–475.

Wolfram, D. (2003). Applied informetrics for information retrieval research. Westport, CT: Libraries Unlimited.

Yang, K. (2005). Information retrieval on the Web. Annual Review of Information Science and Technology, 39, 33–80. Retrieved April 8, 2006, from http://elvis.slis.indiana.edu/kiyang/pubs/webir_arist.pdf

Yus Ramos, F. (1998). A decade of relevance theory. Journal of Pragmatics, 30, 305–345.

Zipf, G.K. (1949). Human behavior and the principle of least effort. Cambridge, MA: Addison-Wesley.